Applying Embedding Approaches to Chatbot Application
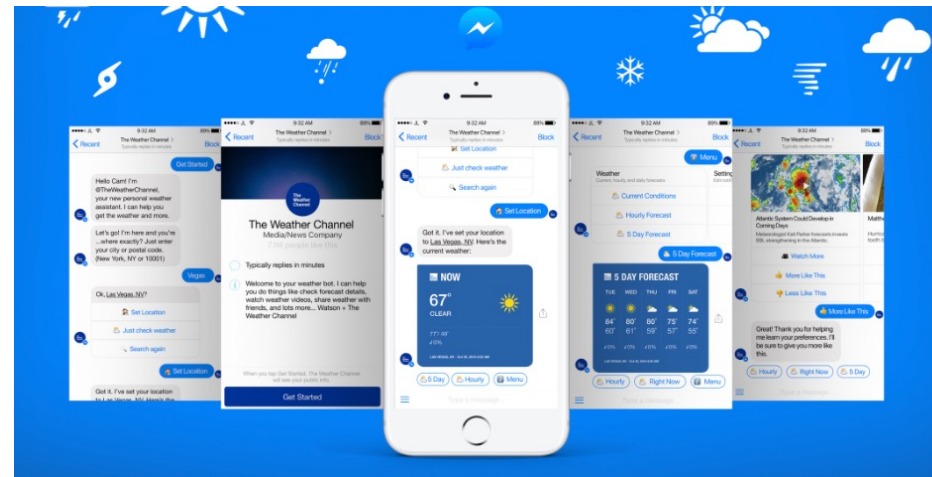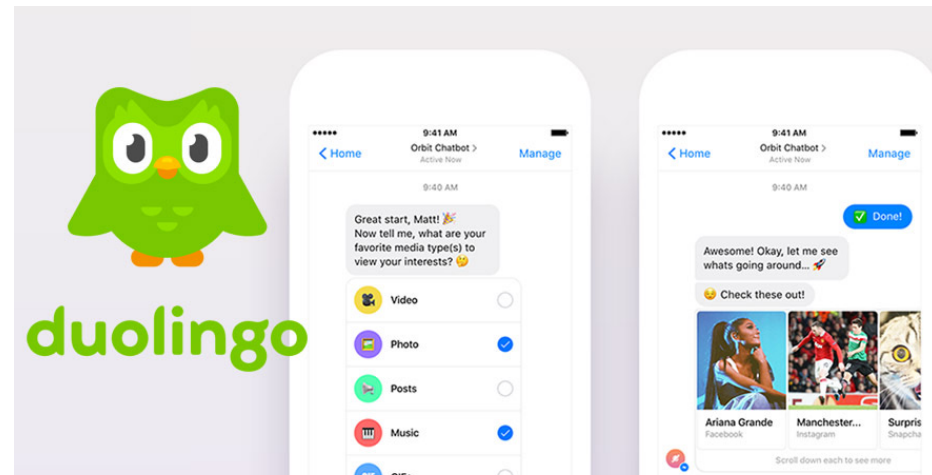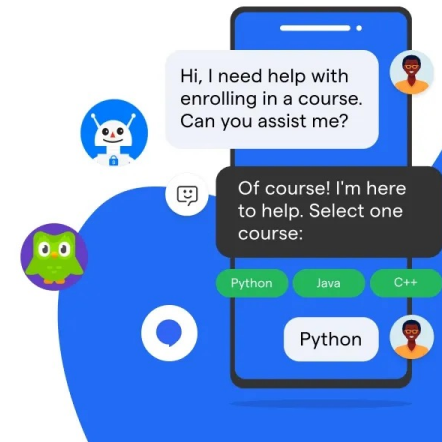
Presenter: Vision(Xin) Wang

Email: xwang314@Binghamton.edu

Chatbot Application

- A chatbot is a **computer program** designed to mimic conversations with actual users, especially online.

- Chatbots can organically and intuitively **communicate with people using text, voice, or even video**.

Chatbots For Education

Embedding

Embeddings are often used to **capture the semantic or contextual relationships** between objects, making it easier for algorithms to process and analyze them.

- Word embeddings: numerical representations of words in a continuous vector space.

