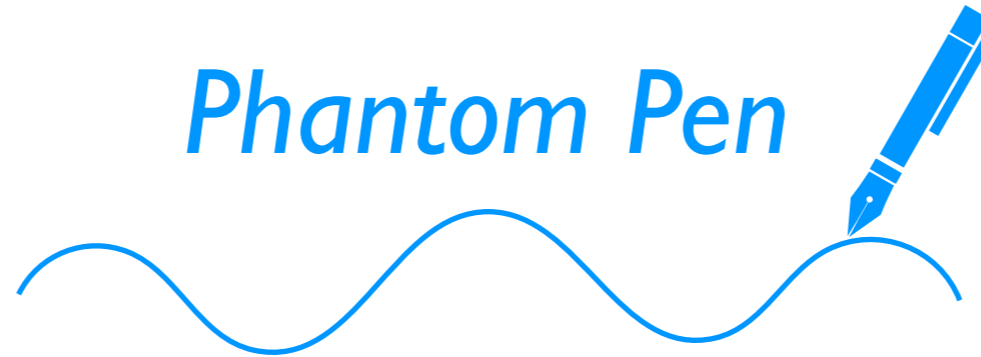


Phantom Pen



Your digital ghost writer

Mobolaji Williams — Insight Data Science

<https://youtu.be/W-UbsOGW3KQ>

#1 Beginning is Hard

Common Writing Process

1. You try to start writing something
2. Manage to write one or two sentences
3. Then have no idea how to continue

(Future Pulitzer Prize Winner)



Need #1

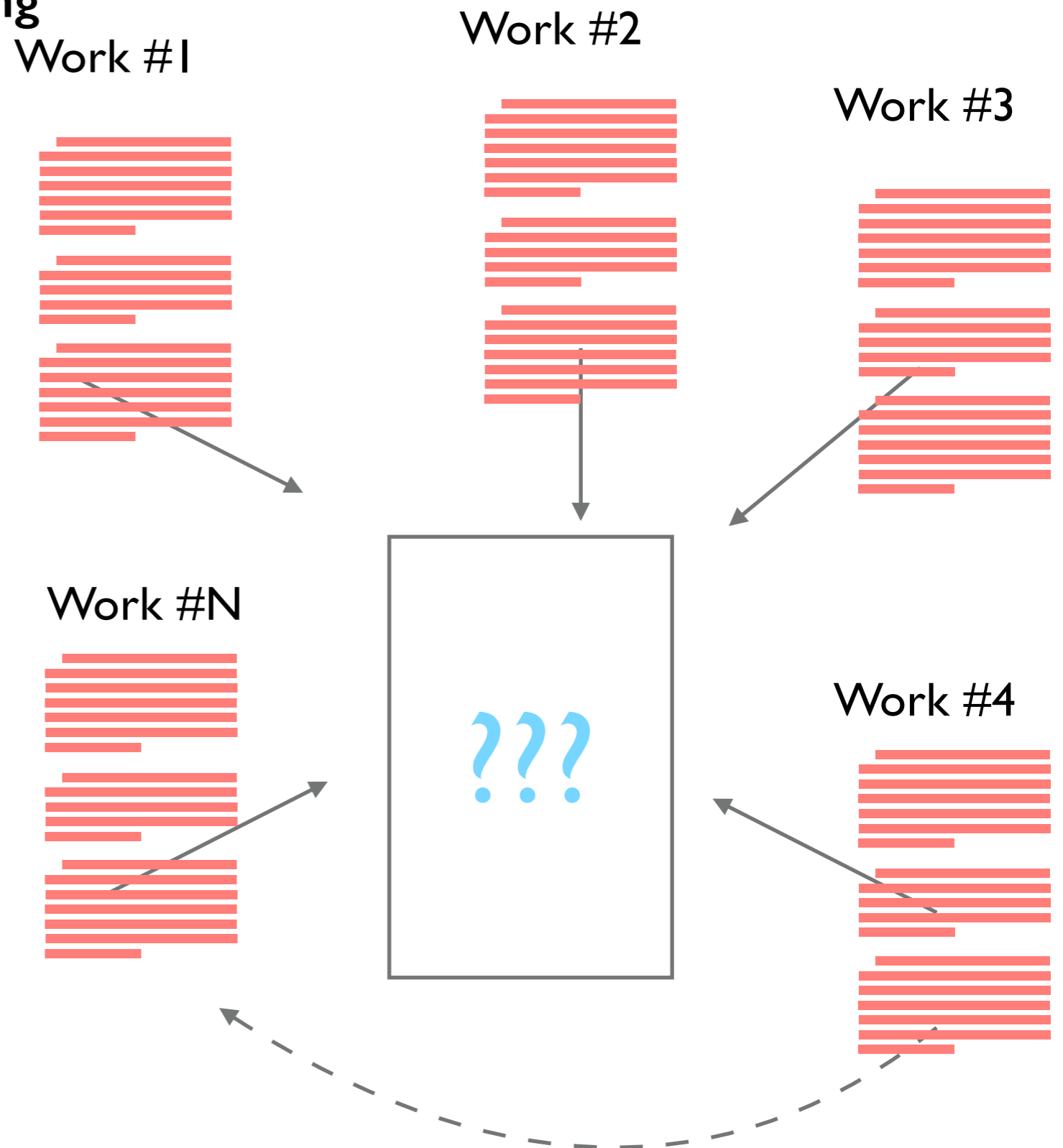
Move past a mental block and generate new ideas for writing

