# IST 402
# Week 5 Notes

Shaun Campbell, Brennan McKendree, Nic Ammazzalorso, Andrew Takahashi

# What’s Wrong with AI and ML

- Lots of things
- Several of them are legitimate issues
  - Have been tackled by ML researchers
  - Led to different emerging fields in these areas
  - These will be covered in more detail in the next couple of weeks
- But depending on what kinds of sources you read from, the AI/ML issues that you might hear of might be completely different.
  - A lot of noise in the news

- Learn how to weed out illegitimate concerns from legitimate ones
- Understand how and why people make mistakes when they point out these legitimate ones

# Concern 1: AI is going to take away all of our jobs

- Case in point:
  - Manufacturing assembly lines
    - Past Humans
    - Now/Future: Machines or AI
  - Cashiers at Fast Food Joints/Grocery Stores
  - Communications for societies
- Taxis and Ubers
  - Truck drivers - Otto
- Maid Services Vacuuming
  - Roombas
- Marketing and advertising
  - Ad exchanges
- Robots to check inventory in departmental stores
  - Amazon Go
- Stock Markets (NYSE, Nasdaq)
- Call center operations (IVRS systems)

# Concern 2: Artificial General Intelligence is Near

We will build autonomous agents that operate much like beings in the world

- Lots of news stories that AGI is just around the corner

Modern day AGI research is not doing at all well

- Mostly seems stuck on the same issues in reasoning

The Singularity is Near

- 2029 is when we would be able to simulate the functioning of the entire human brain
  - Millions of neuron cells, and billions of connections within these cells
- Refers to a point where AI is better at AI research than humans
  - It will recursively improve itself
  - Will no longer be in control of human beings
- Current State:
  - AI system trying to understand a 100 line C++ program
  - Unable to beat a freshman student who has just taken one month of programming lessons
  - C Elegans
    - Nervous system of this worm has 302 neurons and 6000 connections in between these neurons
    - Over the past 30 years, people have been figuring out the entire wiring pattern of the 302 neurons
    - Modeling the neural system of C elegans is still ongoing - not even halfway there

# Concern 3: Misalignment of values between ML & AI

- Misalignment of values
  - AI is programmed for only one task
  - AI does not know what to do if it cannot accomplish that task
  - Turning the device off will disallow it to complete task
  - AI may program the off switch to not shut off the power
- Examples like this are instances in which there is a misalignment of ML & AI

**Concern 4: Robots will kill us all**

- Not close to becoming a reality anytime soon

# Issues with Deep Learning:

- Is Deep Learning approaching a wall?
  - “ for most problems where deep learning has enabled transformationally better solutions, we’ve entered diminishing returns territory in 2016-2017

- What is Deep Learning good at?
  - Just a statistical technique
  - Has a set of assumptions that it works with
  - Performance is not good when these assumptions are not satisfied:
    - Having enough data
      - Deep learning can work with raw data where standard ML models extract “important” features from this raw data and usually this happens using a hand-designed feature extractor
- No bias in training data
  - DL models are just as likely to suffer against bias data

- Data from the real world should be similar to your training data
  - Training data should be a good enough representation of the type of data you will see in the real world
  - The distributions of your training and test data should be the same (or highly similar)

# Limit 1 - Deep Learning is Data Hungry

- If you having training data ➜ DL works well

- Contrapositive of this statement
  - DL doesn't work well ➜ ?
  - In real life you often don't have enough data
- Interpolation
  - If your test data is coming from the same distribution, your DL model should be able to interpolate between things that it has seen before
- Extrapolation
  - If your test data is not coming from the same distribution, DL model needs to extrapolate knowledge that it has currently learnt
  - IMPORTANT: no way to extrapolate currently
- Lacks mechanism to learn abstractions through verbal explicit definition

# Limit 2 - Deep Learning is Shallow

- Does not learn any hidden abstractions similar to human beings
  - These abstractions allow us to transfer knowledge
  - DL can’t do that

# Limit 3 - No Way to Deal with Hierarchical Structure of English

- RNNs represents sentences as sequences of words
  - Ignore hierarchical structure
  - Longer sentences constructed recursively using smaller sub-sentences
- Issue: No hierarchy among set of features, all of them are flat.  We draw correlations among them

# Limit 4 - Open Ended Interface

- Inference has been limited to Squad (Stanford Question Answer Database) type queries
- Given a question and a piece of text
  - Infer answer to question by reading text
    - Assumption: answer is present in text
- Thing that have not been done:
  - Multi-hop inference
    - Locate answers by combining multiple pieces of text
    - Combine text with background knowledge
    - Open Ended Inference example: I think you need to mind your own business
      - Question: What is the mood of the person?
  - Human beings can do this opened ended inference
  - Deep learning cannot

# Limit 5 - Lack of Transparency

- Deep learning is a black box
- Millions or billions of weights
  - All you can get is the values of these learned weights
  - How do you interpret them?
- Why is this even important? In what domains?
  - Viewpoint 1: Depends, if you are just looking for good results, you don’t need transparency, but if you are scientists working at Google who want to understand better, you need transparency
  - Depends on the domain where its being used, if it's being used in regards to people’s health, then you need to understand why a deep learning model is making some prediction
  - Practitioners need to be able to trust the machine learning system that they are using
    - Who is accountable when machine learning makes a mistake? The machine learning model goes scot-free but the doctor gets sued

# Limit 6 - Not Integrated with Prior Knowledge

- No domain knowledge is input
- Standard machine learning used feature extractors which were designed by human experts and contained human insights into the domain
- But you don't have human designed feature extractors in deep learning
- Useful properties of images, text, or whatever kind of data is being used is not present in the deep learning model
- One solid exception
    - convolutional neural networks

# Limit 7 - Unable to Model Causation

- Correlation does not imply causation
- Deep learning system can learn correlations between height and vocabulary
- Will not be able to uncover causation between growth and development to both these variables

# Limit 8 - Assumption of Stationarity

- Deep learning works well with stationary environments
- What if rules of the world continuously change?
  - What about stock prediction? Flu prediction?
- How is this related to extrapolation and difference in training testing data?

# Limit 9 - Deep Learning Can Easily be Fooled

- Deep learning can easily be fooled by simply adding noise to your data