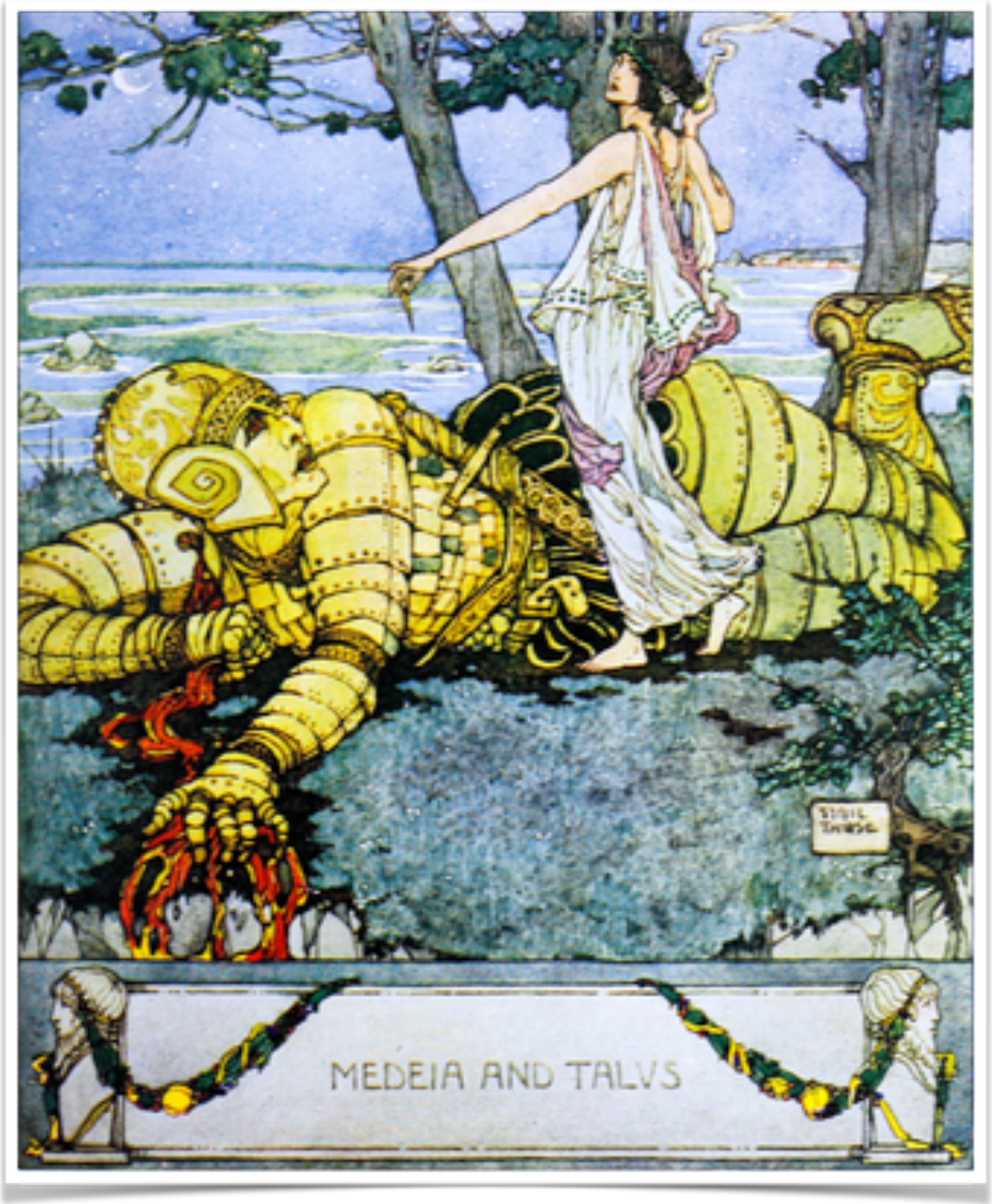
Open Problem: Bias, fairness, and ethics in deep learning

How is deep learning going to change our world

General Intelligence

- The singularity happens
- AI smarter than human
- DOOM!



How far are we from AGI?