#2 Integrating Style is Also Hard

Common Way to Improve Writing

1. Collect many different written texts
2. Extract stylistic features of text
3. Somehow incorporate styles into your own writing

➔ **Need #2**
Integrate the styles of many different writers



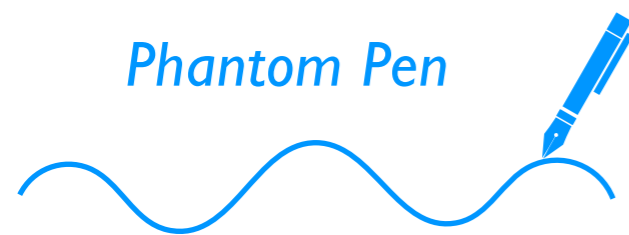
Responding to Writing Needs with Machine Learning

Need #1

Move past a mental block and generate new ideas for writing

Need #2

Integrate the styles of many different writers



Answering the Needs

A program that writes in response to a prompt using the combined styles of many different writers

Author's prompt:



*Created from a compilation of the styles of many writers



How was this built?

Workflow



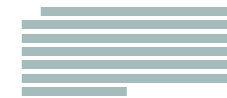
Data Collection



Train Model



Make "predictions"



Today is the first
day of the rest _____

of your life.

Final Essay-Generator

Data Collection

2000+
Articles

Nine Article Categories

- Business
- Education
- Entertainment
- Health
- Ideas
- International
- Politics
- Science
- Technology

Webscraping Objective:
Collect writing of various styles
for different genres

160 *Short
Stories*

62 Authors

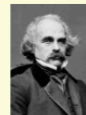
- Anton Chekov
- Henry David Thoreau
- Jack London
- H.P. Lovecraft

20 *Books*
(*Gutenberg*)

Including

- Pride and Prejudice
- The Count of Monte Cristo
- Robinson Crusoe
- Moby Dick

The Atlantic



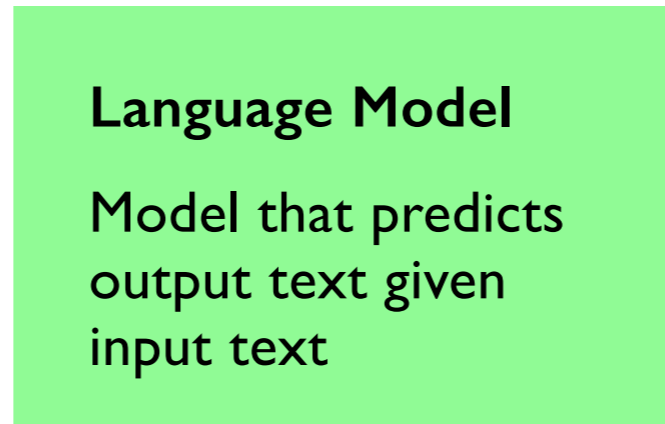
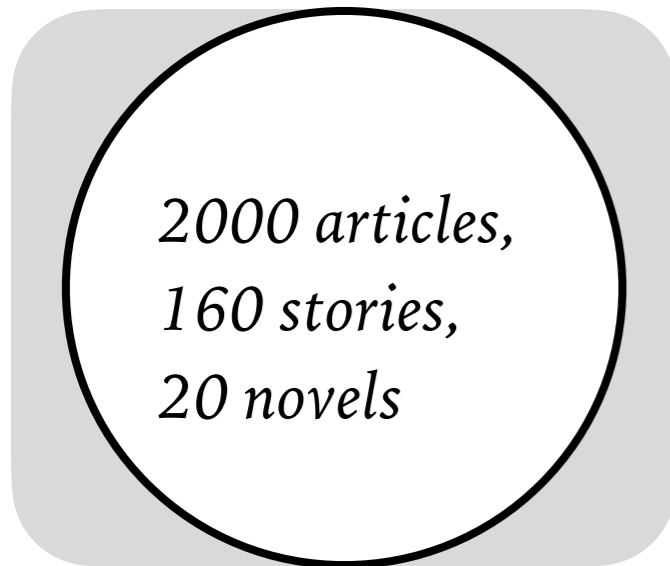
100 Great Short Stories

Okay, I lied. There are so many great short stories that I was unable to trim the list to 100 titles; so here are 160 Great Short Stories for you to enjoy. Click a button to find the best short stories from the authors below. We have a great collection of [Short Stories for Students and Children's Stories](#).

