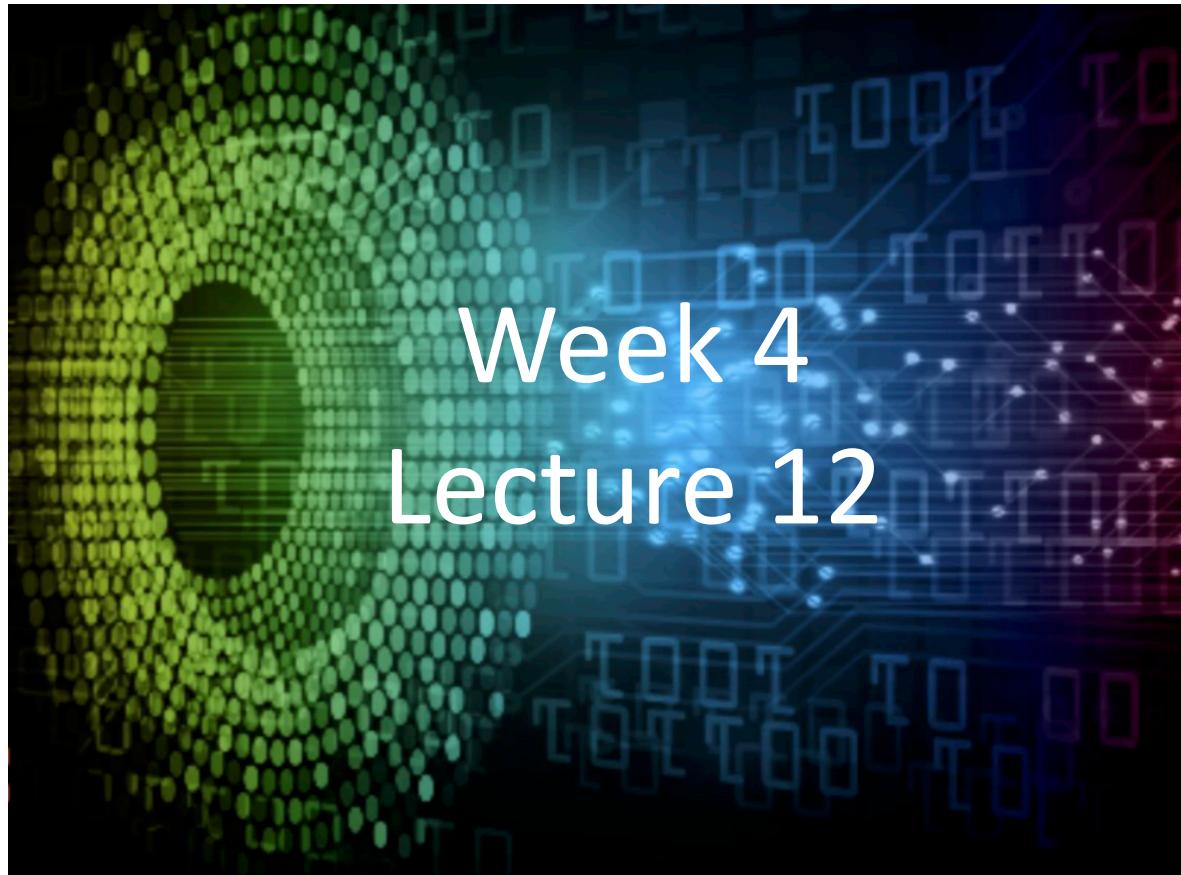


# Introduction to Deep Learning Applications and Theory



AMATH 563

# This Lecture: Project in ML/DL

- **Developing DL Project**
  - Process
  - How to pick a problem
  - Data sources
  - Testing strategies



# Developing a DL Project

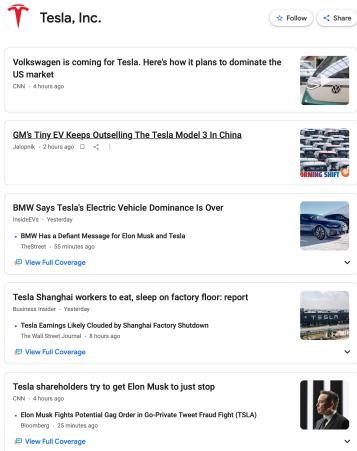
## ➤ Supervised Learning

- Select a Problem
- Get Data
- Choose Architecture
- Train the Model
- Test the Model

