WIM – What'd I Miss?

Victor Geislinger

Finding Relevant Info in a Video is Hard

- Videos are informative
- Playlists help contain info
- If organized well, can find relevant info

Finding Relevant Info in a Video is Hard

- Videos are informative
- Playlists help contain info
- If organized well, can find relevant info

- Titles can only give so much info
- Lots of info to sift through
- No real way to search for related terms/ideas

Solution: AI

- Can get transcripts (**Whisper**)

Solution: AI

- Can get transcripts (**Whisper**)
- Can search through related quotes(**BERT/Encoders via transformers**)

Solution: AI

- Can get transcripts (**Whisper**)
- Can search through related quotes(**BERT/Encoders via transformers**)
- Can summarize information with generated text (**Anthropic’s Claude**)

Solution: AI

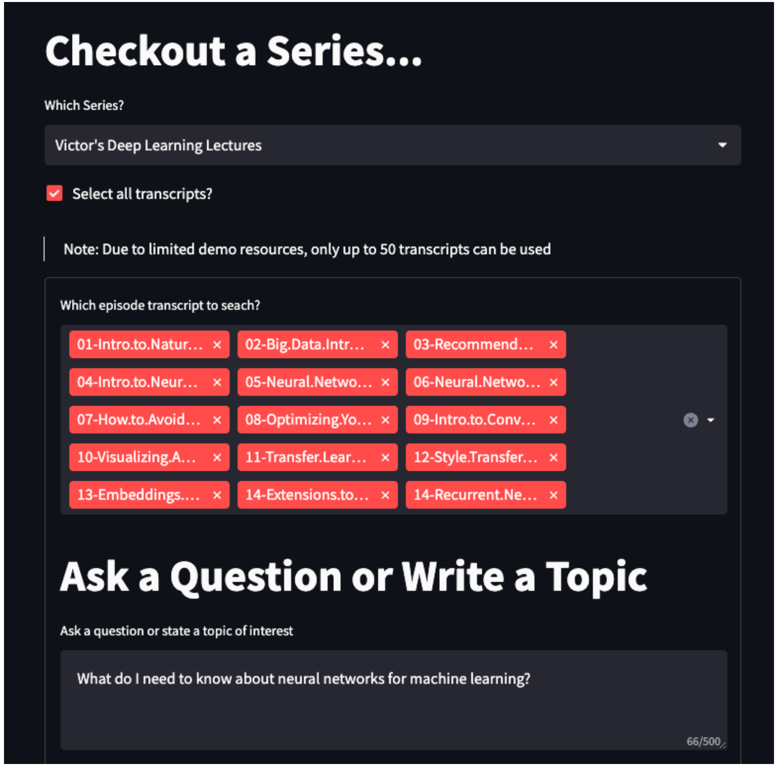
WIM

WIM

Ask pointed questions about a given playlist and get back a summary, key points, and related timestamps generated via AI! 🤖

WIM

Ask pointed questions



Checkout a Series...

Which Series?

Victor's Deep Learning Lectures

☑ Select all transcripts?

Note: Due to limited demo resources, only up to 50 transcripts can be used

Which episode transcript to seach?

01-Intro.to.Natur... ×
02-Big.Data.Intr... ×
03-Recommend... ×
04-Intro.to.Neur... ×
05-Neural.Netwo... ×
06-Neural.Netwo... ×
07-How.to.Avoid... ×
08-Optimizing.Yo... ×
09-Intro.to.Conv... ×
10-Visualizing.A... ×
11-Transfer.Lear... ×
12-Style.Transfer... ×
13-Embeddings.... ×
14-Extensions.to... ×
14-Recurrent.Ne... ×

Ask a Question or Write a Topic

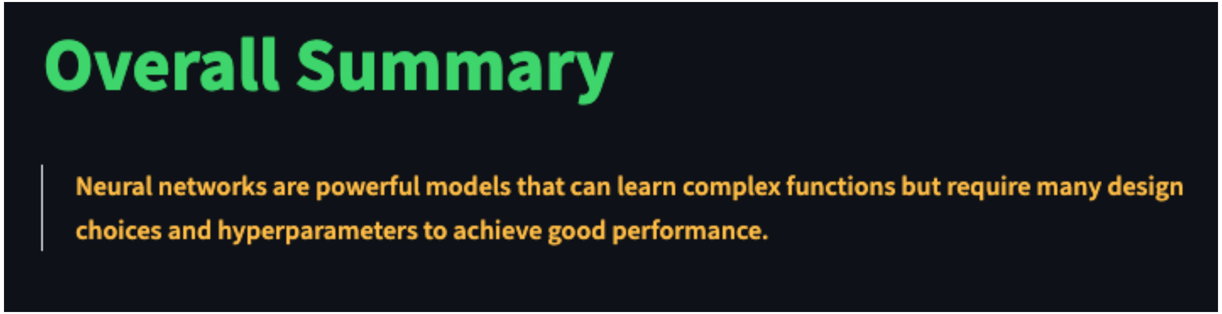
Ask a question or state a topic of interest