**Project
Gutenberg**

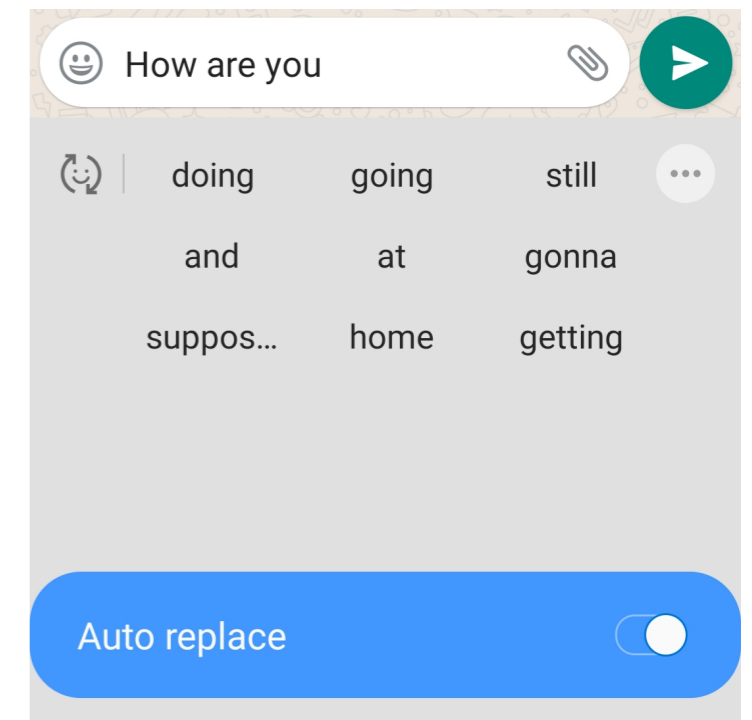
Language Model Options

Collected Data

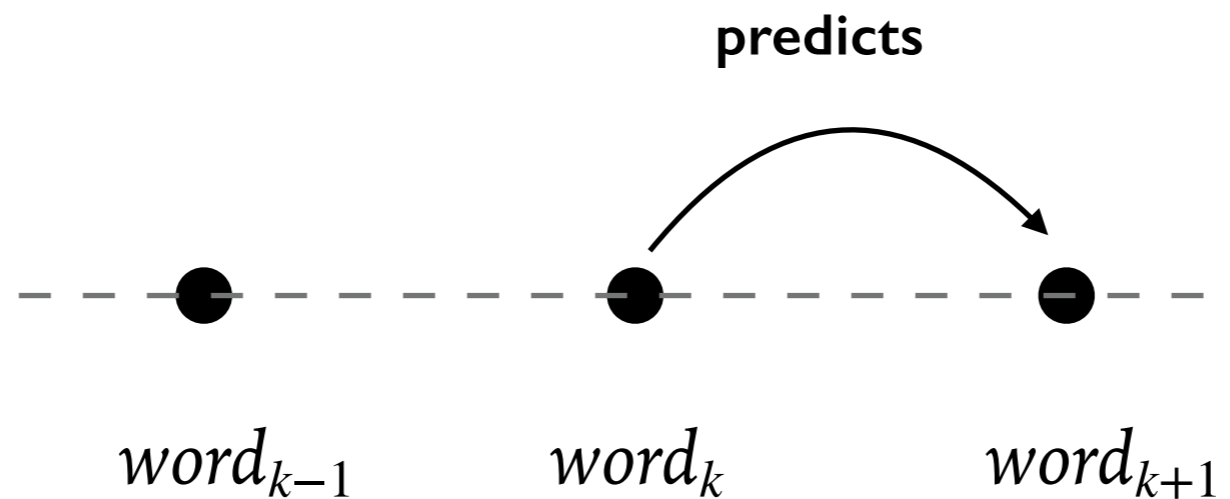


Option #1

Markov Chain



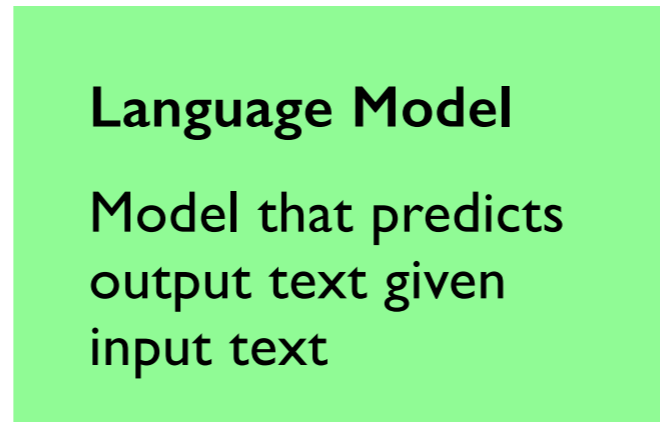
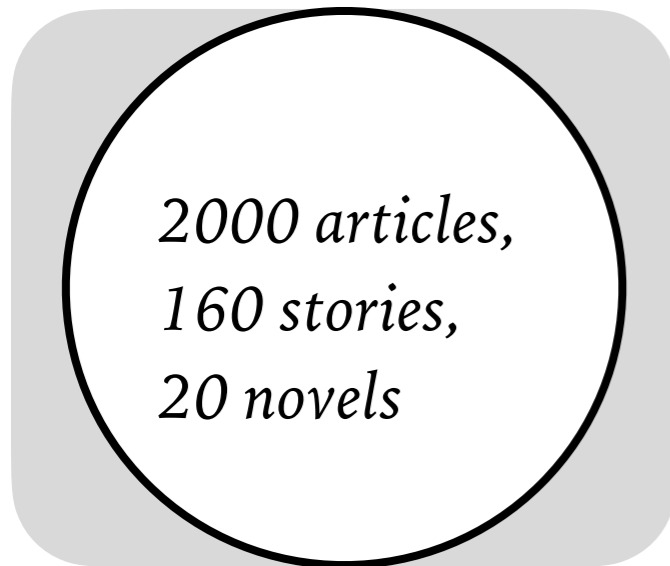
(WhatsApp)



Problem: Cannot recognize context

Language Model Options

Collected Data



Option #2

