

Transformative Reading and Writing Synthetic Archives with Language Models

Jonathan Bradley Gallagher
University of Colorado Boulder
Intermedia Art, Writing, and Performance
May 7th, 2021

Abstract

This paper reflects on Electronic Literature projects I created between 2017 and 2020 through interrogating how each project collaborates with an increasingly complex non-human component. Riffing off of Donna Haraway's concept of significant otherness and making kin I speculate on the differences in the significance of the otherness that is engaged with in projects using methods based on combinatorics/chance, statistical models, and vector semantics (contemporary neural-network based language models like GPT-2). While recognizing that each approach involves a reduction in human agency, this reflective paper focuses on the increasing complexity to which this agency is relinquished, and how to deal with presenting this relationship between human and non-human actors. Culminating in a series of projects using OpenAI's GPT-2, the need for a self-reflexive "transformative reading interface" is introduced as a concrete instantiation of Katherine Hayles' concept of a "technotext." A transformative reading interface links a corpus of text to text generated by a language model based on that corpus. Such an interface serves to provide a source of noisy creativity for writing and a way to explore the materiality of contemporary language models for reading while interrogating and respecting the posthuman nature of these artifacts.

In his 2011 book, *Reading Machines: Towards an Algorithmic Criticism*, Stephen Ramsay posits that "Any reading of a text that is not a recapitulation of that text relies on a heuristic of radical transformation."¹ This critical stance serves as foundational support for his assertion of the congruity between "the narrowing constraints of computational logic—the irreducible tendency of the computer toward enumeration, measurement, and verification—" and the goals of literary criticism.² While Ramsay promulgates the merits of a computational, natural language processing approach to literary criticism, both in a way that compliments and augments traditional theory, he is also quick to point out that this effort is not an attempt to reify literary

¹ Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. University of Illinois Press, 2011. p. 16

² Ibid.

criticism as finally relevant through being computationally defined. He is also interested in the artistry involved in developing concrete implementations of algorithmic criticisms.³

Ramsay's aspirations in 2011 were admirable. He imagined an approach to literary criticism that used the rigor of computer science to critically analyze text while using the rigor of literary criticism to interrogate how computational methods could be applied to reading, writing, and understanding literature. In his closing remarks Ramsay offers the following succinct and simple definition of algorithmic criticism as "an attitude toward the relationship between mechanism and meaning that is expansive enough to imagine building as a form of thinking."

However, in a 2013 review of the book, Susan Brown points out that there is an "odd tension between the insistence in this book on algorithmic criticism as offering 'a different scale and...expanded powers of observation' and the very contained examples Ramsay provides."

Comparing *Reading Machines*, to Moretti's *Graphs, Maps and Trees*, Brown says, "In advocating 'distant reading', Moretti makes a small number of large arguments in relation to the study of the novel. Ramsay makes a large number of small readings in order to mount a very big argument about the nature of texts and critical reading."⁴

Brown's points underscore the difficulty in expanding literary criticism through small algorithmic readings, which in Ramsay's account, amounts to implementing many small natural language processing tasks on equally small and focused datasets—not to mention the problematic expectation that a large number of humanities or even digital humanities scholars would want to engage in "building as thinking" at this low level. The desire for a higher-level of engagement of "building as thinking" is equally palpable in Ramsay's vision as it is in Brown's critical remarks.

³ Ibid. p. 85

⁴ Brown, Susan, "Review: Reading Machines. Toward an Algorithmic Criticism." *Literary and Linguistic Computing*, Vol. 28, No. 3, 2013, pp. 480-482

Since the writing of *Reading Machines* in 2011, significant progress has been made in applying artificial intelligence in the form of neural networks to classifying and generating text. In November of 2019, OpenAI⁵ released the full version of their transformer based model called GPT-2 (General Pre-Trained Transformer). This type of neural network differs from previous architectures, such as recurrent neural networks, in that it consecutively builds connections between single words rather than working with sequences. In text based neural-networks words are represented as high-dimensional vectors (think lists of numbers, often in the range of 50-1000 numbers per word)⁶ called “word embeddings,” and the training of the network could be thought of as finding likely paths through the vectors embedded in this high-dimensional space.

In this paper I argue for exploring using GPT-2 as a heuristic for transforming text and as a model for “building as thinking” that allows for a higher-level of conceptual engagement. I argue that this higher-level of conceptual engagement emerges through collaborating with the nature of the vector space representation of words and how GPT-2’s transformer based model builds connections within that space. Algorithmic collaboration of this kind involves a significant relinquishment of human agency as compared to Ramsay’s many building blocks of algorithmic criticism. However, it is important to understand what that algorithmic collaborator is and if one could/should rely on it to do anything interesting in the space of algorithmic reading and writing.

This paper documents two creative experiments using Open-AI’s GPT-2 language model. These playful experiments aim to explore the potential for using contemporary generative text systems as posthuman collaborators for developing creative and scholarly artifacts. These

⁵ <https://openai.com/blog/tags/gpt-2/>

⁶ Jurafsky, Dan, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed, Pearson Prentice Hall, 2009. Ch 6.

artifacts seek to illuminate, rather than obfuscate, the materiality of these machine learning systems; foregrounding the underlying technology with its strengths and limitations on full display while emphasizing the same for the role of human agency within the posthuman collaboration. In this way, these works are disinterested in the “creative Turing Test” and are aligned with the reasoning in the *electronic book review* article, “The Anxiety of Imitation: On the ‘Boringness’ of Creative Turing Tests,” by Dan Rockmore and Kyle Booten. In their analysis of the Hafez⁷ poetry generation system, which involves both a procedural rule based system and the use of recurrent neural networks for classification, Rockmore and Booten argue that the system itself should be considered a work of literary art—not because of the mimetic merit of its output but that systems such as the Hafez system are:

“...models of minds; they may do things that remind us of how our minds use language, and they may do things that seem utterly alien, mechanical. Beyond the failure or success of an “imitation game,” these models – however imperfect they may be – invite us into a recursive and “unstable” consideration of the ways that our most “jealously guarded psychological attributes” do or do not (and should or should not, could or could not) carry the echo of an algorithm.”⁸

At the same time the authors note that the Hafez system won the 2016 “Turing Tests in the Creative Arts” hosted by the Neukom Computational Institute at Dartmouth College.⁹ This ironic nod of acclaim in an article dedicated to criticizing the imitation approach in Creative AI artworks is indicative of the enchantment with an ethos of aesthetic appreciation surrounding AI generated works that is still ultimately entrenched in a liberal humanist perspective.