- Far, and we're not getting much closer

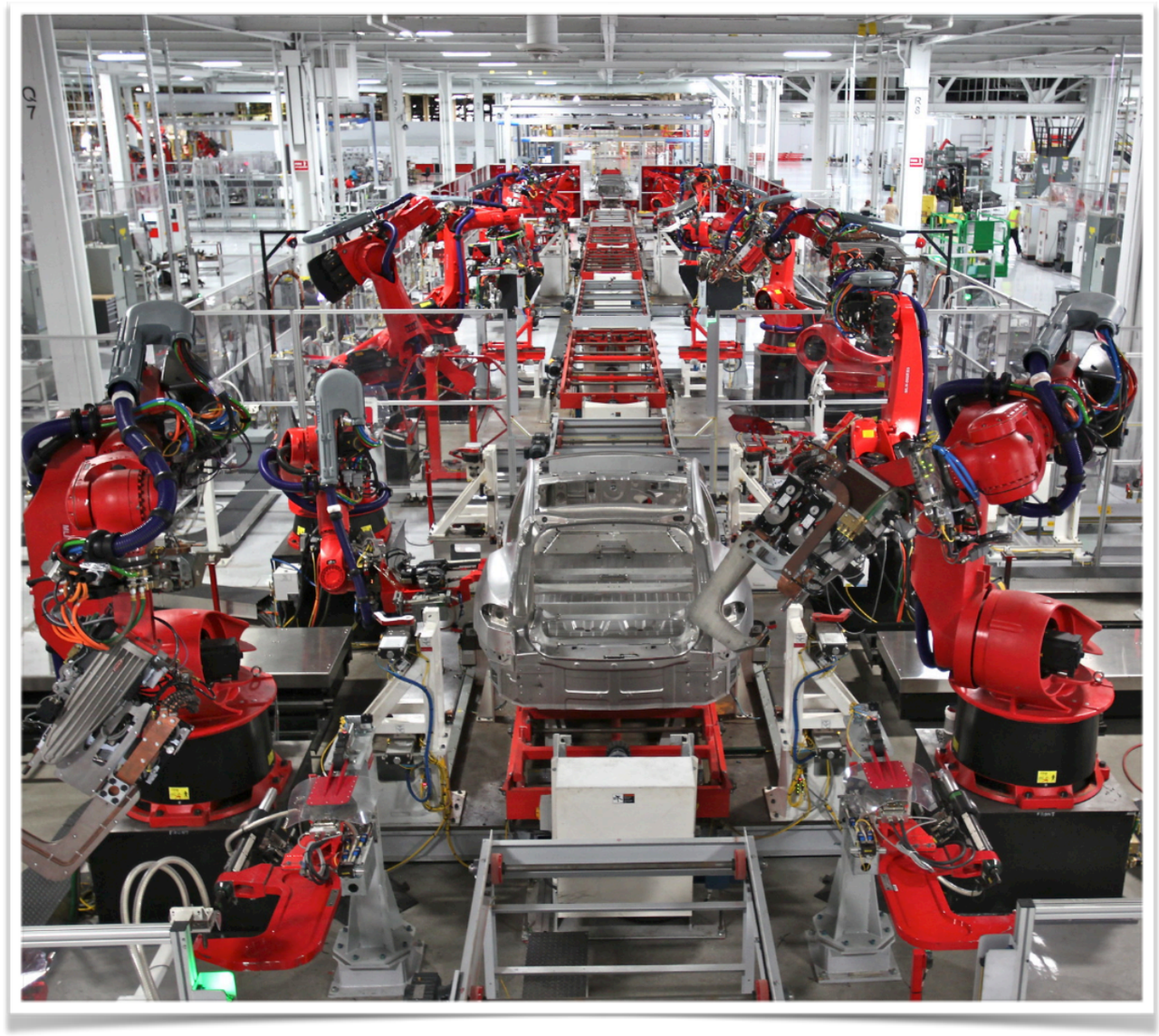
Dangers of dumb AI

- Dumb / narrow AI works
  - How will it change our world?

male
smile
black hair
...