What do I need to know about neural networks for machine learning?

66/500

WIM

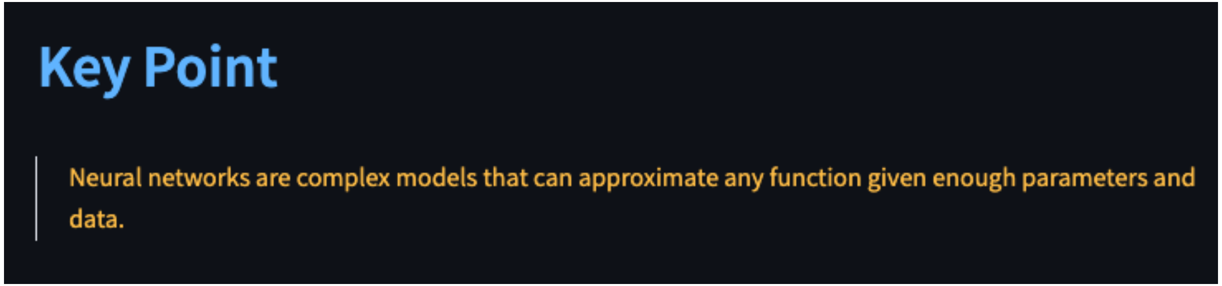
summary



Overall Summary

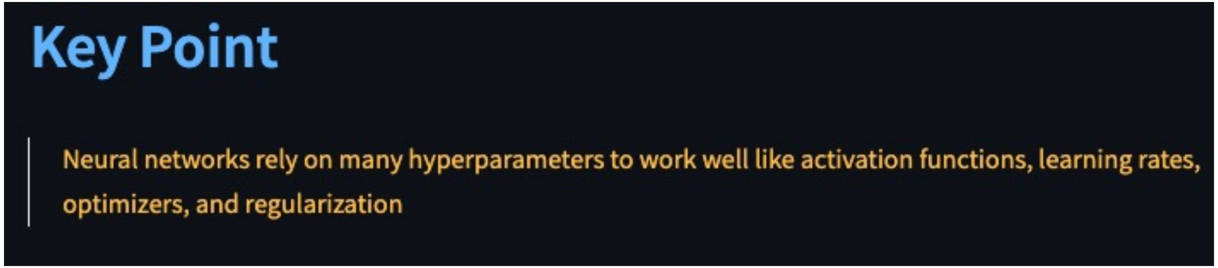
> Neural networks are powerful models that can learn complex functions but require many design choices and hyperparameters to achieve good performance.

key points



Key Point

> Neural networks are complex models that can approximate any function given enough parameters and data.