To counter this liberal humanist impulse these two projects embrace two operational tenets. First, I explore ways of comparing generated text to their corresponding training corpora.

⁷ Ghazvininejad, Marjan, et al. “Hafez: An Interactive Poetry Generation System.” *Proceedings of ACL 2017, System Demonstrations*, Association for Computational Linguistics, 2017, pp. 43–48. *ACLWeb*, <https://www.aclweb.org/anthology/P17-4008>.

⁸ *The Anxiety of Imitation: On the “Boringness” of Creative Turing Tests | Electronic Book Review*. <https://electronicbookreview.com/essay/the-anxiety-of-imitation-on-the-boringness-of-creative-turing-tests/>. Accessed 9 Dec. 2020.

⁹ *Ibid.*

Through employing natural language processing techniques and a bit of text processing, dense sets of hyperlinks are created between generated and original corpora to provide a direct interface for comparing these respective texts. These comparisons reveal both limitations and novel features of the generative text pipeline's behavior. Second, building off of Rockmore's and Booten's idea of "models of minds," I explore designing generative text pipelines inspired by the narratives and relationships that exist between the authors of the training corpora. This exploration is undertaken not to necessarily induce some sort of critical understanding between their works but to serve as frameworks to extend, subvert, and confound Ramsay's notion of a hermeneutics of screwing around.¹⁰ These extensions and subversions are transformative reading and writing with synthetic archives and noisy creativity; concepts that will be elucidated on later in this paper.

Language Models and Materiality

In her 2002 book, *Writing Machines*, Katherine Hayles introduces the term "technotext" to describe a literary work that interrogates the inscription technology that produces it. A technotext has the ability to "mobilize reflexive loops between its imaginative world and the material apparatus embodying that creation as a physical presence."¹¹ The projects discussed in this paper proceed in this spirit and are in part aimed at identifying the attributes of language models and generative text systems that give rise to the salient features of their materiality. While this viewpoint doesn't deny the embodied aspect of this materiality (indeed, special vectorized processors, GPUs, which perform operations on lists of number simultaneously, rather than

¹⁰ Ramsay, Stephen. "The Hermeneutics of Screwing Around; or What You Do with a Million Books." *Pastplay*, edited by Kevin Kee, University of Michigan Press, 2014, pp. 111–20. *JSTOR*, doi:10.2307/j.ctv65swr0.9.

¹¹ Hayles, N. Katherine. *Writing Machines*. MIT Press, 2002, p. 25

serially as on a traditional CPU are necessary for training and running inference with large neural-network based language models) I propose that the significant material feature of language models like GPT-2 are the *representations* of word meanings learned using vector semantics; an unsupervised machine-learning technique that organizes these representations into a high-dimensional vector space. These word embeddings will have the feature that words with similar meanings are clustered around each other in this high-dimensional space, which can be visualized by projecting this space down to three or two dimensions, to produce the “word cloud” visualizations that are often used in the Digital Humanities. This technique is based on the distributional hypothesis, which posits that words that have similar meanings tend to occur in similar contexts—that there is a link between the similarity in the distribution of words and their meanings.¹²

This view of materiality and language models is bolstered by Katherine Hayles’ analysis of the complex nature of materiality and physicality. In her book, *How We Think: Digital Media and Contemporary Technogenesis*, Hayles describes the evolution of human technical innovation as a manifestation of the cognitive skill of attention and its interaction with the interface between physicality and materiality. For Hayles, the physicality of any object is essentially infinite, and it is the role of human attention fussing “with physicality to isolate some particular attribute (or attributes) of interest” from which materiality emerges.¹³ Hayles also highlights the role of conceptual frameworks and methodological strategies that play an essential role in this emergence: “From this infinite array (the various components of a computer in this case) a technotext will select a few to foreground and work into its thematic concerns. Materiality thus

¹² Jurafsky, Dan, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed, Pearson Prentice Hall, 2009. Ch 6.

¹³ Hayles, N. Katherine. *How We Think: Digital Media and Contemporary Technogenesis*. The University of Chicago Press, 2012, p 91.

emerges from interactions between physical properties and a work's artistic strategies." Hayles goes on to say that materiality "depends on how the work mobilizes its resources as a physical artifact as well as on the user's interactions with the work and the interpretive strategies she develops—strategies that include physical manipulations as well as conceptual frameworks."¹⁴

Therefore, it is the focused attention in the form of a concept, the distributional hypothesis, that is instantiated through vector semantics, a mathematical technique implemented in an algorithm, that functions as the significant materiality with which a user can interact and an artist can make strategic decisions. An attractive feature of this conception of the materiality of language models like GPT-2 is how the emergence of the meaning of a word is essentially based on its relation to other words. This synergizes well with the general trend in posthumanism to identify objects, both inanimate and animate, as relational complexes of agential components that cannot be trivially separated from the complex to which they belong. This viewpoint doesn't deny the physical embodiment of the device that performs this computation, but rather, is pointing out that physicality (i.e. the operation of the GPU) would look the same whether it was processing text, images, or sound, and that it is at a higher conceptual level where attention is being deployed in a particular way to bring to the foreground the attributes that give rise to the emergence of the materiality of a language model. For the particular case of language models, it is important to realize that the non-human materiality that we are interacting with doesn't know anything at all about language, grammar, or word meanings in the sense that humans do. Instead, we are making-kin with non-human word embeddings; participating in a sympoietic process like that as described by Beth Dempster by way of Donna Haraway as "collectively-producing systems that do not have self-defined spatial or temporal boundaries. Information and control are