Automation

- AI / ML replaces humans
  - Driving / trucking
  - Factories
  - Radiology
- Fewer humans needed

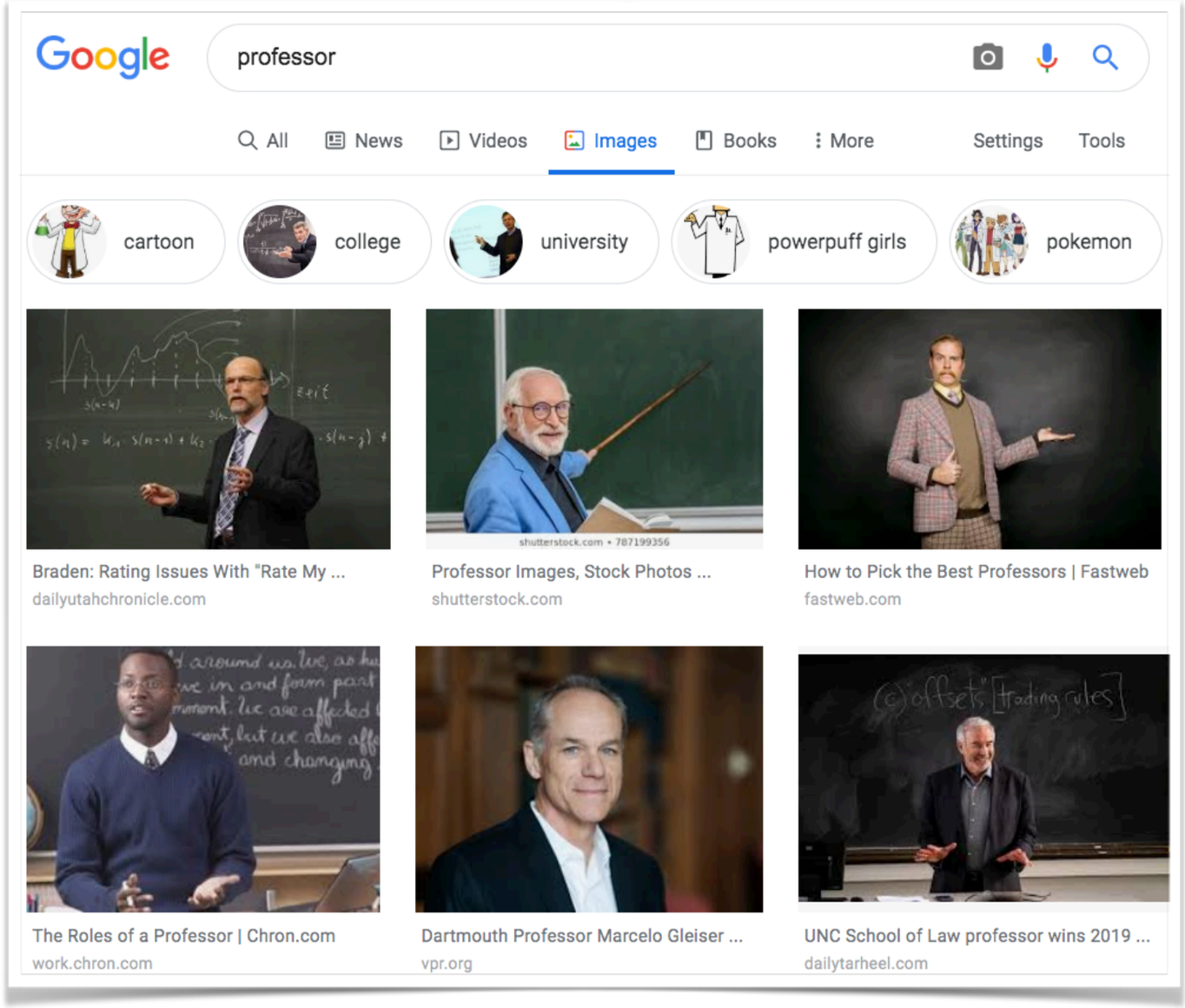


Misuse

- Fake images
  - Propaganda
- State sponsored surveillance

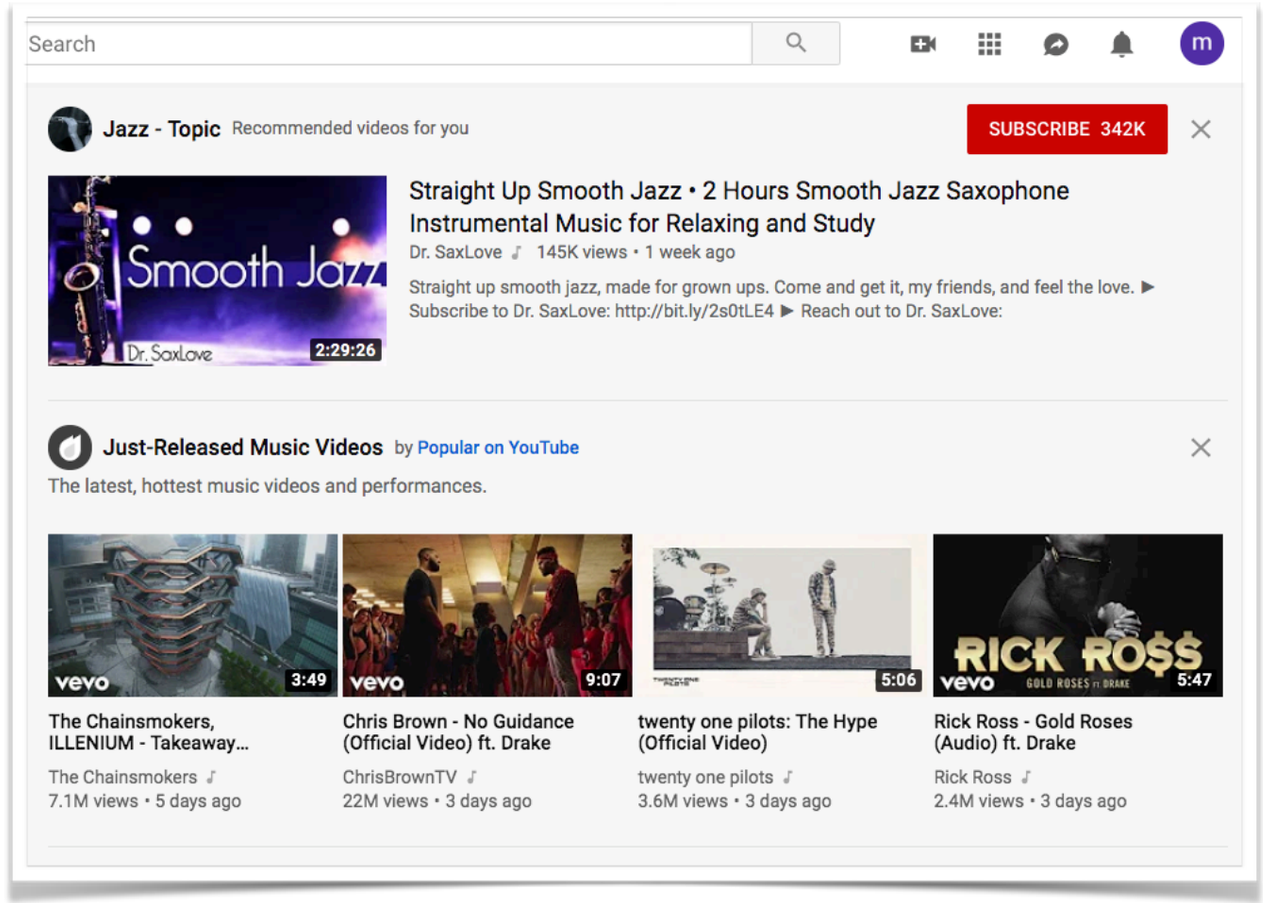
Bias

- DL systems inherit and amplify bias from data
- e.g. google photos
- credit applications
- criminal justice



Recommender systems

- Find user specific content
  - predict users preferences
    - e.g. movies on netflix
    - e.g. products on Amazon
- Optimized to make company money (e.g. click rate, time on screen)



Video-sharing websites

- Recommender systems alter user behavior
    - How do you keep users engaged?
        - Fringe content
        - Conspiracies
        - Alt-left, alt-right

- YouTube Radical: https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html
- Towards neural mixture recommender for long range dependent user sequences, Tang et al., WWW, 2019

Social media

- Recommender systems shape view of world
- What to show users? Real or fake news?

