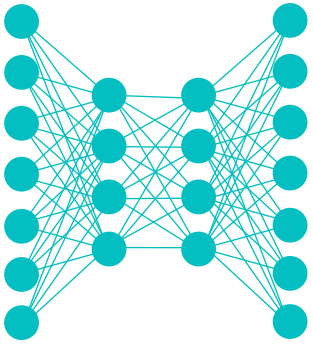
Lecture Notes for

Neural Networks and Machine Learning

Course Introduction

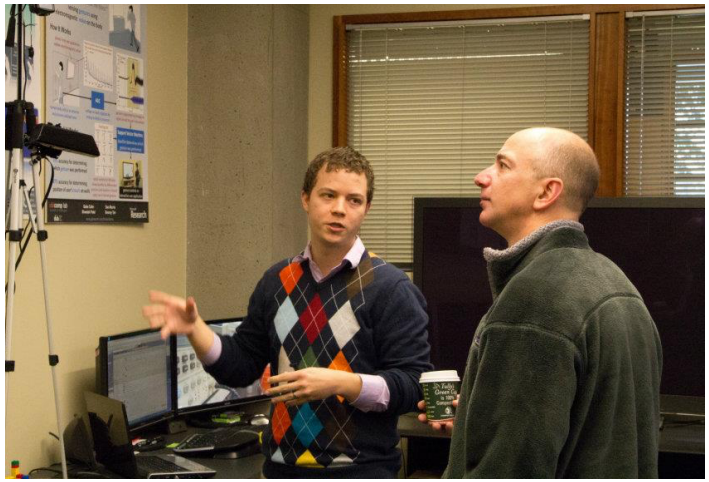


Logistics and Agenda

- Logistics
  - Canvas Access?
- Agenda
  - Introductions
  - Syllabus
  - Presentation Selection

Introductions

- Name
- Department
- Home
- 2 Truths and 1 Falsehood
  - Example: I gave Pitches on Machine Learning to Elon Musk, Bill Gates, and Jeff Bezos

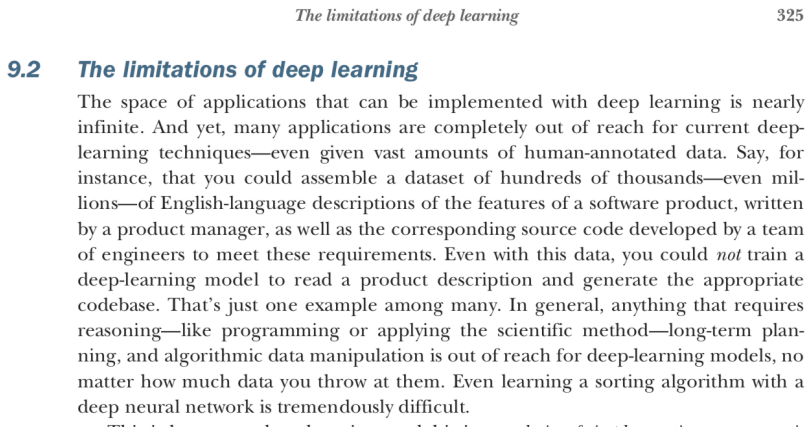


Syllabus

- Reading
- GitHub
- Grading
- Participation
- Course Schedule

Presenting

- First Presentation is Next Week!
- During Semester: Eight Presentations Total
- First Presentation:
  - Section 9.2 of Chollet Book, Limitations of DL
- **Who wants to go first?**
  - 15 Minutes
  - Summarize the Article
  - Make 2-5 Visuals
    - Slides
    - Handouts
    - Notebooks



The limitations of deep learning 325

9.2 *The limitations of deep learning*

The space of applications that can be implemented with deep learning is nearly infinite. And yet, many applications are completely out of reach for current deep-learning techniques—even given vast amounts of human-annotated data. Say, for instance, that you could assemble a dataset of hundreds of thousands—even millions—of English-language descriptions of the features of a software product, written by a product manager, as well as the corresponding source code developed by a team of engineers to meet these requirements. Even with this data, you could *not* train a deep-learning model to read a product description and generate the appropriate codebase. That's just one example among many. In general, anything that requires reasoning—like programming or applying the scientific method—long-term planning, and algorithmic data manipulation is out of reach for deep-learning models, no matter how much data you throw at them. Even learning a sorting algorithm with a deep neural network is tremendously difficult.