¹⁴ Hayles, N. Katherine. *Writing Machines*. MIT Press, 2002, pp. 32-33

distributed among components. The systems are evolutionary and have the potential for surprising change.”¹⁵

Generative Text Systems and (Non)Human Agency

In order to set a foundation for how text generation with a language model like GPT-2 operates at a different scale and provides innovative ways to read and write text, I will first introduce two earlier projects that also function as text generation systems, but in a less complex fashion. Playing off of Donna Haraway’s idea of significant otherness, I am arguing that the complexity of the potential interaction between a person and a language model like GPT-2 is in and of itself significant because of the otherness of the conceptual space that is interacted with; i.e. high dimensional word embeddings. While this conceptual space is a far cry from what could be called conscious, thinking, or even intelligent, I am arguing that it is complex enough and its details unknown enough, in order for interaction with it to contain the same partial connections described in Haraway’s *Companion Species Manifesto*: “patterns within which the players are neither wholes nor parts.”¹⁶ In this way, the black box criticism of neural-networks—the idea that as these models grow bigger and bigger that it becomes harder and harder to actually understand exactly how they are learning what they are learning—is turned on its head. The black box characteristic of neural-networks is the very quality that manifests the potential and opportunity to engage in a posthuman collaboration of significant otherness.

¹⁵ Haraway, Donna Jeanne. *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press, 2016. pg. 33

¹⁶ Haraway, Donna Jeanne. *The Companion Species Manifesto: Dogs, People, and Significant Otherness*. Prickly Paradigm Press, 2003. Pg. 8

The following projects are meant to illustrate an increasing gradient of complexity in the potential depth of significant otherness that is available to interact with based on the underlying materiality of the different text generation systems.

Kinetic Haiku Generator

The Kinetic Haiku Generator is a webcam based project that incorporates computer vision and a 2d physics simulation engine called Box2d. The Kinetic Haiku Generator creates an environment where the user can see themselves and four colored circles that they can “hit” by intersecting their hand with the circles. (Fig. 1) Each circle is attached to a word type: (red) noun, (blue) verb, (green) adjective, (yellow) adverb, and is assigned a number one through four that is its syllable count. When a given circle is hit a random word is selected from a large dataset of words of a given type that have been organized by syllable count and printed to the screen. Provided the user hits the circles such that the five, seven, five, syllable count for the three lines of the haiku are maintained a series of consonant tones are played. If the user violates the syllable count for a given line by hitting a circle with a syllable count that puts it over the limit for that particular line a frequency modulation is applied to the tone. While the piece requires the body to “write,” and a focus of attention that oscillates between stress and flow through the bio-feedback of the tones, the text generation itself is purely combinatorial; relying on pseudo random number generators for variation. In this way its text generating capabilities are based on the same method (and the outputs even more random and nonsensical due to the large word database size) as Allison Knowles’ and James Tenny’s seminal 1967 computational poem “House of Dust.”¹⁷ While mostly relying on randomness, the constraints of the haiku form and

¹⁷ https://nickm.com/memslam/a_house_of_dust.html (link to an implementation of Allison Knowles and James Tenny’s 1967 creation)

the augmented or virtual form of embodiment that mediates the writing (i.e. watching yourself hit virtual circles on a screen) recalls the constraint based approaches of the Oulipian poets that extends the structures and patterns for the potential of literature into the realm of a kind of augmented reality.

(Fig. 1, Kinetic Haiku Generator)



Markov Text Editor

Another project I developed is the “Markov Text Editor.” This project incorporates N-gram based language models and an ordinary text editor to create a tool that provides a “noisy creativity” for poetry generation: a subtractive and additive sculptural approach of erasure and word addition employed to “carve” poems out of the large blocks computer generated text. N-gram language models are familiar additions to many commonly used programs and are responsible for text completion/prediction functionality in applications such as text messaging.

An N-gram model simply keeps the count of unique words (or letters) and the unique words (or letters) that follow a given word (or letter) in a given corpus. So for instance in the following small corpus: “The dog chased the fox that chased the duck who bit the fox.” a 1-gram model for the word “the” would be:

the: dog (1), the: duck (1), the: fox (2)