74. The role of social media is significant. *** has been a useful instrument for those seeking to spread hate, in a context where, for most users, *** is the Internet. Although improved in recent months, the response of *** has been slow and ineffective. The extent to which *** posts and messages have led to real-world discrimination and violence must be independently and thoroughly examined…

Report of the independent international fact-finding mission on Myanmar*, Human Rights Council

- A Genocide Incited on Facebook, With Posts From Myanmar's Military, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

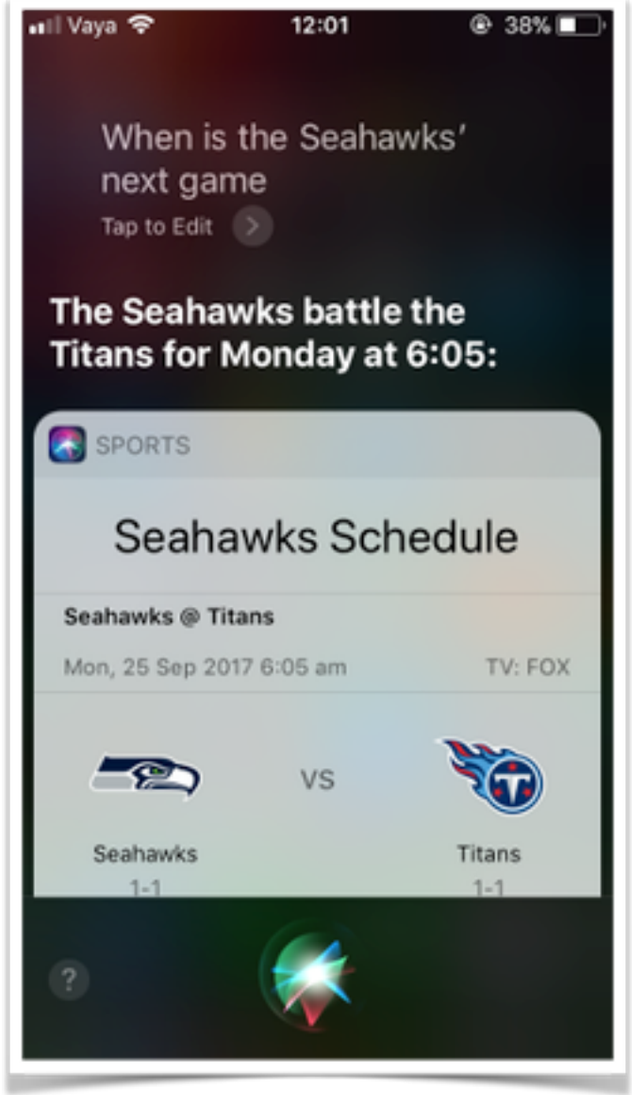
Dumb AI is dangerous too

- How can you help?
  - Understand the technology
  - Know that you see what an AI wants you to see
  - Ask for solutions
  - Help fix the problems

male
smile
black hair
...

Not all deep learning is bad

- AI assistant (Siri, Google, Alexa)
- Home security
- Healthcare
- Driving safety
- Disaster rescue