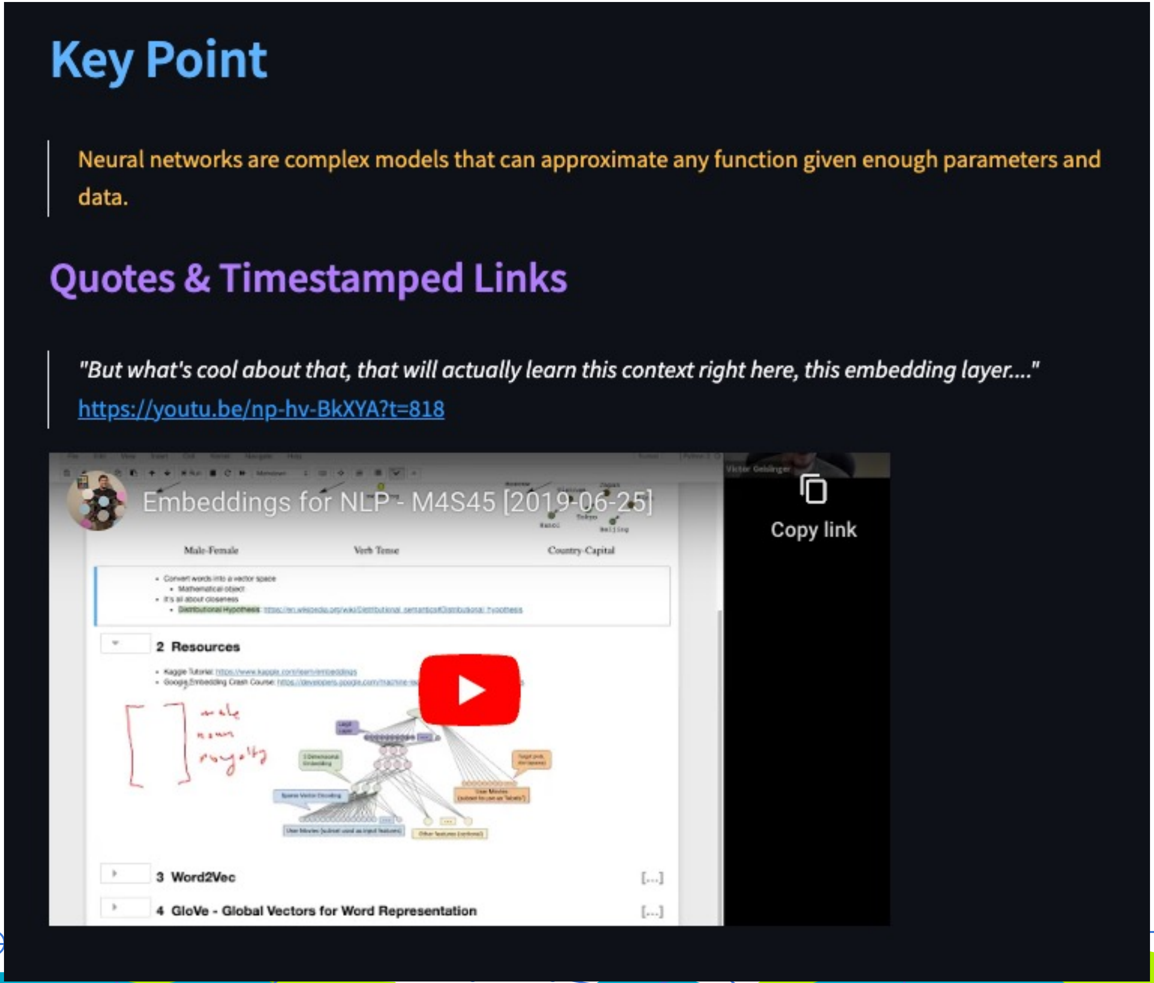


Key Point

> Neural networks rely on many hyperparameters to work well like activation functions, learning rates, optimizers, and regularization

WIM

timestamps



Key Point

> Neural networks are complex models that can approximate any function given enough parameters and data.