Techniques: Word2Vec, GloVe, FastText…

- Sentence embeddings: numerical representations of entire sentences.

Techniques: Sentence Transformer, BERT…

- Image embeddings: numerical representation of images

Techniques: CNNs, ViT, CLIP…

- Knowledge graph embeddings: representation of entities and relationships in a knowledge graph, such as concepts in a domain and their semantic connections.

Techniques: TransE, DistMult…
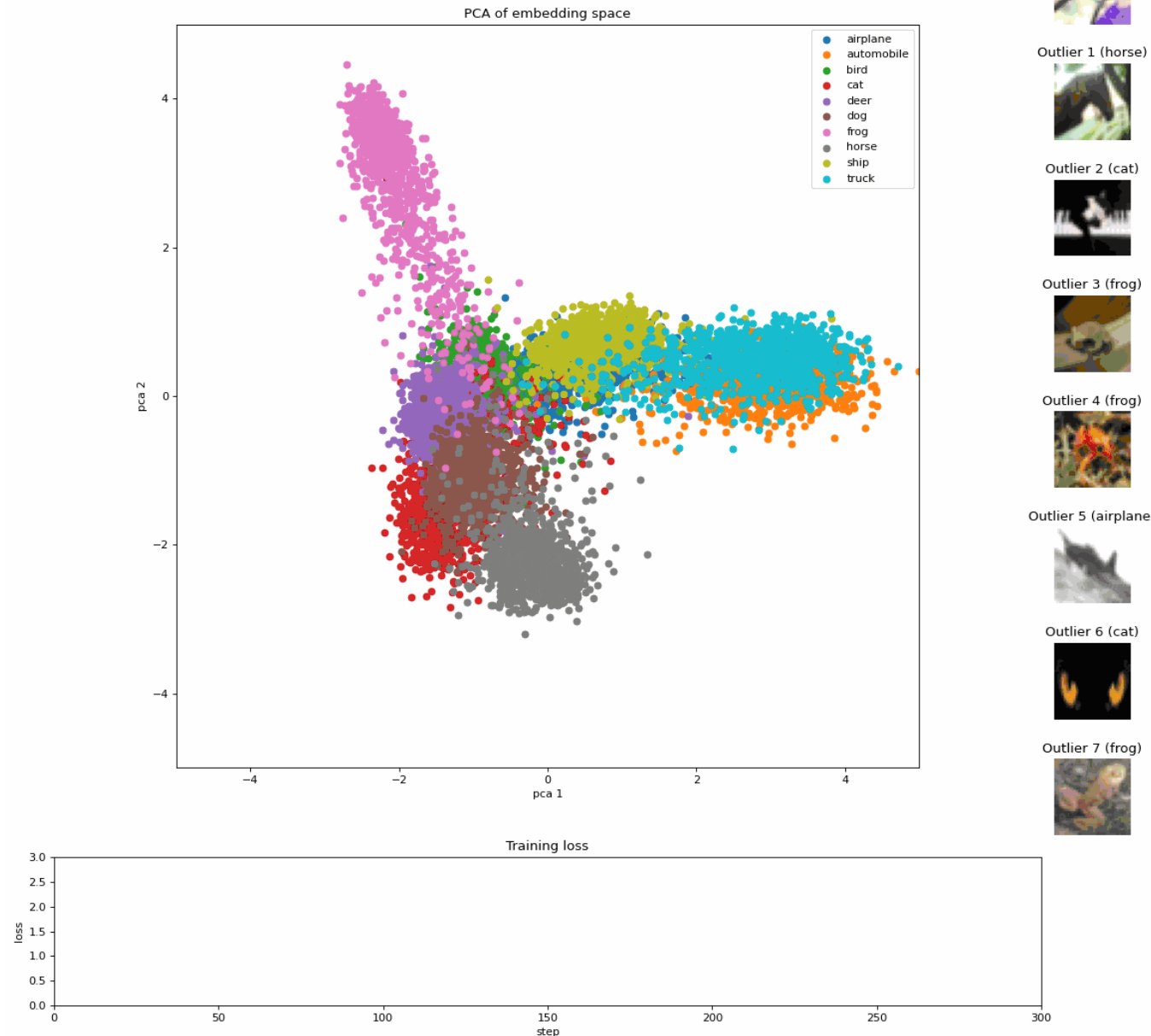
Changes of Embeddings during Fine-Tuning of Vision Transformers (ViT)
Source: https://huggingface.co/blog/MarkusStoll/embeddings-during-fine-tuning-of-vision-transform

Word2vec model

CBOW

- Word2vec is based on a shallow, two-layer neural network, which takes as input a corpus of texts and produces as the result a vector for each word that is represented in the corpus.

