


Generative AI Hands On 1

 Details:

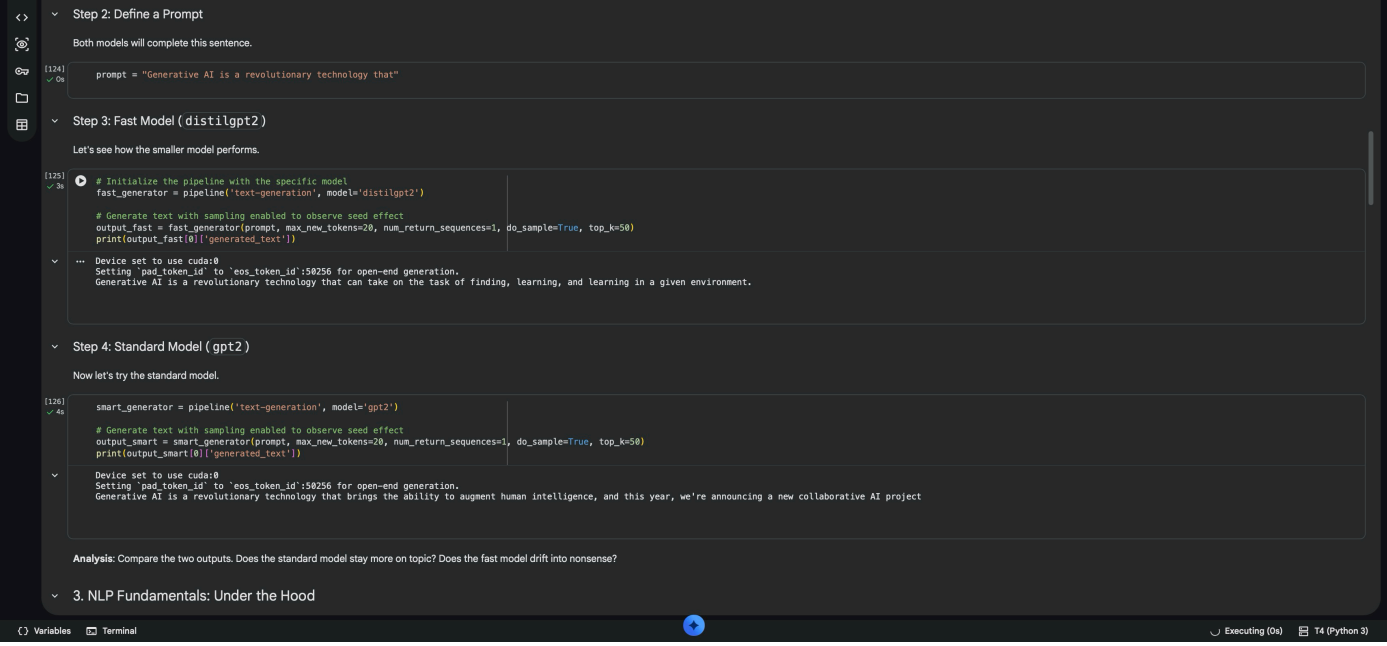
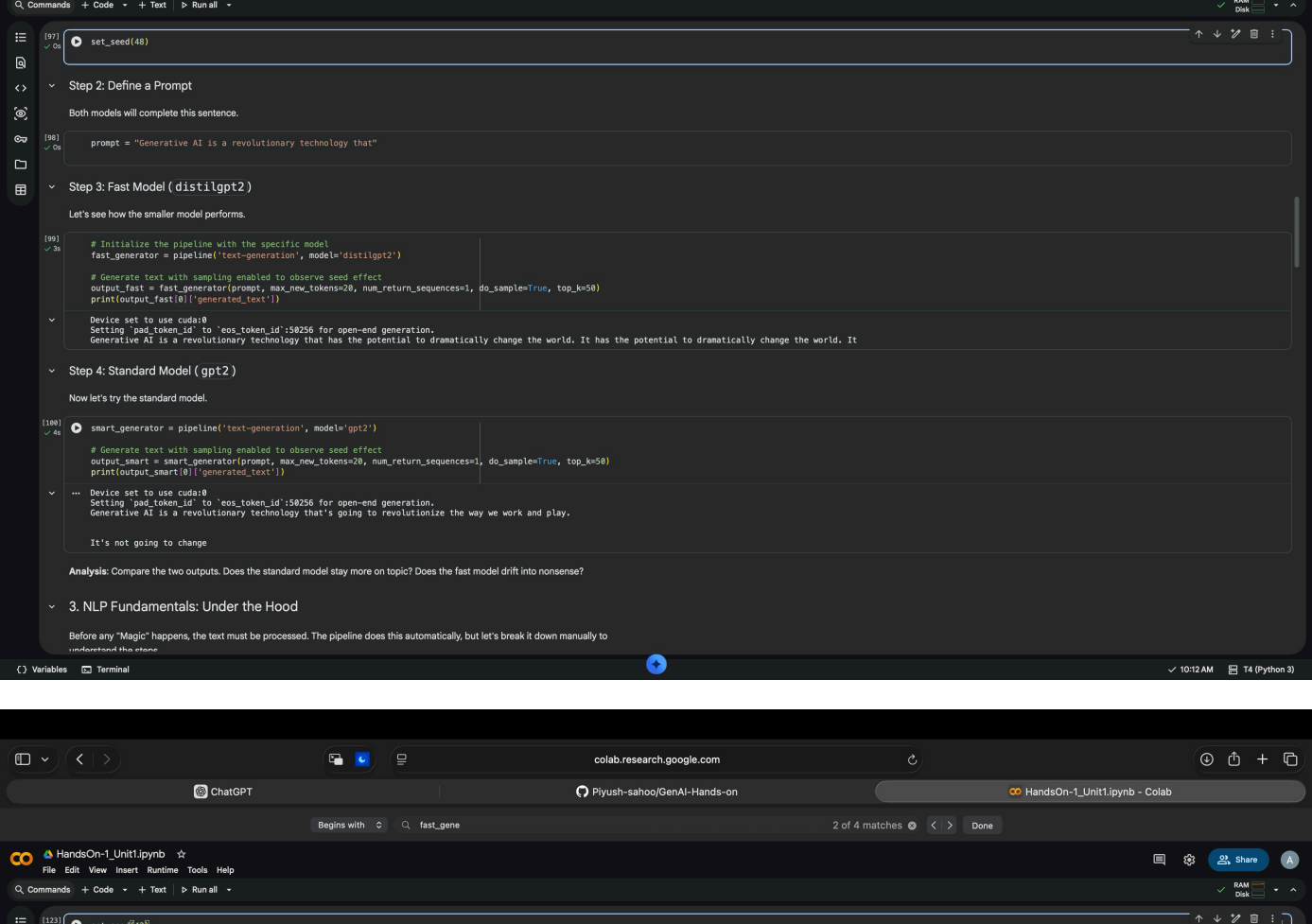
Name: Abhishrut Kaushik