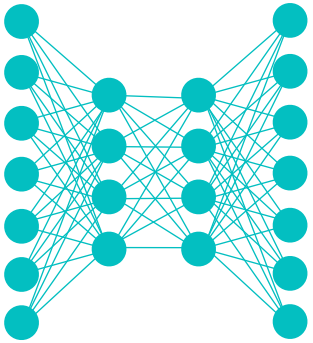
Lecture Notes for

Neural Networks and Machine Learning

Course Introduction

Next Time:
Case Studies in Ethics of ML

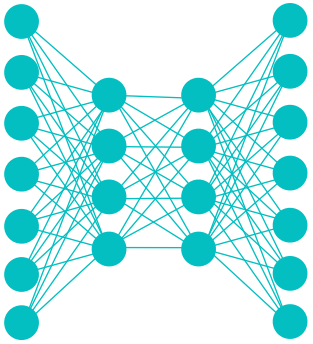
Reading: None



Lecture Notes for

Neural Networks and Machine Learning

Case Studies in Ethical ML



Logistics and Agenda

- Logistics
  - Presentation to start next time
- Agenda
  - Problem Overview of Ethical Machine Learning
  - Case Studies and Discussion

Ethical Considerations in ML

From Thomas Davenport, fellow of MIT Digital Economy

- Make AI ethics a board level issue
  - Not up to one or two developers
- Promote fairness through avoiding bias
  - Black box ML will reflect bias and project bias
- Lean towards disclosure of ML use
  - How will data be used? What data?
- Tread lightly on privacy
  - Facebook?
- Help alleviate employee anxiety
  - Programs for explaining automation
- Work with Humans in the loop
  - You will never catch it all
  - opt for content review and beta feedback

https://sloanreview.mit.edu/article/every-leaders-guide-to-the-ethics-of-ai/

Google as an Example

- Be socially beneficial
- Avoid creating or reinforcing unfair bias
- Be built and tested for safety
- Be accountable to people
- Incorporate privacy design principles
- Uphold high standards of scientific excellence
- Be made available for uses that accord with these principles
- Google will not pursue: Tech likely to cause harm, tech that principally is a weapon, Tech that violates surveillance norms, Tech that contravenes human rights

https://www.blog.google/technology/ai/ai-principles/

How is Google doing?

FeiFei Li, in an email to other Google Cloud employees:

"Avoid at ALL COSTS any mention or implication of AI. Weaponized AI is probably one of the most sensitized topics of AI — if not THE most. This is red meat to the media to find all ways to damage Google."

Opinion: There's more to the Google military AI project than we've been told

by TRISTAN GREENE — 7 months ago in ARTIFICIAL INTELLIGENCE

What about Facebook?

Machine Learning – Facebook Research

https://research.fb.com/category/machine-learning/ ▼

Our **machine learning** and applied **machine learning** researchers and engineers ... The **Facebook** Field **Guide** to **Machine Learning**, Episode 6: Experimentation.

Missing: ~~ethics~~ | Must include: ethics

Case Studies In Ethical ML

Case Study: ML Generated Reviews

- Which of these are fake:
  - "I love this place. I have been going here for years and it is a great place to hang out with friends and family. I love the food and service. I have never had a bad experience when I am there."
  - "I had the grilled veggie burger with fries!!!! Ohhhh and taste. Omgggg! Very flavorful! It was so delicious that I didn't spell it!!"
  - "My family and I are huge fans of this place. The staff is super nice and the food is great. The chicken is very good and the garlic sauce is perfect. Ice cream topped with fruit is delicious too. Highly recommended!"
- Does this violate any ethical guidelines?
- "While this study focuses only on creating review text that appears to be authentic, Yelp's recommendation software employs a more holistic approach," said a spokesperson. "It uses many signals beyond text-content alone to determine whether to recommend a review."
- Does the mere presence of this cause problems of trust?

