for more accurate
representation of original LLM terms cont:
↓ words

~ Lemmatization: a text normalization technique
in NLP where we reduce words to their dictionary
form known as Lemma. insted of chopping suffix/prefix
it considers the context of the whole word and applies
Morphological analysis to find the base dictionary form
Ex: running → run, ate → eat, children → child, worst → bad

- Stemming: a text normalization technique in NLP to
reduce words to base or root form known as "stem" this
is done by removing suffixes like ("ing", "-ed", "s")
from words the goal is to treat words with similar meaning
as the same (for efficiency in NLP tasks)
Ex: running → runn, happily → happi, cars → car

Semi Supervised learning: a type of ML where training data
has both Labeled and unlabeled data. The goal is to leverage
the information from the unlabeled data to improve the
performance of a model that would otherwise be trained
solely on labeled data. Good when labeld data is hard to find but
                                        unlabeled data is redily avalible

Self supervised learning: a type of ML where the model learns
from unlabeled data by creating its own supervisory signals this is
achived by masking parts of the input data, and training the
model to predict the masked portion based on the remaning data
in essence the data itself provides the labels allowing
model to learn without human annotation for d ch.