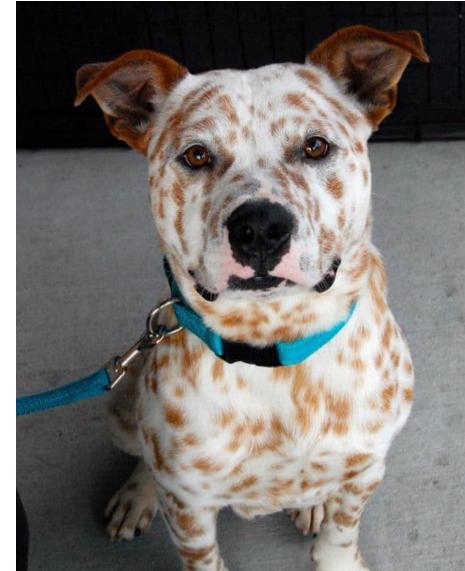
## ➤ Unsupervised Learning

## ➤ Fundamentals; DL Theory

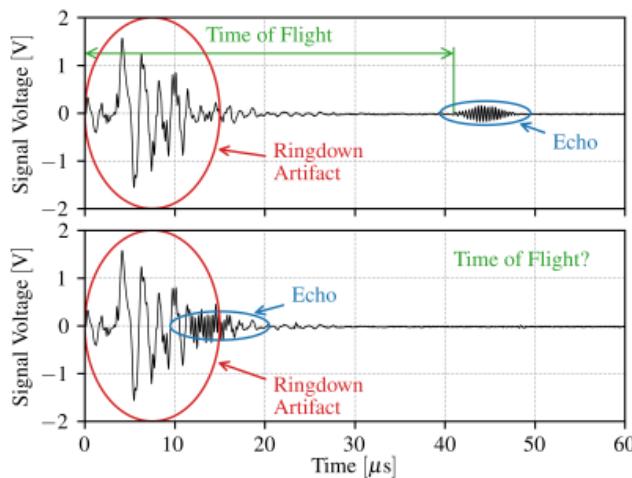
# “Sample” Projects



## Tesla News Bot



## Freckles Counter



## UltraSound Ringdown Artifact Removal

# Survey: Key Aspects

Q1. List Key Aspects for a Successful DL/ML/NN Project.

# Selecting a Problem

- Relevance / Interest
- Data Availability
- Domain Knowledge
- Usefulness / Utility
- Feasibility

**“AI Ethics”**

**Limitations/Risks/Biases/Uncertainty**

# Options

- Search recent projects **SOTA**
- Contests



 Microsoft | Microsoft AI

**AI for Good  
Idea Challenge**

 **eScience Institute**  
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

# Survey: Domains of Expertise

Q2. What is your Domain Knowledge/Interests?

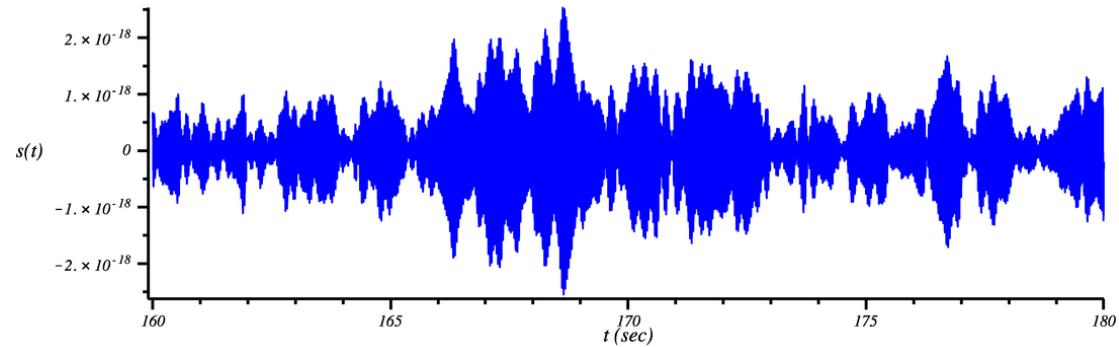
# Survey: Data Collection

**Q3. How **many days** would you spend on collecting data?  
Where will you collect the data from?**

# Data Collection and Training

- Concepts to take into account
  - **Small batches** of data
  - **Establish needed** components
  - **Overfit** the network initially
  - **Train and Enhance**

# Unstructured Data



audio



images



videos

"Hello, congratulations  
you have won \$100000000"

natural language

# Datasets

0 0 0 0 0 0 0 0 0 0 0 0 0 0  
1 1 1 1 1 1 1 1 1 1 1 1 1 1  
2 2 2 2 2 2 2 2 2 2 2 2 2 2  
3 3 3 3 3 3 3 3 3 3 3 3 3 3  
4 4 4 4 4 4 4 4 4 4 4 4 4 4  
5 5 5 5 5 5 5 5 5 5 5 5 5 5  
6 6 6 6 6 6 6 6 6 6 6 6 6 6  
7 7 7 7 7 7 7 7 7 7 7 7 7 7  
8 8 8 8 8 8 8 8 8 8 8 8 8 8  
9 9 9 9 9 9 9 9 9 9 9 9 9 9

MNIST



ImageNet



MS COCO



CIFAR10



IMDB Reviews



Free Music Archive

# Data Resources



amazon mechanical turk

# Data Processing

- Transformation
  - (audio: frequencies vs. raw)
- Sufficient number of features
  - (video: is high res needed?)
- Classical (non ML) vs. NN?

# Dilemma:

Example: freckles; image resize

- Non ML
- Small NN to perform preprocessing

# Comparison

	Non ML	Small NN
Time	30min	?
Domain experience	?	?
Data Changes	yes	no

# Testing / Deploying

- Efficiency bar
- Data Variability
- Adversarial examples
- How will be used/ deployed