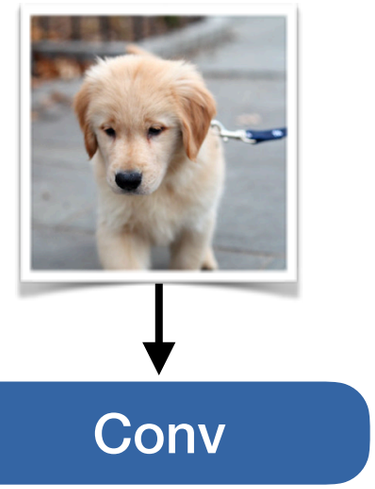


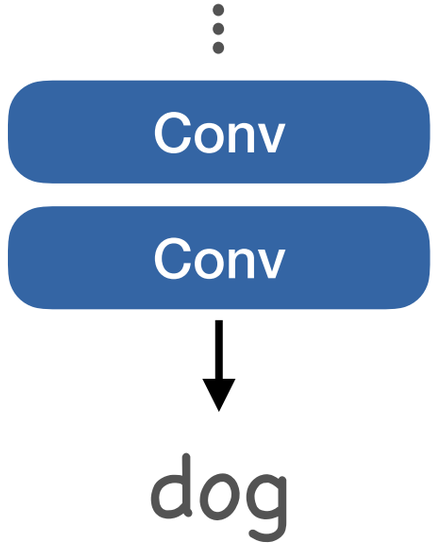
Course summary and further topics

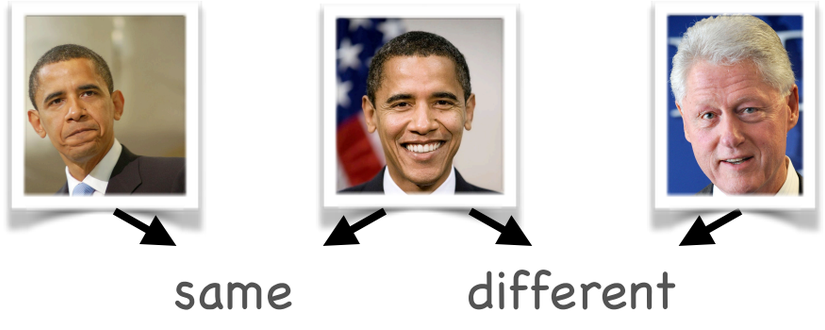
This course

- Fundamentals
  - How to build, train, use deep networks
- Few applications

```
In [ ]: %pylab inline
        import torch
        import sys
        sys.path.append('..')
        sys.path.append('../..')
        from data import load
        train_data, train_label = load.get_dogs_and_cats_data(resize=(128,128), n_images=10)
        device = torch.device('cuda') if torch.cuda.is_available() else torch.device('cpu')
        print('device = ', device)

In [ ]: class ConvNet(torch.nn.Module):
            class Block(torch.nn.Module):
                def __init__(self, n_input, n_output, stride=1):
                    super().__init__()
                    self.net = torch.nn.Sequential(
                      torch.nn.Conv2d(n_input, n_output, kernel_size=3, padding=1, stride=stride),
                      torch.nn.ReLU(),
                      torch.nn.Conv2d(n_output, n_output, kernel_size=3, padding=1),
                      torch.nn.ReLU()
                    )

                def forward(self, x):
                    return self.net(x)

            def __init__(self, layers=[32,64,128], n_input_channels=3):
                super().__init__()
                L = [torch.nn.Conv2d(n_input_channels, 32, kernel_size=7, padding=3, stride=2),
                     torch.nn.ReLU(),
                     torch.nn.MaxPool2d(kernel_size=3, stride=2, padding=1)]
                c = 32
                for l in layers:
                    L.append(self.Block(c, l, stride=2))
                    c = l
                self.network = torch.nn.Sequential(*L)
                self.classifier = torch.nn.Linear(c, 1)

            def forward(self, x):
                # Compute the features
                z = self.network(x)
                # Global average pooling
                z = z.mean(dim=[2,3])
                # Classify
                return self.classifier(z)[:,0]

        net = ConvNet()
```

