

Empower Your AI Journey

Hands-on Machine Learning with fastai for Graphics and NLP

David vonThenen

Software Engineer/Developer Advocate

   [@dvonthenen](https://github.com/dvonthenen)

David vonThenen

- Are you Human or AI?
- I want 5 Kubernetes
- Virtual Machines are Real
- Replacing Myself with Bots...
- Cloudy, cloudy, cloudy...
- There is storage for that!



[@dvonthenen](https://twitter.com/dvonthenen)



Agenda

- What is fastai?
- Image Classification
- Natural Language Processing
 - Dataset Considerations
- Resources
- Q&A

What is fastai?

Accelerate Model Creation

fastai: Deep Learning Library for All

fastai - <https://github.com/fastai/fastai>

- Deep Learning Abstraction on top of PyTorch
- Top-Down Approach to Learning AI/ML
- Fast and Easy Neural Network Training
 - ML Model is Your Output
- Usability Without Sacrificing Performance
- Provides High-level Components
- Enables Transfer Learning with Minimal Code
- Audience: Everyone!
 - Beginners in Machine Learning
 - Rapid Prototyping for Experts



Great! What Do I Need to Know?

Core Components:

- Data Loaders - Flexible Data Loading and Transformations
- Learner - Training Models

Wide Variety of Domains (these are Data Loaders!):

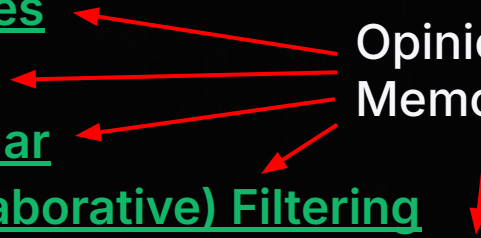
- Images
 - Text
 - Tabular
 - (Collaborative) Filtering
 - Data Block (Generic/Custom)
 - And more!
- Opinionated Data, Pipeline,
Memory Management
- 



Image Classifier

Discussion and Demo

Problem to Blueprint

Image Classifier for Dogs (Boxer, German Shepherd, Golden Retriever)

- **Need to Create a Dataset**
 - 30 Unique Images of Each Type of Dog
 - Make Images Uniform
 - Incorrectly Download Dogs?
- **Build and Train the Model**
 - Create Dataset Based on Categories
 - Augment Pre-Trained Model with Layers
 - Train the New Model (Apply Weights)
- **Test the Model with Images**
 - Success? Probability?



Demo

Who Doesn't Love Dogs?!?

Natural Language Processing

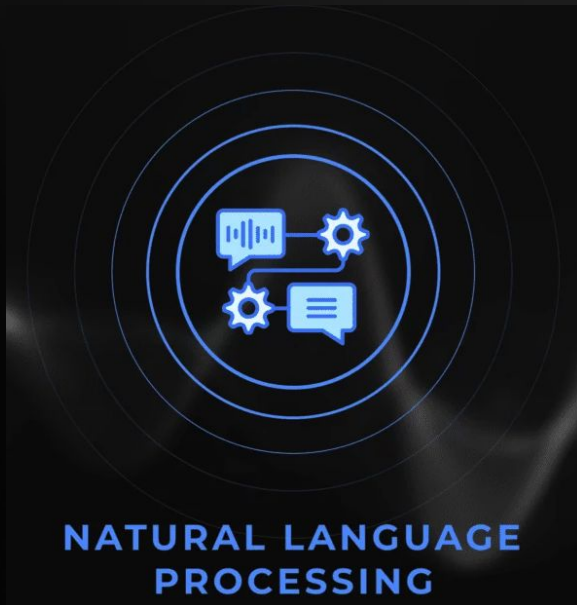
Discussion and Demo

<https://youtu.be/DC9Q6hIQiTs>

Question vs Non-Question Sentences

How to Approach This Problem:

- Need to Create a Dataset
 - Big Problem: Where to Get Data?
 - Qs: Stanford Question Answering Dataset
 - Remove the Answers
 - Non-Qs: Use Answers OR Rand Wiki Pages
- Build and Train the Model
 - Using Tabular/DataFrame
 - 2 Cols: Sentence, Is Question
- Test the Model
 - Create a Q and Non-Q and Test!