- CBOW (continuous bag of words): to predict the target word based on patterns in the input data

- Skip-gram: to predict the surrounding words from the target word.

Skip-gram
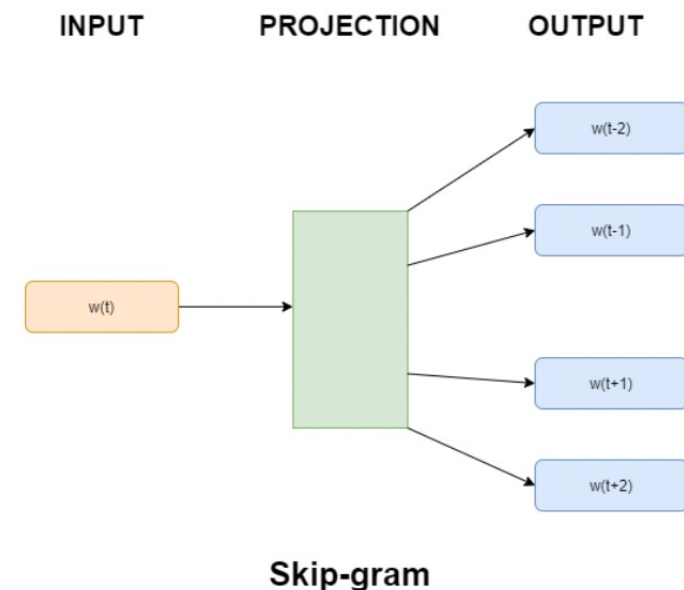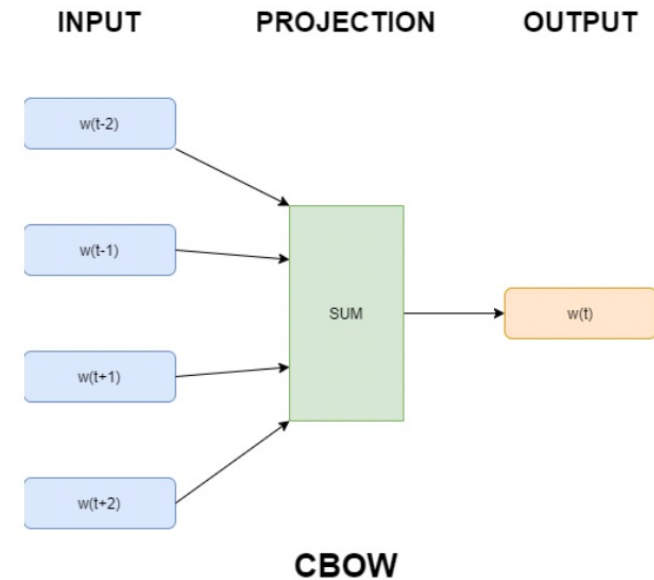
Vector Search

