Some applications of DL

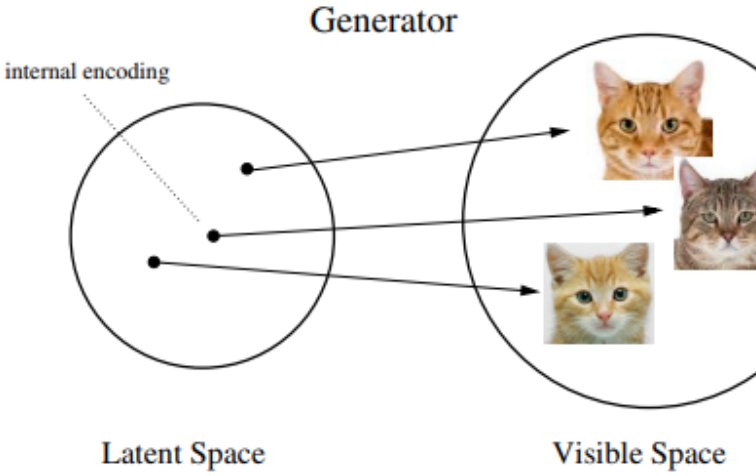
NLP

Key Technologies

- **tokenization**: splitting the input sentence into relevant lexical components (characters/words/subwords), and coding them into numbers. Byte-Pair Encoding, WordPiece, SentencePiece, etc. see this tokenizer summary
- **transformers**: a feed-forward deep learning model adopting self-attention for weighting the mutual significance of tokens in the sentence
- **word embeddings** a semantic embedding of words, mostly used for text similarity, text retrieval, code search, etc. Examples are Word2Vec, Glove. Transformers do not use them: they learn their own embeddigs. see this blog for a comparison of state-of-the-art text embeddings

Generative modeling

Train a *generator* able to sample original data similar to those in the training set,



implicitly learning the *distribution of data*.
- the randomicity of the generator is provided by a *random seed* (noise) received as input. -