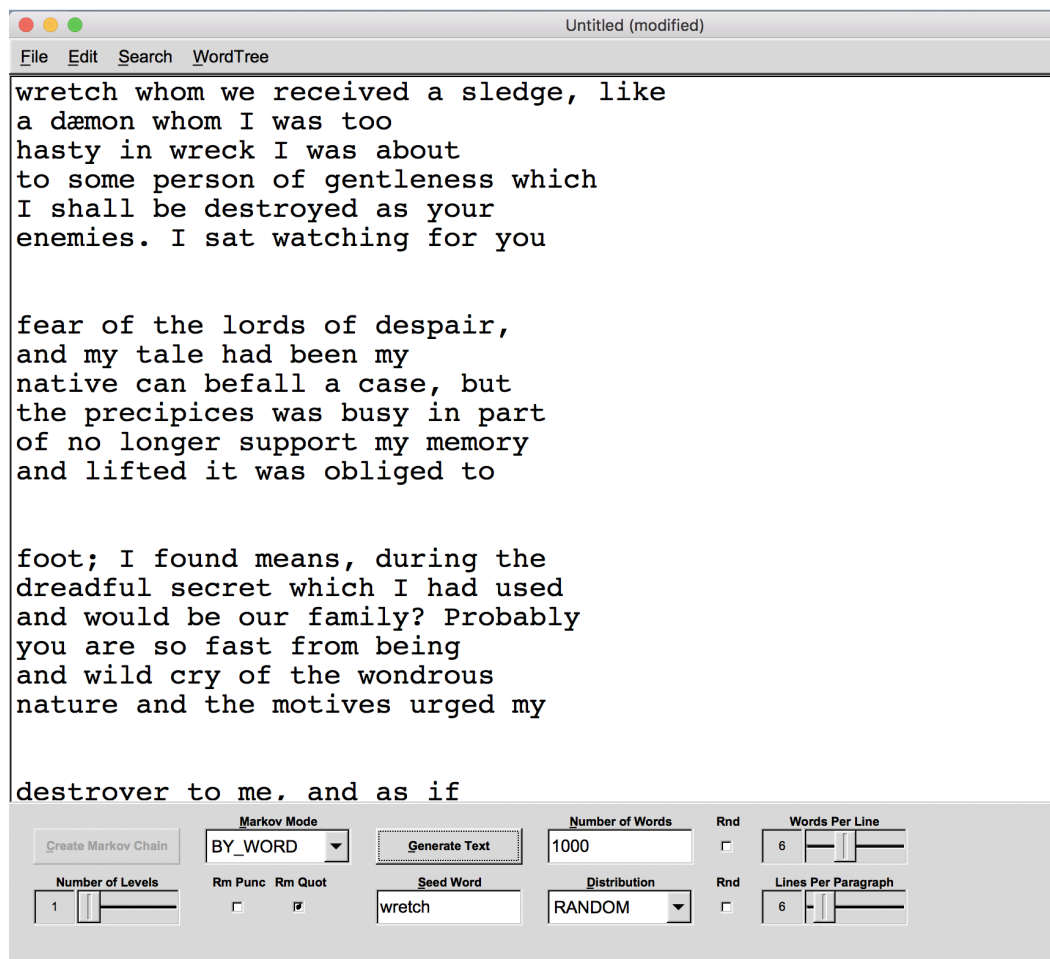
meaning there is a 50% chance that “fox” follows “the” and a 25% chance for “duck” and “dog” to follow “the.” Once such a model has been built up from a large corpus of text, the probability distribution of the model is sampled, usually by using a pseudo-random number generator in order to generate text. Claude Shannon, the founder of information theory, was the first to describe this process known as “Shannon’s Game” in his investigations with using N-grams to compute approximations of English word sequences.¹⁸ Compared to the Kinetic Haiku Generator, the resulting output will be much more readable as only words or letters that appeared in the original corpus can possibly appear in the generated text. In fact, as the N-gram model increases in size, it will approach reproducing the corpus text verbatim. However, divergences from meaningful sentences do occur, especially when you think of root words like “the” that have many possible different words that could potentially follow them.

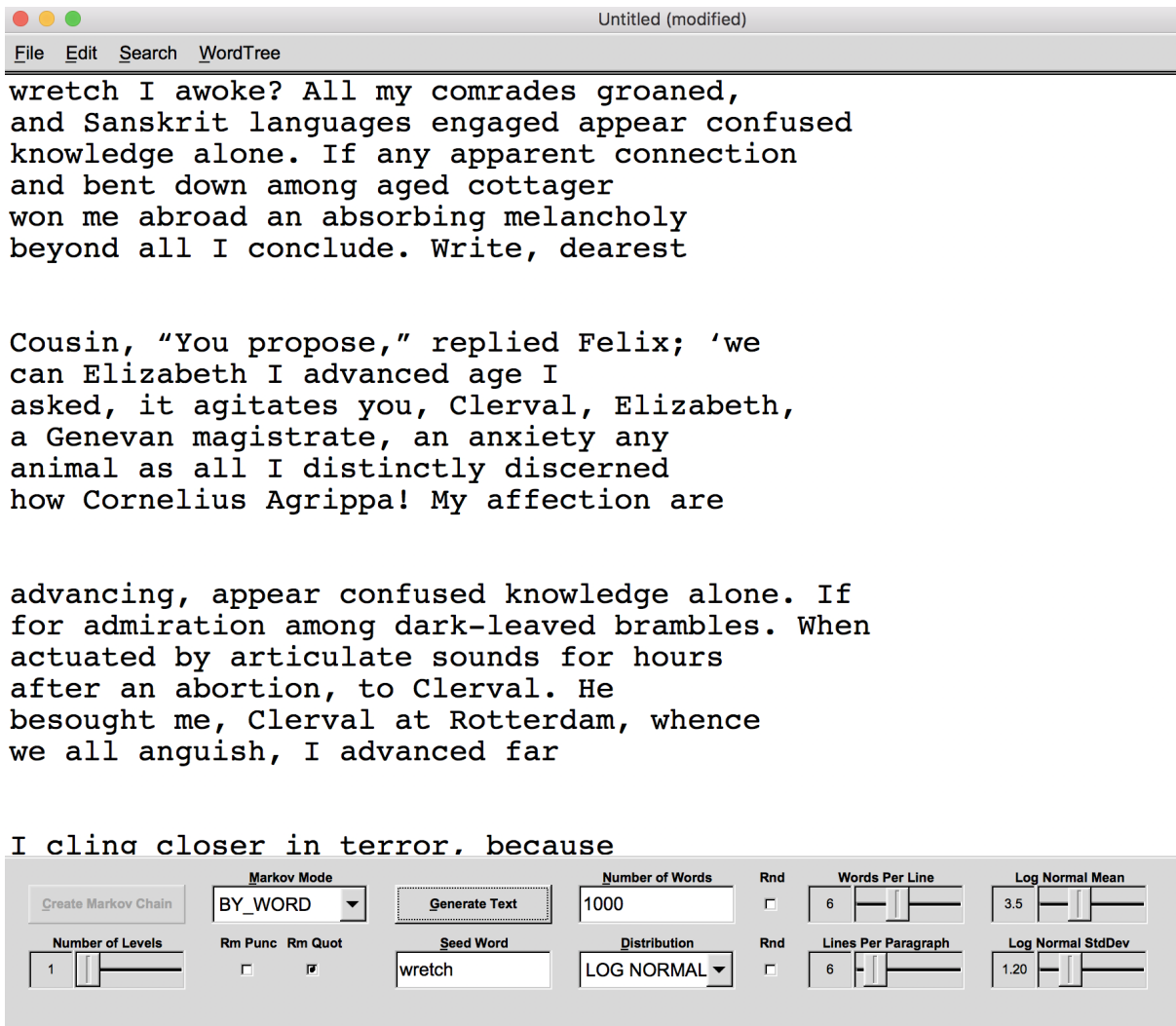
In an effort to explore ways in which the agency of the machine could be extended in the Markov Text Editor, I experimented with using different distributions of numbers to sample the probability distributions of the N-gram models. Some interesting cases are the effect of sampling an N-gram probability distribution with a log normal distribution which I have found creates alliterative sequences of text. Another case is using the distribution of numbers that emerge out