https://www.theverge.com/2017/8/31/16232180/ai-fake-reviews-yelp-amazon

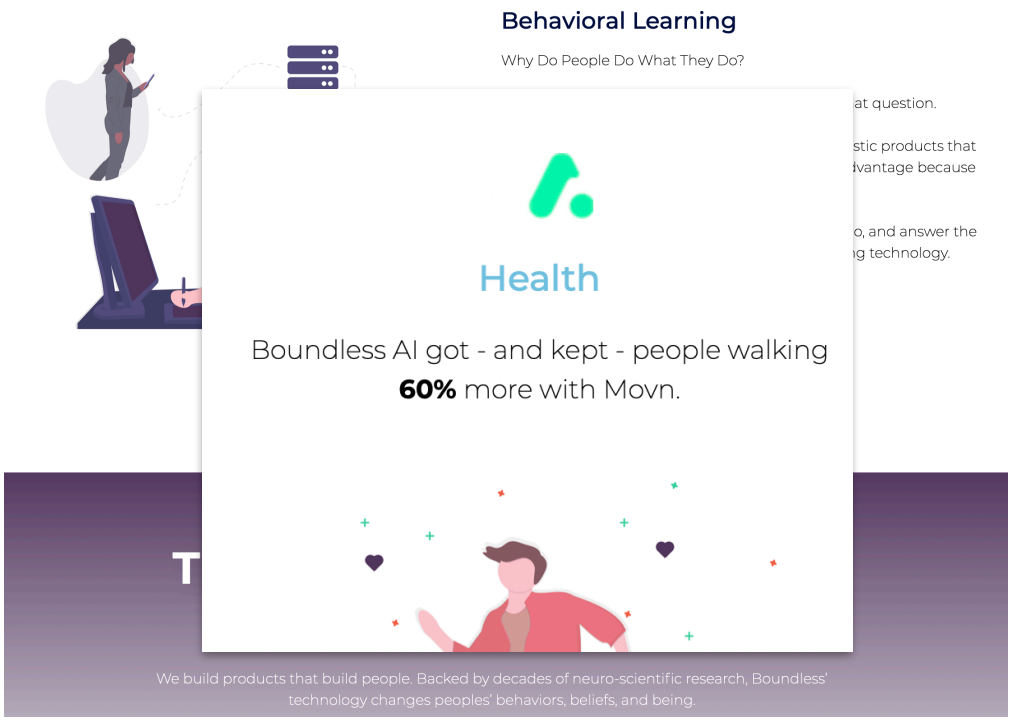
Case Study: Face Swapping

- Does the mere presence of this cause problems of trust?

https://www.youtube.com/watch?v=gLol9hAX9dw

Case Study: Reinforcing App Addiction

- Identifying behavior to keep users in your app
- Does this violate any ethical guidelines?



Ultimately, Dopamine Labs predicts they can add 10 percent to a company's revenues. In practice, their numbers are a bit all over the map, with some companies seeing bounces of more than 100 percent in terms of user interactions with, in or on an app. For other companies the boost could be around 8 percent.

https://techcrunch.com/2017/09/08/meet-the-tech-company-that-wants-to-make-you-even-more-addicted-to-your-phone/

Case Study: Reinforced Gender/Race Bias

- Not a new problem in technology:
  - Example: Crash Test Dummies, Because most crash tests have male "dummies" female had a 20 to 40 percent risk of being killed or seriously injured, compared to 15 percent for men.
- But can also be more subtle:

Does this violate any Ethics Principles?

Internet Culture

Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you.

"It's part of a cycle: How people perceive things affects the search results, which affect how people perceive things," Cynthia Matuszek, Professor of Computer Ethics at UMD

https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/?noredirect=on&utm_term=.055bff1a94ad

Case Study: Predictive Policing

- Once a crime has happened, can it be classified as a gang crime?
  - Used partially generative NN for classifying if a crime was gang related, with the aim at predicting gang retaliation. Trained on LAPD data 2014-2016
- Does this violate any ethical guidelines?