Recurrent Neural Network

Dinner next week

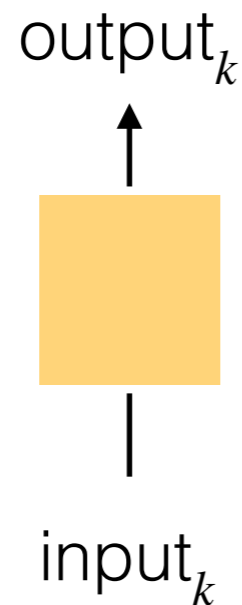
Evan Brown, Maalika Patel

Dinner next week

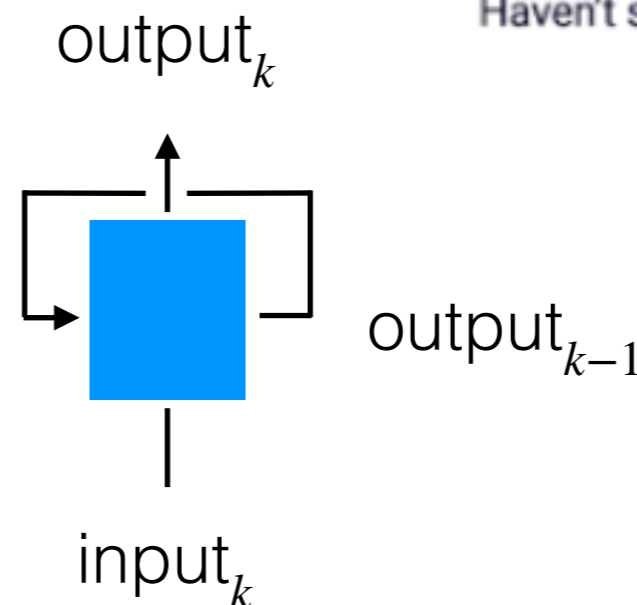
Hey Evan and Maalika,

Haven't seen you in a while!

(Gmail Smart Compose)



Vanilla Neural Network

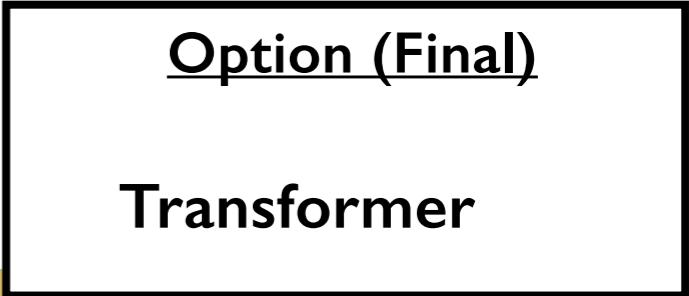
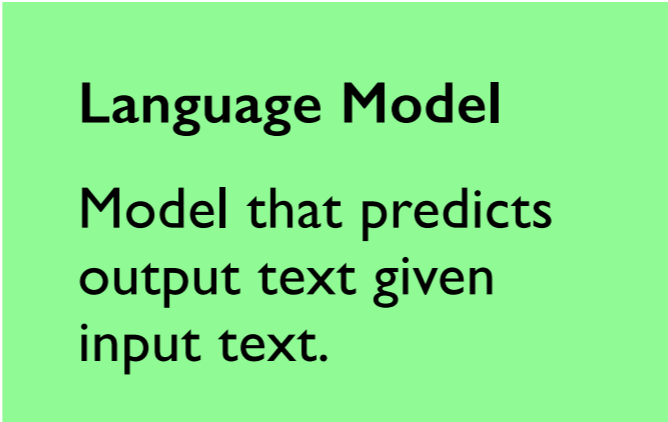