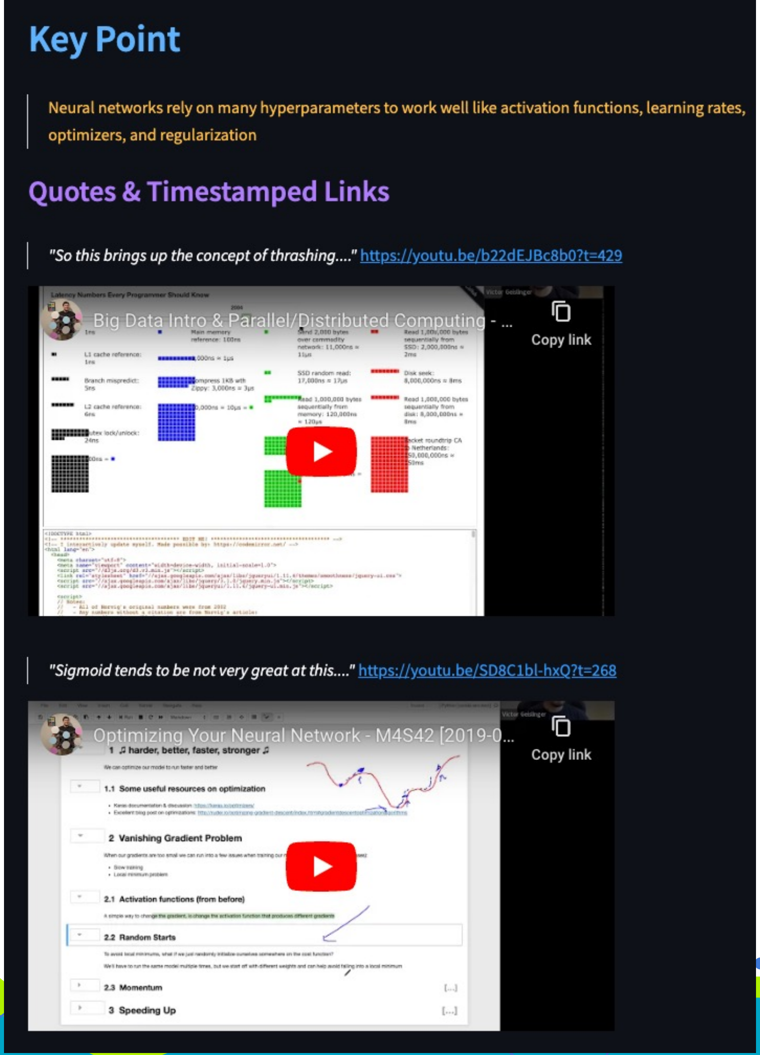
Quotes & Timestamped Links

> *"But what's cool about that, that will actually learn this context right here, this embedding layer...."*
> https://youtu.be/np-hv-BkXYA?t=818

Embeddings for NLP - M4S45 [2019-06-25]
Copy link
Male-Female
Verb Tense
Country-Capital
Convert words into a vector space
Mathematical object
It's all about closeness
Distributional Hypothesis
2 Resources
Kaggle Tutorial
Google Embedding Crash Course
male
noun
royalty
3 Word2Vec
4 GloVe - Global Vectors for Word Representation
[...]
[...]

WIM

timestamps



Key Point

> Neural networks rely on many hyperparameters to work well like activation functions, learning rates, optimizers, and regularization

Quotes & Timestamped Links

> *"So this brings up the concept of thrashing...."* https://youtu.be/b22dEJBc8b0?t=429

Big Data Intro & Parallel/Distributed Computing - ...
Copy link
Latency Numbers Every Programmer Should Know

> *"Sigmoid tends to be not very great at this...."* https://youtu.be/SD8C1bl-hxQ?t=268

Optimizing Your Neural Network - M4S42 [2019-0...
Copy link
1 ♫ harder, better, faster, stronger ♫
1.1 Some useful resources on optimization
2 Vanishing Gradient Problem
2.1 Activation functions (from before)
2.2 Random Starts
2.3 Momentum
3 Speeding Up

WIM

Ask pointed questions about a given playlist and get back a summary, key points, and related timestamps generated via AI! 🤖

How Does it Work?

- Transcripts generated via **Whisper**
  - https://github.com/MrGeislinger/whisper-extract
  - Technically can be created with any other tool

How Does it Work?

- Transcripts’ sentences compared via embeddings (**BERT**)
  - BERT or other encoding transformer
  - Selects a subset of sentences from given transcripts

How Does it Work?

- Subset of transcripts’ & user’s question fed to AI (**Anthropic’s Claude**)
  - Generated summaries and key points
  - Model chooses relevant quotes
  - Quotes cross-checked with subset to provide links w/timestamps

The Future?

- Deployed with more resources
  - Currently Streamlit deployment limits RAM to 1GB
  - Allows more comparisons (more transcripts)

The Future?

- Deployed with more resources
  - Currently Streamlit deployment limits RAM to 1GB
  - Allows more comparisons (more transcripts)
- Use vector database (such as Chroma)
  - Currently local
  - Vector database will speed up comparisons on a database server

The Future?

- Deployed with more resources
  - Currently Streamlit deployment limits RAM to 1GB
  - Allows more comparisons (more transcripts)
- Use vector database (such as Chroma)
  - Currently local
  - Vector database will speed up comparisons on a database server
- Sentence embeddings treated differently between question and transcripts
  - Fine-tuning of model to create embeddings

The Future?

- Deployed with more resources
  - Currently Streamlit deployment limits RAM to 1GB
  - Allows more comparisons (more transcripts)
- Use vector database (such as Chroma)
  - Currently local
  - Vector database will speed up comparisons on a database server
- Sentence embeddings treated differently between question and transcripts
  - Fine-tuning of model to create embeddings
- Adjust prompting (always different for different LLMs!)

Demo Time

WIM

wim.victorsothervector.com