

Generative AI Hands On 1

① Details:

Name: Abhishek 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:

The screenshot shows two separate runs of the same Python code in a Jupyter Notebook. Both runs start with the same prompt: "Generate AI is a revolutionary technology that". The first run uses a seed of 42, and the output is: "Generate AI is a revolutionary technology that has the potential to dramatically change the world. It has the potential to dramatically change the world. It". The second run uses a seed of 42, and the output is: "Generate AI is a revolutionary technology that's going to revolutionize the way we work and play. It's not going to change". This demonstrates that changing the seed value can lead to different outputs.

The screenshot shows two separate runs of the same Python code in a Jupyter Notebook. Both runs start with the same prompt: "Generate AI is a revolutionary technology that". The first run uses a seed of 42, and the output is: "Generate AI is a revolutionary technology that can take on the task of finding, learning, and learning in a given environment.". The second run uses a seed of 42, and the output is: "Generate AI is a revolutionary technology that brings the ability to augment human intelligence, and this year, we're announcing a new collaborative AI project". This demonstrates that changing the seed value can lead to different outputs.

Key Observations:

1. Seed Value Impact : When we use the `set_seed()` function, running the code multiple times with the same seed produces the same random results
2. Generative AI Quality : The 'smart' model (gpt2) generally produces more coherent, on-topic, and higher quality text compared to the fast model (distilgpt2). The smaller/faster model might tend to drift into less sensible or less relevant text.
3. 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 words and sometimes they are split into subword units, and special characters like 'Ġ' (for space) indicate word boundaries.
4. POS Tagging : Each word in a sentence is assigned a Part of speech tag, which helps understand the grammatical structure of the sentence.
5. 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).
6. 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).
7. Question Answering: Given a context and a question, the model can extract precise answers directly from the provided text
8. 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. It's 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 capture complex language patterns and generate high-quality, coherent text. Pros are that it generally produces high-quality, fluent, and contextually relevant text, but the cons are that it requires significant computational resources and 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 used:

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

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

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

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

- **VBD** means *verb, past tense*

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

- **JJ** means *adjective*
- **JJR** means *comparative adjective*
- **JJS** means *superlative adjective*
- **RB** means *adverb*
- **RBR** means *comparative adverb*
- **RBS** means *superlative adverb*

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

- **PRP** means *personal pronoun*
- **DT** means *determiner*
- **IN** means *preposition*
- **CC** means *coordinating conjunction*
- **UH** means *interjection*

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

Common POS Tags Explained

Nouns

- Names a person, place, thing, or idea

Verbs

- Shows action or state of being

Adjectives

- Describes a noun

Adverbs

- Describes a verb, adjective, or another adverb

Preposition

- Shows relationship (position, time, direction)

Conjunction

- Joins words or sentences

Interjection

- Expresses emotion

Determiner / Article

- Limits or defines a noun

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

- **NN** means *singular noun*
- **NNS** means *plural noun*
- **NNP** means *proper noun (name)*
- **NNPS** means *proper noun (name)*

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

- **VB** means *base form*
- **VBD** means *past tense*
- **VBG** means *gerund / present participle*
- **VBN** means *past participle*
- **VBP** means *present tense (non-3rd person)*
- **VBZ** means *present tense (3rd person)*

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

- **JJ** means *adjective*
- **JJR** means *comparative adjective*
- **JJS** means *superlative adjective*
- **RB** means *adverb*
- **RBR** means *comparative adverb*
- **RBS** means *superlative adverb*

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

- **PRP** means *personal pronoun*
- **DT** means *determiner*
- **IN** means *preposition*
- **CC** means *coordinating conjunction*
- **UH** means *interjection*

3) Example (Sentence Breakdown)

Sentence:

Rahul bought two books yesterday.

Word	POS Tag	Meaning
------	---------	---------