Vector search is an approach in information retrieval that stores numeric representations of content for search scenarios. Because the content is numeric rather than plain text, the search engine matches on vectors that are the most similar to the query, with no requirement for matching on exact terms.
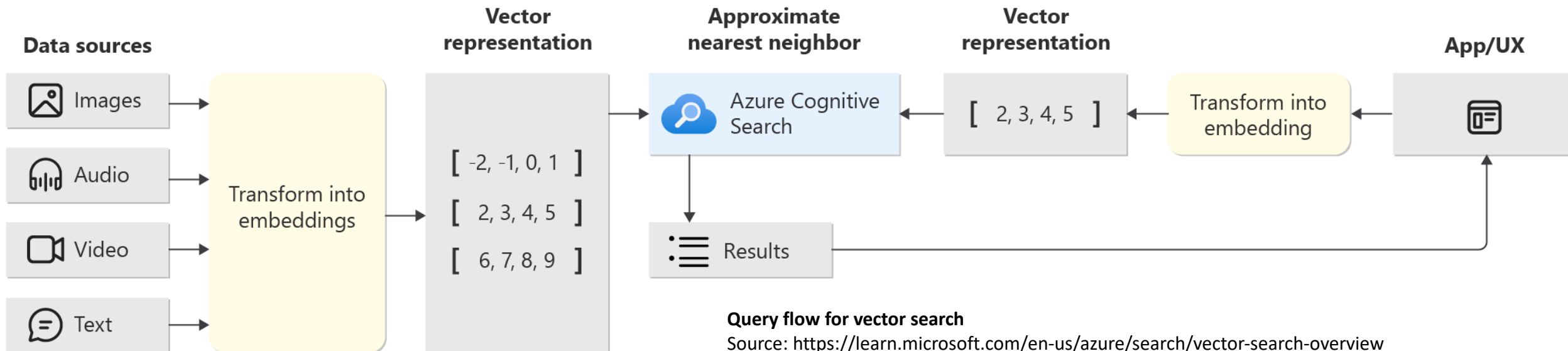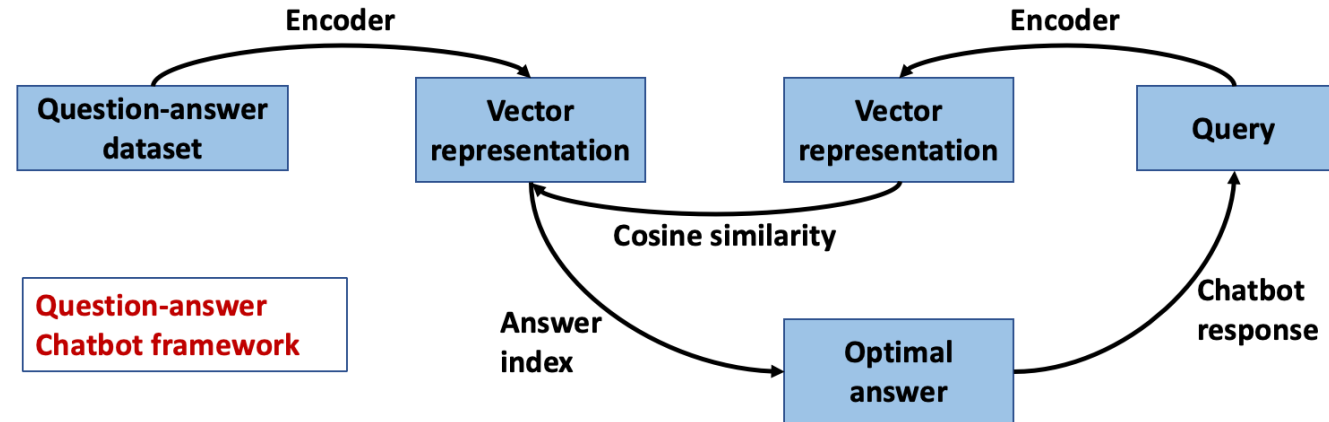
Encoder

Encoder

| Question-answer dataset | → | Vector representation | | Vector representation | → | Query |

Question-answer Chatbot framework

Cosine similarity

Answer index

Optimal answer

Chatbot response

Data sources

Vector representation

Approximate nearest neighbor

Vector representation

App/UX

Images

Audio

Video

Text

Transform into embeddings

[-2, -1, 0, 1]

[2, 3, 4, 5]

[6, 7, 8, 9]

Azure Cognitive Search

[2, 3, 4, 5]

Transform into embedding

Results

Query flow for vector search
Source: https://learn.microsoft.com/en-us/azure/search/vector-search-overview