¹⁸ Jurafsky, Dan, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed, Pearson Prentice Hall, 2009. Ch 3.

of the logistic map; a famous iterative equation from chaos theory. This distribution of numbers oscillates between regularity and randomness. Sampling an N-gram language model with this kind of distribution creates sequences of repetitive text followed by more heterogeneous text. All of these techniques are examples of what I call *noisy creativity*; sources of non-random, but noisy and novel patterns. This is especially true when the N-gram language model is applied to characters which allows for the spontaneous creation of new words (see the poem *Southerine*, below (Fig 3)).

(Fig. 2a, n-gram language model of Mary Shelley's *Frankenstein*, sampled randomly)





(Fig. 2b. N-gram model of Mary Shelley's *Frankenstein* sampled with a log normal distribution, notice the alliteration, especially of "a" words)

Southerine

shit's outsiderably litterin'
with Southerine

they've been sick and shiftin',
sittin' kitchen, blowin' smoke,
bitchin' blind,
bad on silver sheen

mornings, I stay
at the park
across the street
man my submarine

afternoons, I play
to scare off, to get off,
to take one off, every mix

their sweat rolls over
they stop, they piss

Stella, promise me flight

excise me
unite this string of smoke
there is no sound

just an irregular heartbeat
fake.

with a love for this room
to be anywhere else

she lives, entwines,
dilates lines of courtships down
so, start counting anywhere,
start today

Stella says
wave at puffy eyes
for what water ricochets
birds peck clean

toilet flush and swirl
powdered pearl
bubbles drown, these mean:

stares and streaks
with a white clay pipe

that icky-sticky heat

close your palms
keep your eyes shut tight

see the kudzu crop
the swirl of dark

can you feel it creep?

open your eyes, look up—do you see
neverender's
lake?

let us go to the park
across the street
take pictures of your feet—
mount animal forms
clench diagonal traces
a flight back and forth
before we eat

I know this time
buffeting below
much trouble whistles

but for now we are free from fight
the red facade
the leaves on tight
the sounding wall
what sinking sunlight twilight
makes.

before sure hidden beat

animals here
shrinking-seen, choking-cape
ESCAPE-IT-APE

in a desert
land me

every autumn
strand be

Stella-lala-lala
boom-a-rang me

speed me
back and forth

I know this time
I jump off the swing

(Fig. 3, poem “carved” out Markov Text Editor output)

Do GPT-2's Dream of Electric Poetry?

GPT-2 (General Pretrained Transformer) is known as an attention based neural network. While the base model was trained on 20 million web pages, its attention based architecture makes it especially good at transfer-learning. This allows for fine-tuning the large model on a much smaller corpus of text; creating a new model that leverages the lower-level features learned by the base model (things such as likely word order, apparent grammar, etc.) while taking on the style and content of the fine-tuning corpus. Other neural-networks like traditional recurrent neural networks can be used to generate text, but can also be used to classify it; i.e. identify if a text was written by a certain author once trained on a subset of that author's corpus. These experiments use both of these types of neural-networks to create what I call "hyper-carving" text generation pipelines.

The first GPT-2 experiment discussed references to the title of Philip K Dick's book "Do Androids Dream of Electric Sheep?" not from any concern for whether or not GPT-2's trained on poetry dream of poetry, or of anything at all (they of course don't dream) but with the general anxiety surrounding the social impact of the creative potential of artificial intelligence. In Dick's book the narrator is concerned with the social implications of owning an electric sheep in a world where animal ownership is a symbol of socio-economic status. The title creates a double ambiguity that creates a posthuman crisis of identity: would androids be sophisticated enough to carry the shame of having an electric sheep in lieu of a real one as a status symbol, or is the narrator actually an android and already in this absurd situation?

In this project I am building off the idea of "carving" text out of the output from N-gram language models. This concept was eloquently captured by David Jhave Johntson in his book, *Aesthetic Animisms*:

“A block of A.I.-generated text, massive and incomprehensible, can exude the presence of solid stone. Here, the cursor exists like a chisel; I called this human-editing part of the process, carving.”¹⁹

Hyper-carving

Building off this concept I set out to create what I call a “hyper-carving” pipeline that creates end to end computer generated text. The design of this pipeline is informed by the social and critical relationship between the poets John Ashbery, W. H. Auden, and Wallace Stevens.