Recurrent Neural Network

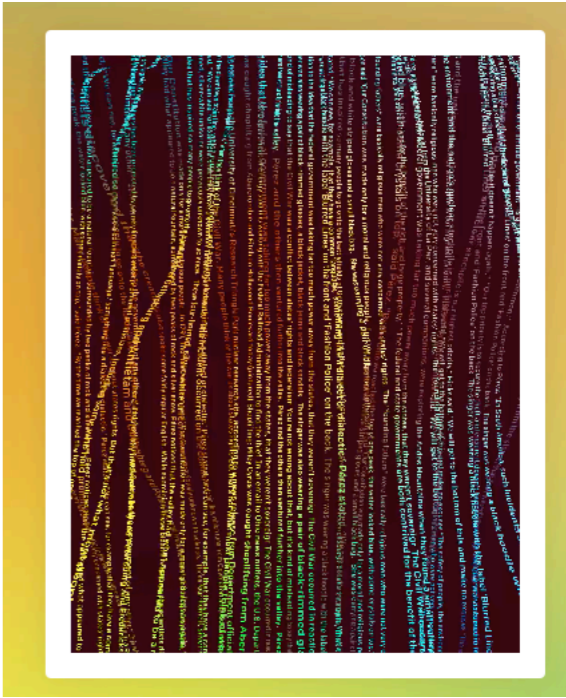
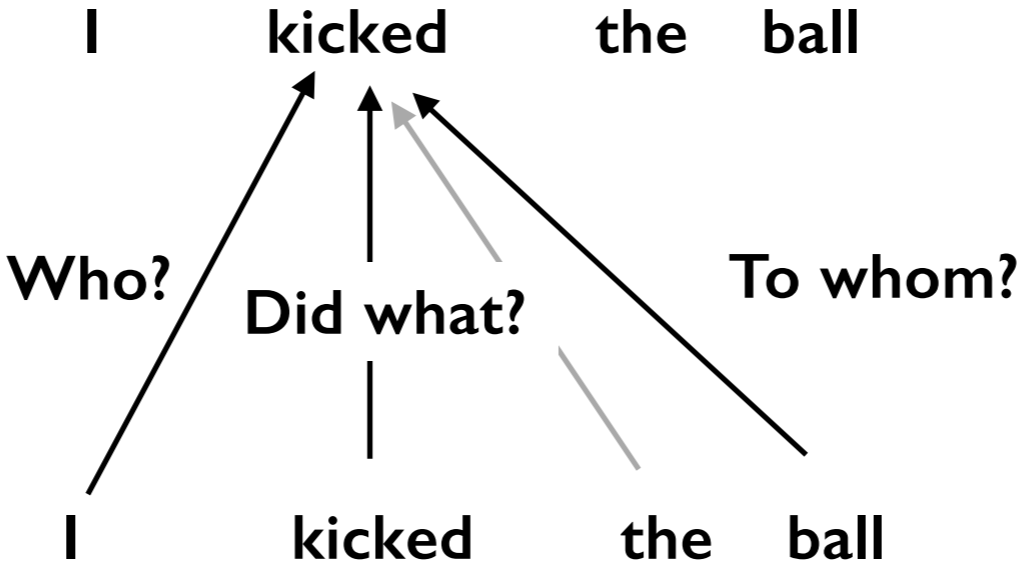
Problem: Memory
too short for long-
term coherence

Language Model Options

Collected Data



**Better Language Models
and Their Implications**



(Open AI)

Problem: Takes a long time to train

Benefit: Can reproduce semantics and structural relationships in words

**Generalized
Pre-trained
Transformer-2**

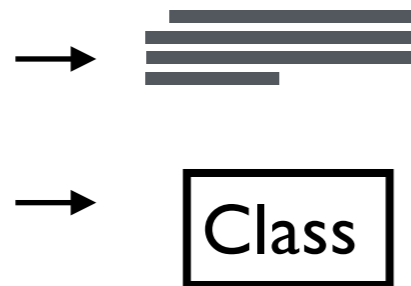
Functions of application

MAIN FUNCTION

SIMPLE GENERATE

Generates an essay given a **prompt** and a chosen **class**

- User entered prompt



Re-trained GPT-2



New Essay



- User chosen class

Business — Education — Entertainment — Ideas
— International — Health — Politics — Science
— Technology — Short Story — Gutenberg