CLIP embedding

(1) Contrastive pre-training

Pepper the aussie pup → Text Encoder → T_1, T_2, T_3, ..., T_N

Image Encoder → I_1, I_2, I_3, ⋮, I_N

	T_1	T_2	T_3	...	T_N
I_1	$I_1{\cdot}T_1$	$I_1{\cdot}T_2$	$I_1{\cdot}T_3$	...	$I_1{\cdot}T_N$
I_2	$I_2{\cdot}T_1$	$I_2{\cdot}T_2$	$I_2{\cdot}T_3$	...	$I_2{\cdot}T_N$
I_3	$I_3{\cdot}T_1$	$I_3{\cdot}T_2$	$I_3{\cdot}T_3$	...	$I_3{\cdot}T_N$
⋮	⋮	⋮	⋮	⋱	⋮
I_N	$I_N{\cdot}T_1$	$I_N{\cdot}T_2$	$I_N{\cdot}T_3$	...	$I_N{\cdot}T_N$
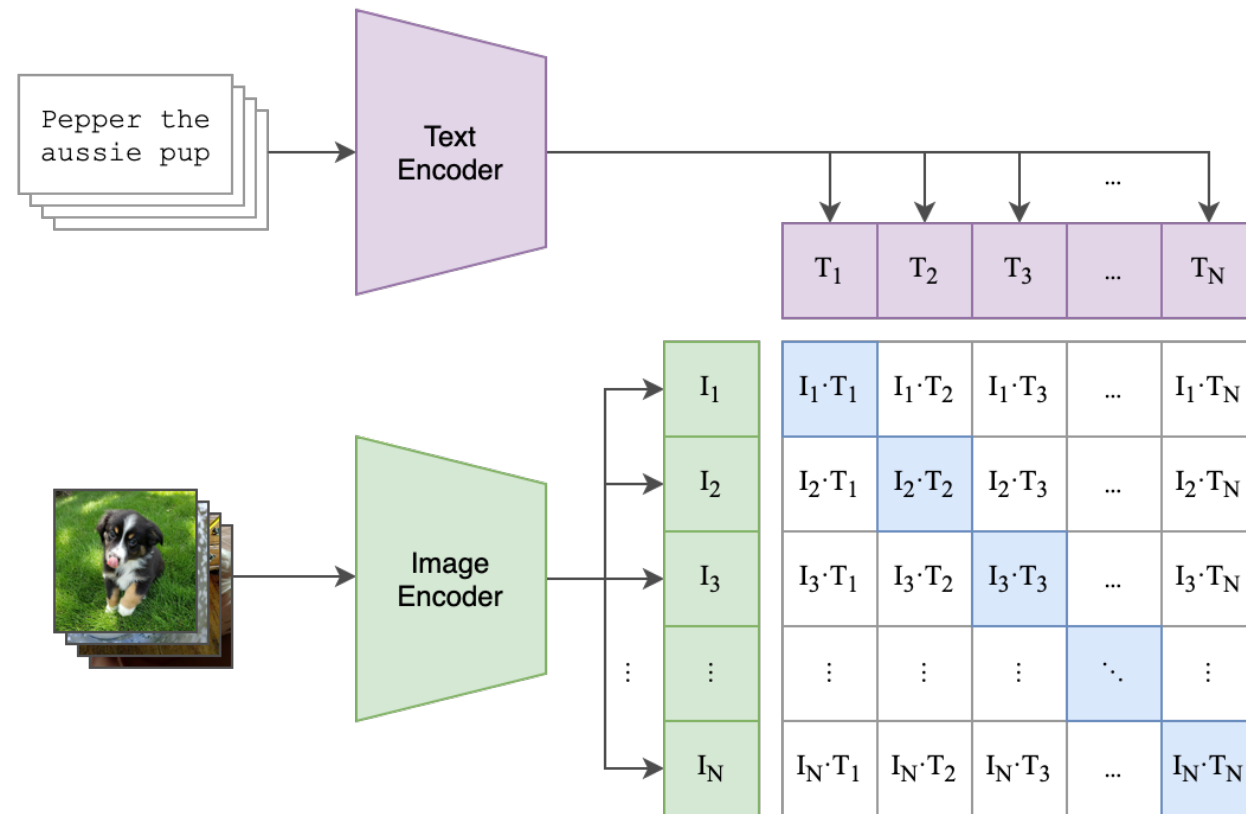
(2) Create dataset classifier from label text

plane, car, dog, ⋮, bird → A photo of a {object}. → Text Encoder → T_1, T_2, T_3, ..., T_N

(3) Use for zero-shot prediction

Image Encoder → I_1

$I_1{\cdot}T_1$	$I_1{\cdot}T_2$	$I_1{\cdot}T_3$	...	$I_1{\cdot}T_N$

A photo of a dog.