In a 1980 interview with David Remnick, John Ashbery describes the formative impact that the poetry of W. H. Auden had on his writing: “I am usually linked to Wallace Stevens, but it seems to me Auden played a greater role. He was the first modern poet I was able to read with pleasure...”²⁰ In another interview Ashbery identifies Auden as “one of the writers who most formed my language as a poet.”²¹ For Auden’s part there was a mutual yet mysterious appreciation for the younger poet’s work; Auden awarded Ashbery the Younger Yale Poets prize for his book of poems *Some Trees* with the caveat: “...that he had not understood a word of it.”²²

Playfully building off this narrative I devised a text generation pipeline that involved fine-tuning GPT-2 models on the poetry of John Ashbery and W. H. Auden. These two subsequently produced models were then put into conversation with each other: the output of the Ashbery model serves as an input prompt to the Auden model and vice versa in a feedback loop. The generated text is then classified with three separate recurrent neural networks: one trained on the poetry of Ashbery, one trained on the poetry of Auden, and one trained on the poetry of

¹⁹ Johnston, David Jhave. *Aesthetic Animism: Digital Poetry’s Ontological Implications*. The MIT Press, 2016.

²⁰ Ashbery, John, Remnick, David, (September 1980). John Ashbery in conversation with David Remnick. *Bennington Review*

²¹ Pattel, Cyrus R, (Ed.). (1994). *The Cambridge History of American Literature: Volume 8, Poetry and Criticism, 1940-1995*, pg. 214

²² Orr, D. Smith, D. (2017, Sept. 3rd). John Ashbery Is Dead at 90; a Poetic Voice Often Echoed, Never Matched. *The New York Times*.

<https://www.nytimes.com/2017/09/03/arts/john-ashbery-dead-prize-winning-poet.html>

Wallace Stevens. (Fig. 4) This classification part of the pipeline serves to filter the large amount of generated text into smaller subsets that have different characteristics. For instance, texts generated from the Ashbery and Auden models that have the highest score from the Wallace trained RNN may have more novelty, and outputs that have high scores from the RNNs of the author they were trained on may be overfit.

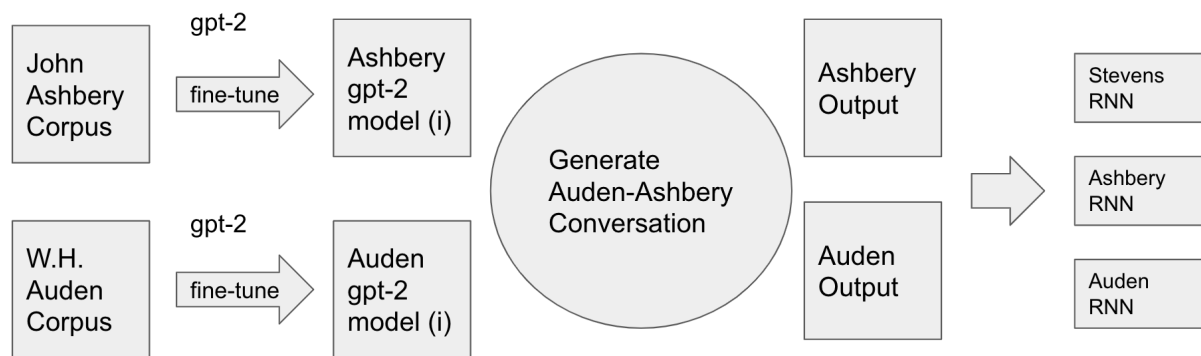
One output of this process was a short, partially machine curated, and partially human curated list of poems that made use of these classification networks to find novel poems by searching for poems generated by the Ashbery model, for instance, that had a higher Auden score and an even higher Stevens score. A presentation of these poems can be seen here:

<http://jbgallag.ddns.net/aa/>

This sort of output is uncomfortably close to a creative Turing Test, and while the success of the hyper-carving method using recurrent neural networks trained on the different poets to filter a large amount of GPT-2 output is a useful methodology to keep in the toolkit, I was left unsatisfied with this type of output as a presentation of the process.

Transformative Reading and Writing, Synthetic Archives, and Noisy Creativity

This dissatisfaction led to the development of a concept to present the generative output of GPT-2 models with the corpora that they were fine-tuned on. With such an approach it is possible to address practical matters such as whether or not the training was overfit (which will reproduce the learned text verbatim) and to explore the deeper interconnections between how the generated text and the original corpus are converging and diverging. This is accomplished by



(Fig. 4 A hyper-carving pipeline for outside classification)

creating a dense set of hyperlinks between the generated text and the original corpus that links all occurrences of significant words between them. These links can have multiple destinations, i.e. a word in a generated text may link to multiple locations in the original training corpus and vice versa, so the interface for traversing them cycles through these possibilities. Currently these transformative readings are presented in a web interface that allows for viewing generated text and the original corpus side by side with bi-directional multi-links between them. In the case of the “Do GPT-2s Dream of Electric Poetry?” project, it is also possible to filter the generated output based on each generated poem’s classification score for each of the three recurrent neural networks trained on Ashbery, Auden, and Stevens. The current version of this project can be interacted at the following links.