Helping with Writer's Block and Style Integration

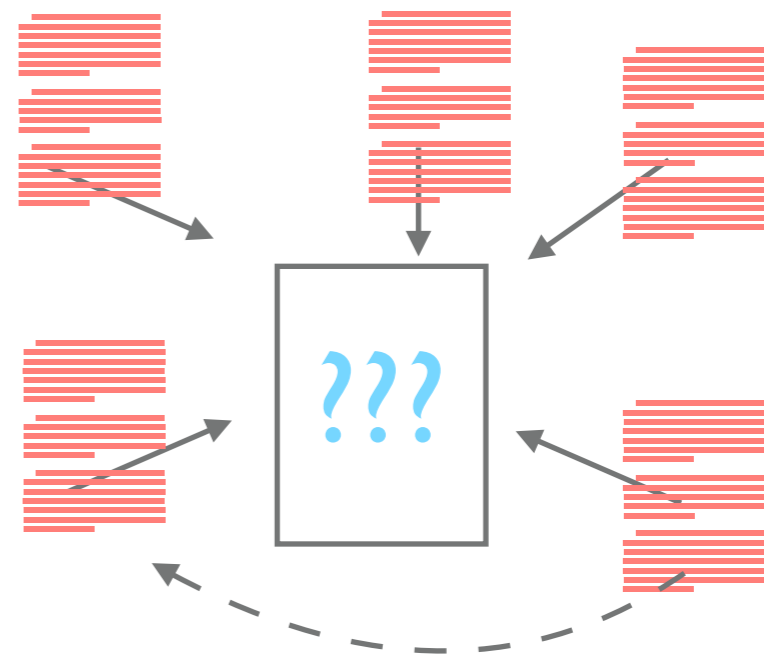
Need #1

Move past writer's block and generate new ideas for writing



Need #2

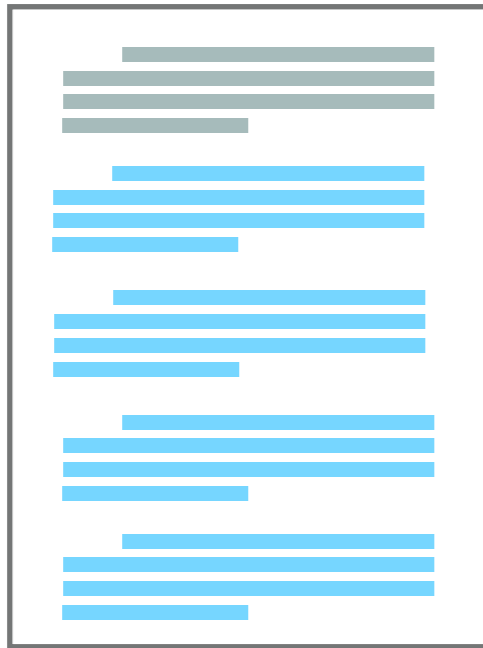
Integrate the writing styles of many different writers



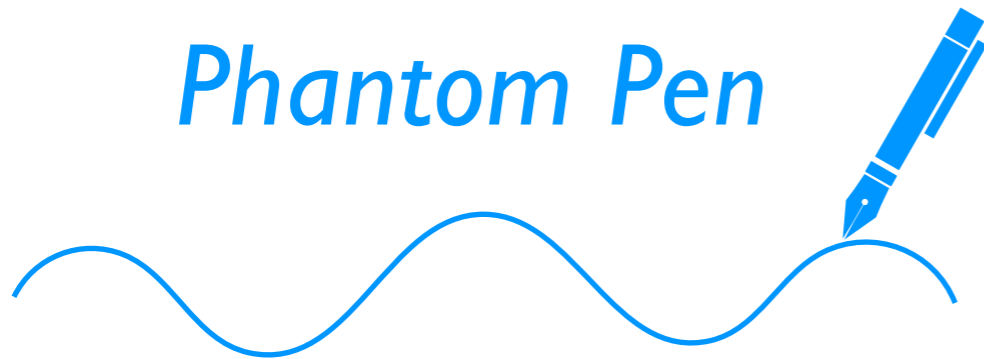
Helping with Writer's Block and Style Integration