Blake Lemoine AI Google Researcher On Bias in ML

But researchers attending the AIES talk raised concerns during the Q&A afterward. How could the team be sure the training data were not biased to begin with? What happens when someone is mislabeled as a gang member? Lemoine asked rhetorically whether the researchers were also developing algorithms that would help heavily patrolled communities predict police raids.

Hau Chan, a computer scientist now at Harvard University who was presenting the work, responded that he couldn't be sure how the new tool would be used. "I'm just an engineer," he said. Lemoine quoted a lyric from a song about the wartime rocket scientist Wernher von Braun, in a heavy German accent: "Once the rockets are up, who cares where they come down?" Then he angrily walked out.

https://www.sciencemag.org/news/2018/02/artificial-intelligence-could-identify-gang-crimes-and-ignite-ethical-firestorm

Case Study: ML Generated Products

- Online generation of content to facilitate buying behavior

It's not about trolls, but about a kind of violence inherent in the combination of digital systems and capitalist incentives. It's down to that level of the metal. This, I think, is my point: The system is complicit in the abuse.

And right now, right here, YouTube and Google are complicit in that system. The architecture they have built to extract the maximum revenue from online video is being hacked by persons unknown to abuse children, *perhaps not even deliberately*, but at a massive scale.

These videos, wherever they are made, however they come to be made, and whatever their conscious intention (i.e., to accumulate ad revenue) are feeding upon a system which was consciously intended to show videos to children for profit. The unconsciously-generated, emergent outcomes of that are all over the place.

—James Bridle

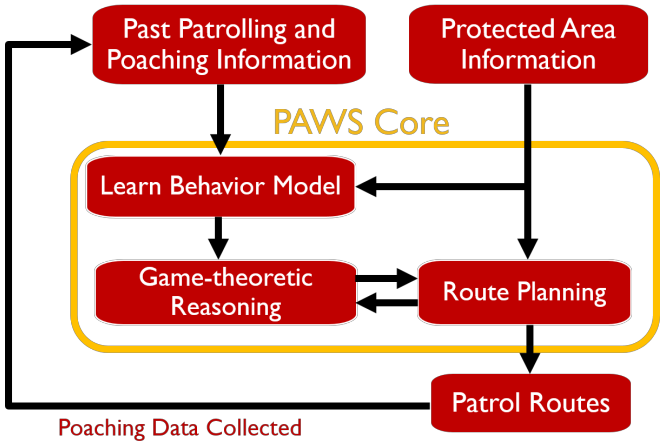
https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2

A Counter Example

- PAWS: Prevent Tiger Poaching in Malaysia

"Even if you can predict some sort of poaching activities, it's not always good to just go to areas with high predicted poaching activity!"

Prof. Fei Chang, CMU



"None of these aspects can be addressed with a publicly available commercial tool, or directly addressed by sitting in an office… That means we need to talk to experts, understand the problem, and propose solutions to it."

https://www.fastcompany.com/90157255/you-can-now-take-a-class-on-how-to-make-ai-that-isnt-evil

How to Make a Racist AI without Really Trying

Robyn Speer, 2017
http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/

Debiasing: Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings
Bolukbasi et al., NeurIPs 2016
https://arxiv.org/pdf/1607.06520.pdf

ConceptNet 5.5: An Open Multilingual Graph of General Knowledge
Speer et al., AAAI 2017
https://arxiv.org/pdf/1612.03975.pdf

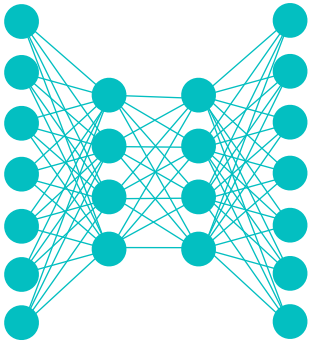
Lecture Notes for

Neural Networks and Machine Learning

Case Studies in Ethical ML

Next Time:
CNN Visualization

Reading: Chollet Article



Backup slides

Title Between Topics

Example Slide

Title

Subtitle

`Follow Along: Notebook Name`