<http://jbgallag.ddns.net/poemData> (Do GPT-2s Dream of Electric Poetry)

<http://jbgallag.ddns.net/GPC> (Gnarly Posthuman Conversations: John Ashbery, W. H. Auden, Wallace Stevens and GPT-2, which is a new iteration of the project)

Media Archaeology Language Model

In this project I am interested in exploring the potential of using the transformative reading interface as a creative tool for exploring scholarly texts. The goal for the development of this system is to create a synthetic archive of generated text that is densely linked back to the original archive which is created by asking the fine-tuned model questions and propagating the responses in the form of an exponential graph. In the first iteration of this project around 900,000 tokens of text related to the field of Media Archaeology (see website below for references) was used to fine-tune a GPT-2 model. The subsequent model is prompted with a question which produces four responses. The four responses are used as prompts to each generate four more responses, etc, up to a certain exponent. In the current version of the process, this is done five times, which creates 4096 unique combinations of interconnected (outputs of previous steps serving as prompts for subsequent steps) responses to a single question. This creates a nonlinear, branching structure that is inspired by Zielinski's *Deep Time of the Media*; his analogy between Thomas Hutton's discovery of unexpected strata of slate under layers of granite and the stratification of the history of media; a deep time interpretation that resists a linear understanding of a "predictable and necessary advance from primitive to complex apparatus."²³

The hope is that a transformative reading system like this can be used to build up a rich series of interconnected responses to questions that probe a given research topic, which are in turn linked back to the original texts, enabling an environment that contains a noisy creativity that can provide an innovative and novel way to approach creating critical discourse. This could be a tool for creating a traditional paper by collaborating with such a system, or simply by creating such a system itself.

²³ Zielinski, Siegfried. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*. MIT Press, 2006. pp. 4-7

The anti-disciplinary spirit of media archaeology and its tendency to blend disparate disciplines (I'm thinking of Zielinski's appropriation of geological methodology and Thomas Elsaesser's appropriation of the laws of thermodynamics to create a new and powerful way to think about the history of the cinema²⁴ as poignant examples) makes the field and its archive of discourse an apt partner to the techniques that can be employed with GPT-2. In particular, I have been interested in exploring a research project that looks at media archaeology through the lens of punctuated equilibrium, an alternative theory of evolution put forth by Stephen J. Gould, that states that once a species emerges they often stay the same until they go extinct. Speciation on the other hand seems to occur in punctuated, disruptive spurts, that are often associated with the sub population of a species becoming geologically isolated from the main population. I am interested in looking at the evolution of media technology through this methodology. A transformative reading environment to support this could involve creating separate GPT-2 models, one based on media archaeology scholarship and the other on scholarship concerning punctuated equilibrium, and putting them into conversation with each other as with the poetry project. Another idea would be to train a GPT-2 model first on media archaeology scholarship, and then train the same model again in a transfer learning approach on the punctuated equilibrium material to create a blended model.

In further iterations of this project I am interested in extracting the questions that have happened to occur in the generated text and adding them as prompts to the process; allowing for a natural source of feedback that allows GPT-2 to directly change the course of the synthetic archive.

MALM: <http://jbgallag.ddns.net/MALM/>

²⁴ Roberts, Ben, and Mark Goodall. *New Media Archaeologies*. Amsterdam Univ. Press, 2018. ElsaEsser, Thomas, *Cinema, Motion, Energy and Entropy pp.* 107-132

Conclusion:

The projects surveyed in this paper represent small steps towards developing new ways of reading, writing, and interacting with text using language models. Building off of Stephen Ramsay's idea of reading as a type of transformation, I have begun an investigation in using generative language models like GPT-2 to create environments of transformative reading, where synthetic archives of text and the corpora they are trained on can be viewed together to be scrutinized, explored, and appreciated as works that are the result of posthuman collaboration. I'm arguing for the use of neural-network based language models over the sort of nuts and bolts natural language processing techniques promulgated by Ramsay in *Reading Machines*, because they afford us a way to work in a conceptual space. We can create conversations between dead poets, and create synthetic archives of exponential answers to questions in a scholarly field and produce this text at scale.

In Zielinski's investigation into the deep time of media, he describes the type of media history he is interested in as curiosities: "By curiosities, I mean finds from the rich history of seeing, hearing, and combining using technical means: things in which something sparks or glitters—their bioluminescence—and also points beyond the meaning or function of their immediate context of origin."²⁵ As a prompt, Zielinski's quote engages me to respond with a couple of questions: As language models continue to evolve (OpenAI has already developed a GPT-3 model that is 100 times bigger than GPT-2) could the spirit of Zielinski's "curiosities" be embodied in the development of synthetic histories on which a simulated discourse is performed? Could the resultant synthetic discourse provide anything useful for understanding the past and the future from a posthuman perspective?

²⁵ Zielinski, Siegfried. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*. MIT Press, 2006. pp. 34

Annotated Bibliography

1. Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. University of Illinois Press, 2011.

Reading Machines provides a great set of examples of how natural language processing techniques can be applied to critical discourse to perform what he calls algorithmic criticism. His detailed approach pays homage to the artistry behind the design of such systems, but still presents a rather low-level approach that while he successfully argues is connected to traditional critical analysis, seems unlikely to be widely practiced. For my purposes his idea of all reading being a “heuristic of transformation” and the idea of “building as a type of thinking” were fantastic jumping-off points for my arguments about the higher-level type of “building as thinking” that can be done with contemporary transformer language models like GPT-2.

2. Brown, Susan, “Review: Reading Machines. Toward an Algorithmic Criticism.” *Literary and Linguistic Computing*, Vol. 28, No. 3, 2013, pp. 480-482

This review was useful for situating the difficulty in the adoption of Ramsay’s techniques and helped establish the notion of a desire for a higher conceptual level of interaction within algorithmic criticism.

3. <https://openai.com/blog/tags/gpt-2/>

This is the main site for the GPT-2 language model. The language model uses 1.5 billion parameters, these are the weights (single numbers in the end) whose particular values are the result of the training of the initial model and the fine-tuning process. These weights describe the cumulative prediction of what the next word in a sequence might be, in this way they describe likely paths through the vector space representation of the words in a corpus.