Demo

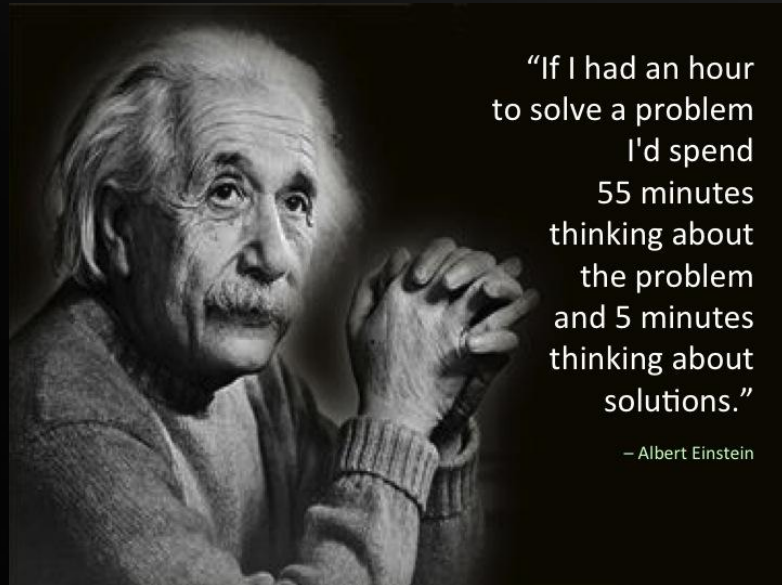
Identifying Questions in Text

<https://youtu.be/ijzgaRlrYSg>

Right Data for the Right Problem

Why Did We Fail At First?

- **Dataset Needs to be Representative**
 - Obtain the Right Kind of Data
 - Verify the Data is Correct
 - Modify the Data to Fit?!?
- **Verify the Model**
 - Test and Retest
 - More Data or More "Alignment"
 - Repeat!

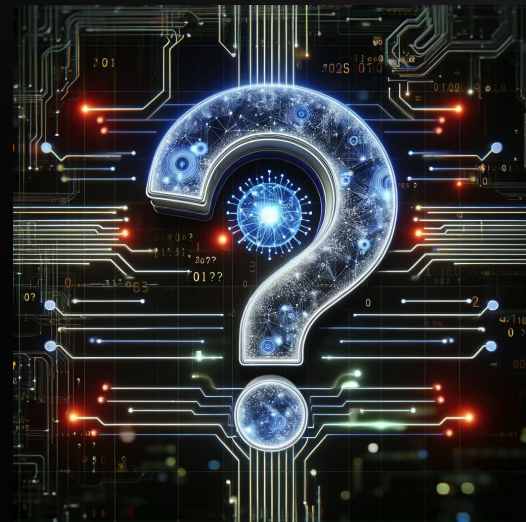


Getting Answers to Questions

Use Case: In Meetings, Phone Calls, etc, Use an LLM to Answer Questions

- Useful Because:
 - Provide "Level Set" on Participant Knowledge
 - Get Answers Now vs "Taking Offline"
 - Provide Other Lesser Known Facts
- LLM Could Be:
 - Geared Toward a Domain (ex. Medical)
 - Trained on Company Docs
 - Generalized

Demo: Deepgram STT + OpenAI + Deepgram TTS



Demo

Answering Questions with LLMs in Real-Time
Using Deepgram and OpenAI

<https://youtu.be/NCDLFGmK9Sw>

Resources

Materials, Links, Docs, Extras, etc

All Of The Things

[\[CLICK HERE\] for All Material Contained in this Session \[CLICK HERE\]](#)

All the Things in This Presentation

- [Example: Which Breed of Dog?](#)
- [Example: Question vs Non-Question](#)
- [Example: LLM Answering Questions in Meetings](#)
- STT & TTS: [Deepgram Speech-to-Text and Text-to-Speech](#)
- Additional Readings:
 - [Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter by Wes McKinney](#)
 - [Deep Learning for Coders with fastai and pytorch by Howard Gugger](#)

All Of The Things (part 2)

[\[CLICK HERE\]](#) for All Material Contained in this Session [\[CLICK HERE\]](#)

How To Get Started

- Tutorials
 - FREE Online Courses - <https://course.fast.ai/>
 - Examples via Jupyter Notebook - <https://github.com/fastai/fastbook>
- Image Processing
 - [Kaggle Jupyter Book](#) by [Jeremy Howard](#)
 - [FastAI Data Tutorial - Image Classification](#) by [Julius Simonelli](#)
- Natural Language Processing
 - [NLP from Scratch with PyTorch, fastai, and HuggingFace](#) by [Amar Saini](#)
 - [A Hackers' Guide to Language Models](#)



Thank you

David vonThenen
Software Engineer/Developer Advocate

   **@dvonthenen**

AI Dev
Summit

San Francisco, CA
May 29-30, 2024