Uni-Model: Text to Text, Image to Image

Multi-Model: Text, Image

CNN: GAN

NLP: summarization, classification, Translation, generation, etc.

Language Modeling: RNN, LSTM, GRU

Prompt: zero-shot, few-shot, change of taugh (COT)

Video generation model: OpenAI SORA, NUWA (Microsoft), CogVideo, Runway Gen-2, Text2Video-zero

Stemming:

Definition: Stemming is the process of reducing a word to its root form by removing suffixes or prefixes, often without considering the word’s meaning.

Key Characteristics:

* Usually rule-based or algorithms
* The output may not always be a valid word.
* Focuses on quick and computationally efficient reduction of words.

Example:

Words like **running**, **runner**, and **ran** might be reduced to the stem **run**.

Similarly, **better** might be stemmed to **bett.**

Common stemming algorithms:

* Porter Stemmer
* Lancaster Stemmer
* Snowball Stemmer

Advantages:

* Simple and fast.
* Useful in scenarios where slight inaccuracies are acceptable, like search engines.

Disadvantages:

* Often produces stems that aren’t actual words (eg: studies 🡪 studi).

Lemmatization:

Definition: Lemmatization reduces a word to its base form (lemma) while considering the context and the word’s.

Key Characteristics:

* More sophisticated and linguistically informed.
* The output is always a valid word.
* Requires tools like a dictionary or vocabulary resource to find the base form.

Example:

* Running 🡪 run
* Better 🡪 good (context – aware, since “better” is the comparative form of “good”.

Common Lemmatization Tools:

* WordNet Lemmatizer in NLTK
* Spacy (built-in lemmatizer)

Disadvantage:

* Slower and computationally more expensive than stemming.
* Requires additional information, such as part-of-speech tags.

**Vocabulary:** A collection of unique words or distinct words.