Need #1

Move past writer's block and generate new ideas for writing

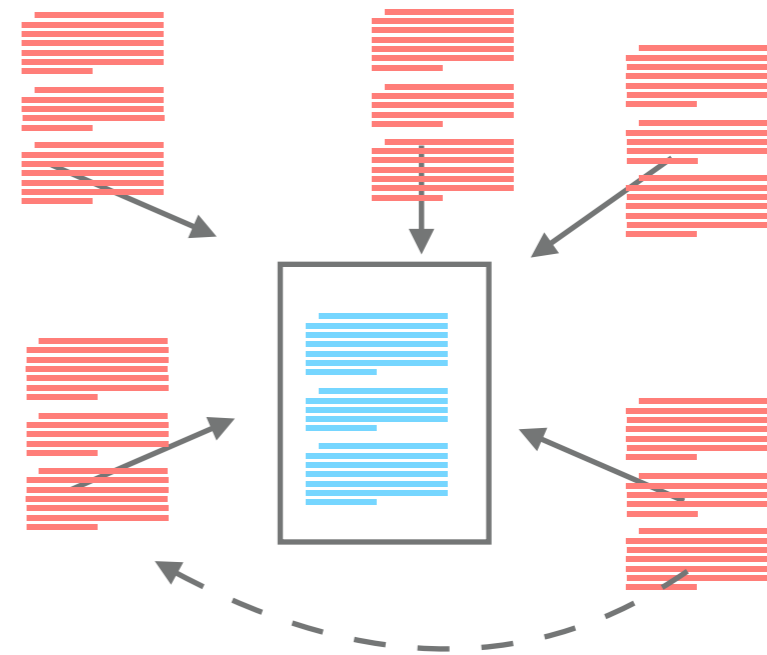


Phantom Pen



Need #2

Integrate the writing styles of many different writers



— Application that completes essay prompts

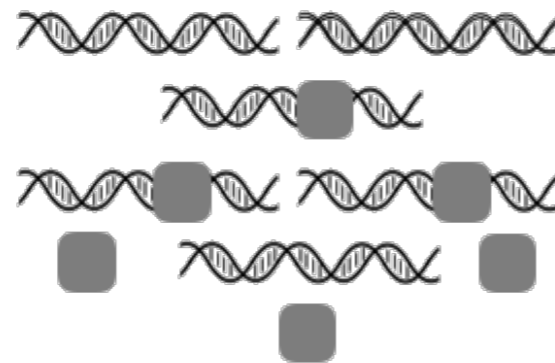
— using various corpora culled from a compilation of writers

Program for the User

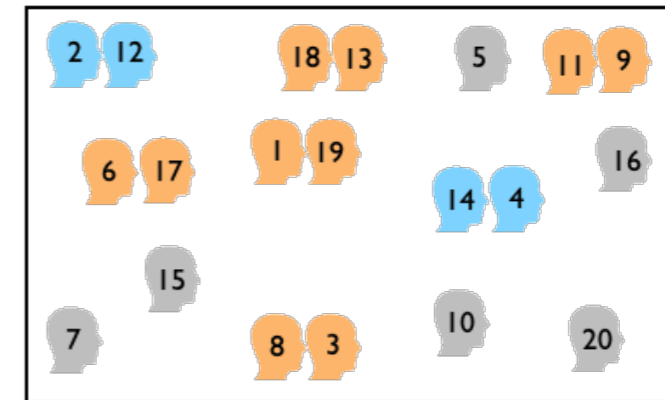
<http://phantompen.us/>

Mobolaji Williams

► PhD Physics, Statistical Physics (2019)



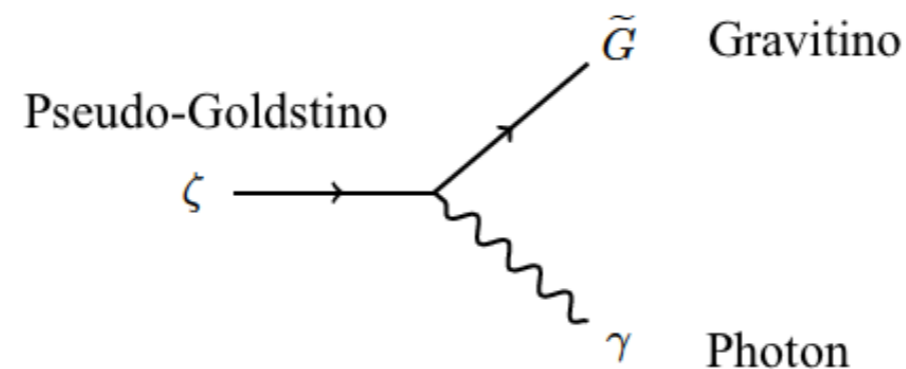
Protein-DNA System



Couples in a Dance Hall

Used combinatorics to build theoretical models of biomolecule interactions

► BSc Physics, Particle Physics (2013)