Section: 6A

SRN: PES2UG23CS023

(SO HERE ARE THE OBSERVATIONS AND SCREENSHOTS FROM THE HANDSON_1.ipynb file. For other insights and observations I have included them in a new code file in the repo)

Changing the Seed values and getting different outputs:



Key Observations:

- Seed Value Impact : When we use the `set_seed()` function, running the code multiple times with the same seed produces the same random results
- Generative AI Quality : The 'smart' model (gpt2) generally produces more coherent, on-topic, and higherquality text compared to the fast model (distilgpt2). The smallerfaster model might tend to drift into less sensible or less relevant text.
- Tokenization : Text is broken down into smaller pieces (tokens), which are then converted into numerical id's that the model can process. Words are not always tokenized as full wordss and sometimes they are split into subword units, and special characters like 'Ġ' (for space) indicate word boundaries.
- POS Tagging : Each word in a sentence is assigned a Partofspeech tag, which helps understand the grammatical structure of the sentence.
- Named Entity Recognition (NER): Models can identify and categorize specific entities in text, such as organizations (ORG), miscellaneous entities (MISC like AI, LLMs, Transformer), and people (PER - though not explicitly shown in the output, it's a common NER category).
- Summarization Efficiency vs. Quality: Similar to text generation, faster summarization models (distilbart-cnn-12-6) might produce summaries quicker but potentially with less detail or fluency compared to larger, quality-optimized models (bart-large-cnn).
- Question Answering: Given a context and a questionthe model can extract precise answers directly from the provided text
- Masked Language Modeling : Models can predict missing words in a sentence based on the surrounding context, indicating their understanding of language patterns and semantics.