4. Jurafsky, Dan, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed, Pearson Prentice Hall, 2009.

This book provides a comprehensive overview of natural language processing both from a technical perspective and with great examples of underlying linguistic motivations for various NLP techniques. The book covers everything up to recurrent neural networks. In this paper I relied on it primarily for its section on vector semantics and word embeddings. The link between word embeddings and neural networks is important because these serve as the input to the neural-network, and it is the underlying structure in the word embeddings that is being learned in an unsupervised way, in the case of GPT-2. All the other operations of a neural network, the feed forward calculations and back propagation methods that adjust the weights are pretty similar for any type of neural

network. The motivation for using word embeddings is based on the distributional hypothesis, which became an important concept for describing the important aspect of materiality in these language models, from the perspective of this paper.

5. Ghazvininejad, Marjan, et al. “Hafez: An Interactive Poetry Generation System.” *Proceedings of ACL 2017, System Demonstrations*, Association for Computational Linguistics, 2017, pp. 43–48. *ACLWeb*, <https://www.aclweb.org/anthology/P17-4008>.

This paper describes the “Hafez” poetry system, which is the canonical example of a hyper-carving pipeline as far as I know. The system generates sonnets through a partly procedural and partly machine-learning based pipeline. Sonnets are constructed procedurally through databases connecting rhyming words and rules are applied to enforce iambic pentameter. The resulting poems being largely combinatorially created are often nonsensical. A recurrent neural network trained on song lyrics (and other types of corpora) is used to filter out these nonsensical outputs.

6. Ramsay, Stephen. “The Hermeneutics of Screwing Around; or What You Do with a Million Books.” *Pastplay*, edited by Kevin Kee, University of Michigan Press, 2014, pp. 111–20. *JSTOR*, doi:10.2307/j.ctv65swr0.9.

While Ramsay describes a hermeneutics of screwing around as a valid way to deal with a million books, one of the questions I am asking in this paper is what it means to extend that hermeneutics to creating synthetic archives and works of literature, which could vastly exacerbate the situation Ramsay is describing in this essay?

7. Hayles, N. Katherine. *Writing Machines*. MIT Press, 2002

I take Hayles idea of technotext as the inspiration for the transformative reading interface I develop in these projects. These projects are an initial attempt to instantiate the idea of a literary work that interrogates the inscription process that produces it and “mobilizes reflexive loops between its imaginative world and the material apparatus embodying that creation as a physical presence.”

8. Hayles, N. Katherine. *How We Think: Digital Media and Contemporary Technogenesis*. The University of Chicago Press, 2012

Hayles complex description of materiality, one that relies on both physicality and the focusing of attention to create the emergent phenomena of materiality, was essential in trying to decide where the important aspect of materiality in language models like GPT-2 lies. Hayles allowing for conceptual frameworks and artistic strategies as aspects of that focusing of attention gave me the permission to look at the distributional hypothesis in the form of vector word embeddings as the important aspect of materiality that emerges in the way Hayles describes. GPT-2 is a neural network, and as such it has a large amount of weights that are adjusted through the training process which often involves a feed forward phase and back propagation phase, which are fairly ubiquitous to any type of

neural network. It is the word embeddings, which serve as the input into models like GPT-2, where the emergence of materiality occurs through the focusing of attention on a particular (virtual) physicality.

9. Haraway, Donna Jeanne. *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press, 2016.

Haraway's development of sympoiesis helps situate the type of relationship and context of interpretation that I see existing in the interaction with language models like GPT-2.

10. Haraway, Donna Jeanne. *The Companion Species Manifesto: Dogs, People, and Significant Otherness*. Prickly Paradigm Press, 2003.

I riff off of Haraway's idea of Significant Otherness, drifting ever so slightly away from the sense of the word significant as I believe she intended. Through surveying two other text generation projects that I did in the past, that did not use neural-networks or word embedding models of words, I attempt to demonstrate that the "significance" of the otherness that I interact with in both the "Kinetic Haiku Generator" and the "Markov Text Editor" is of a lesser value, or level of complexity for potential interaction, as compared to GPT-2 whose nexus of material interaction is the black box of a neural network.

11. https://nickm.com/memslam/a_house_of_dust.html (link to an implementation of Allison Knowles and James Tenny's 1967 creation)

Included as the quintessential example of a combinatorial, randomly sampled, generated text system.

12. Johnston, David Jhave. *Aesthetic Animism: Digital Poetry's Ontological Implications*. The MIT Press, 2016.

Jhave's quote is important for situating the model in which algorithmically assisted writing has generally followed. In the hypercarving approach this view of the late night sculptor is excised. This view of carving still places the human as the final designer, and augments of reality, and is still overly drenched in liberal humanism.

13. Zielinski, Siegfried. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*. MIT Press, 2006.

Zielinski's ideas of the deep time of media and curiosities help situate why such an enterprise as transformative reading and writing with synthetic archive may have any value at all. As Zielinski hopes to find curiosities, innovations and moments of time that history has looked over, the process of generating text with an advanced language model like GPT-2, and the techniques of dealing with that output, could possibly lead to a system that can be used to simulate the sort of fissured, stratified, and buried histories that Zielinski is interested in develop methods for uncovering.

14. Roberts, Ben, and Mark Goodall. *New Media Archaeologies*. Amsterdam Univ. Press, 2018. Elsaesser, Thomas, *Cinema, Motion, Energy and Entropy* pp. 107-132

Thomas Elsaesser's "Cinema, Motion, Energy, and Entropy," the laws of thermodynamics are blended with the media theory and history of the cinema. Thinking of the cinema in terms of energy exchanges, both in its physical operation, in the images it depicts, in the physiological stimulation of the viewer, to the nineteenth century preoccupation with maximizing the energy of human labor, a powerful language for Elsaesser emerges that allows him to conceive of cinema as "a discourse engaged in training the body and the senses in such a way that we experience as entertaining what society requires from us as a necessity, in order for its particular form of energy transformation -in this case, capitalist production methods - to function."