Explored models for light production from dark matter

End



Backup Slides

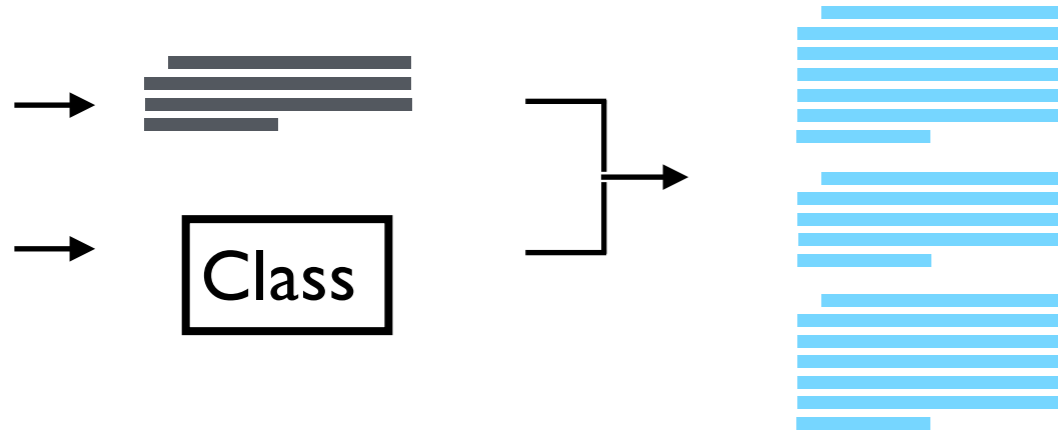


Functions of application

MAIN FUNCTION

SIMPLE GENERATE

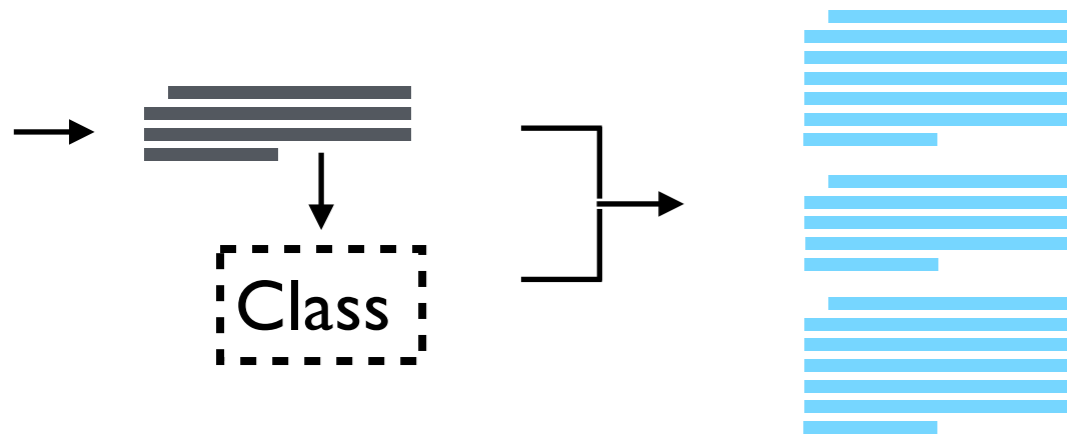
Generates an essay given a **prompt** and a chosen **class**



ADDITIONAL FUNCTION #1

CLASSIFY AND GENERATE

Generates an essay given a **prompt** and the class determined from the prompt

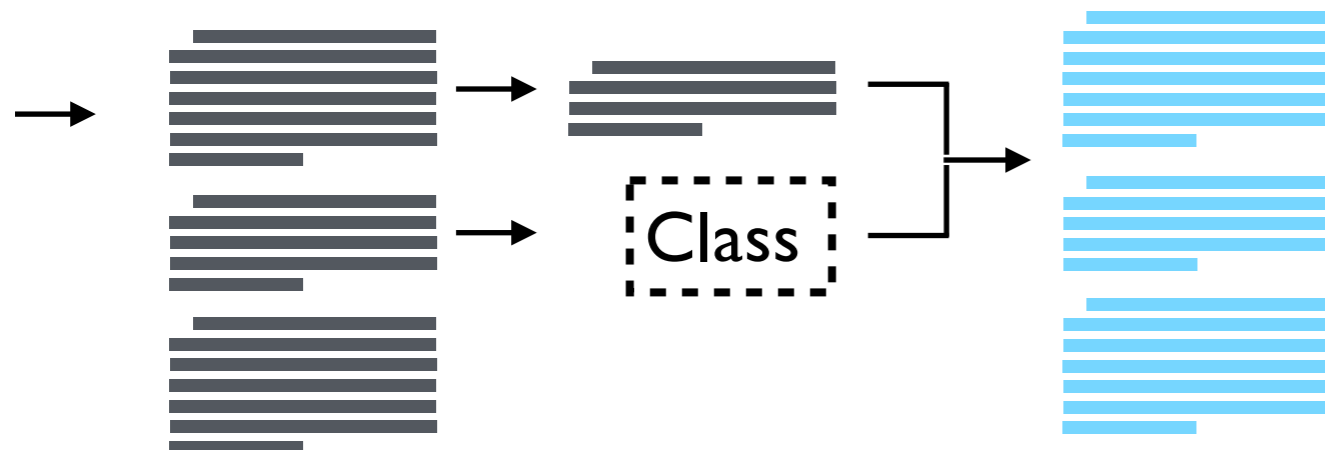


ADDITIONAL FUNCTION #2

CLASSIFY, EXTRACT, AND GENERATE

Generates an essay given a **previous essay**

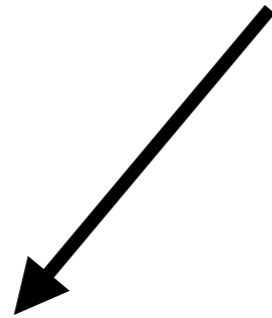
- Classifies and extracts the important sentences and generates essay from both



Most Essay Generators are Stupid

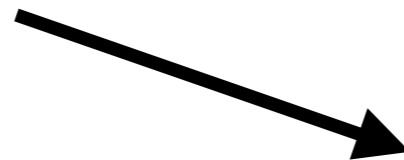
Markov Chain
Text Generator

“Memory less”



Recurrent
Neural Network

Exhibits “Short
Term” Memory



Generalized Pretrained
Transformer

Out performs
the RNN