Difference Between Distilled and Normal Models (distilgpt2 vs gpt2):

gpt2:

- gpt2 is the standard version of the GPT-2 model. Its a powerful LLM trained on a big dataset to predict the next word in a sequence. It consists of many layers and parameters, allowing it to capturee complex language patterns and generate high quality, coherent textt. Pros are generally produces high quality, fluent, and contextually relevant text and lastly cons are that it requires significant computational resources is slower for inference due to its size

distilgpt2 :

- distilgpt2 is a distilled version of gpt2 and this model distillation is a technique where a smaller and simpler model is trained to mimic the behavior of a larger complex model. The student model learns to reproduce the outputs and internal representations of the teacher model, effectively transferring knowledge from the large model to the small one. Pros are that it is significantly smaller, faster, and requires less memory, making it more suitable for deployment on devices with limited resources or for applications requiring high inference speed and the cons are that while it retains much of the teacher model's performance, it typically sacrifices some degree of accuracy, coherence, or nuance compared to the full-sized model.

Now we enlist some pos tags commonly usedd:

Parts of Speech (Basics)

1. Noun

Names a **person, place, thing, or idea**

2. Pronoun

Replaces a noun

3. Verb

Shows action or state of being

4. Adjective

Describes a noun

5. Adverb

Describes a verb, adjective, or another adverb

6. Preposition

Shows relationship (position, time, direction)

7. Conjunction

Joins words or sentences

8. Interjection

Expresses emotion

9. Determiner / Article

Limits or defines a noun

2) What are NNS, VBD, NNP? (basically all of the POS Tags)

Common POS Tags Explained

Nouns

Tag	Meaning	Example
NN	Noun, singular	<i>dog, car</i>
NNS	Noun, plural	<i>dogs, cars</i>
NNP	Proper noun, singular	<i>India, Rahul</i>
NNPS	Proper noun, plural	<i>Americans</i>

- NNS** means *plural noun*
- NNP** means *proper noun (name)*

Verbs

Tag	Meaning	Example
VB	Base form	<i>run</i>
VBD	Past tense	<i>ran</i>
VBG	Gerund / present participle	<i>running</i>
VBN	Past participle	<i>eaten</i>
VBP	Present tense (non-3rd person)	<i>run</i>
VBZ	Present tense (3rd person)	<i>runs</i>

- VBD** means *verb, past tense*

Adjectives and Adverbs

Tag	Meaning	Example
JJ	Adjective	<i>big</i>
JJR	Comparative adjective	<i>bigger</i>
JJS	Superlative adjective	<i>biggest</i>
RB	Adverb	<i>quickly</i>
RBR	Comparative adverb	<i>faster</i>
RBS	Superlative adverb	<i>fastest</i>

Others

Tag	Meaning	Example
PRP	Personal pronoun	<i>he, she</i>
DT	Determiner	<i>the, a</i>
IN	Preposition	<i>in, on</i>
CC	Coordinating conjunction	<i>and, but</i>
UH	Interjection	<i>wow!</i>

3) Example (Sentence Breakdown)

Sentence:

Rahul bought two books yesterday.

Word	POS Tag	Meaning
Rahul	NNP	Proper noun
bought	VBD	Past tense verb
two	CD	Number
books	NNS	Plural noun
yesterday	RB	Adverb

Also POS tags are used in parsers, syntax validators, and NLP, they help with:

- Tokenization
- Building syntax trees
- Grammar checking
- Compilers and language processing projects