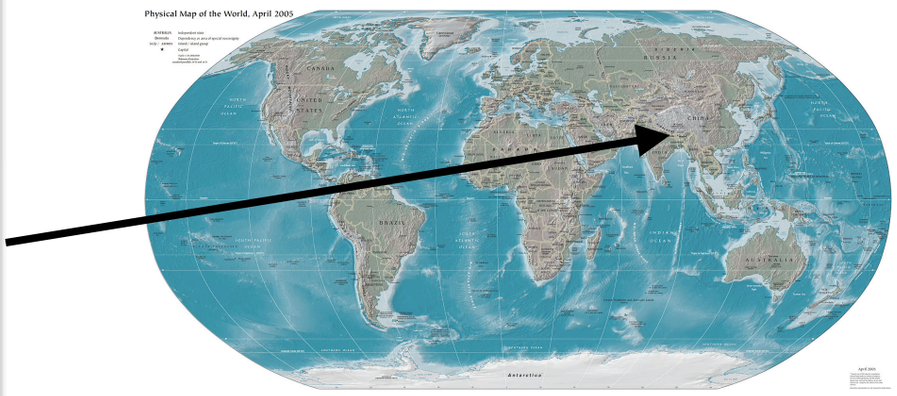


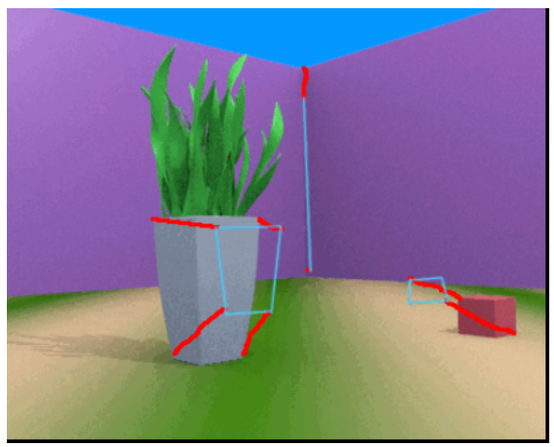




Computer vision

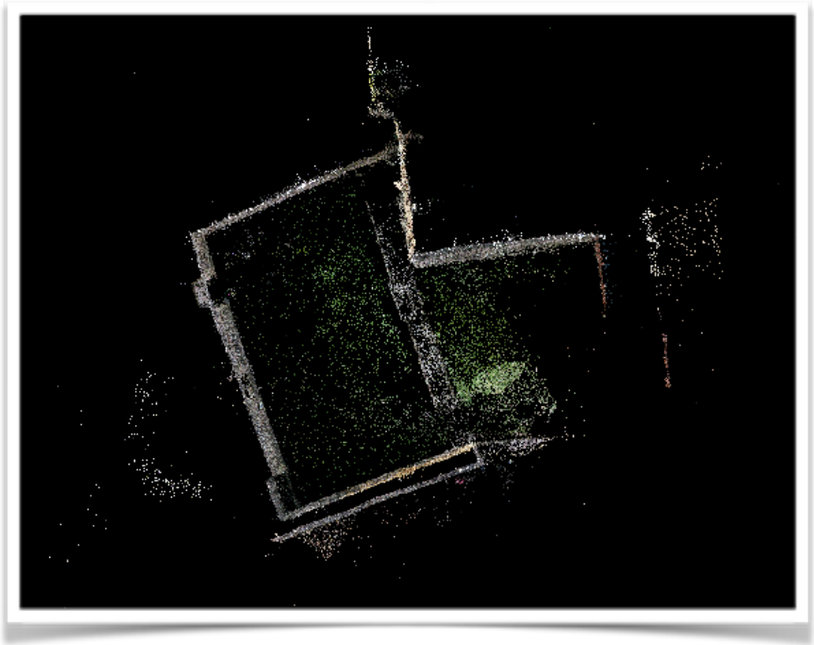
- Geolocalization
- Pose estimation
- Tracking
- Scene layout estimation
- Visual odometry
- …

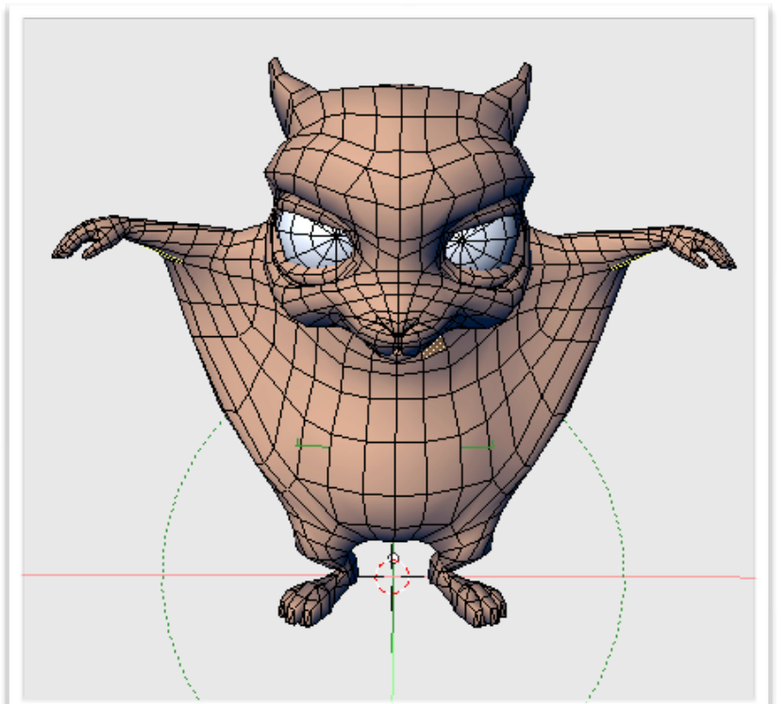




3D vision

- Point cloud or volume based networks
- Applications
- Reconstruction
- 3D recognition
- Surface representation
- …



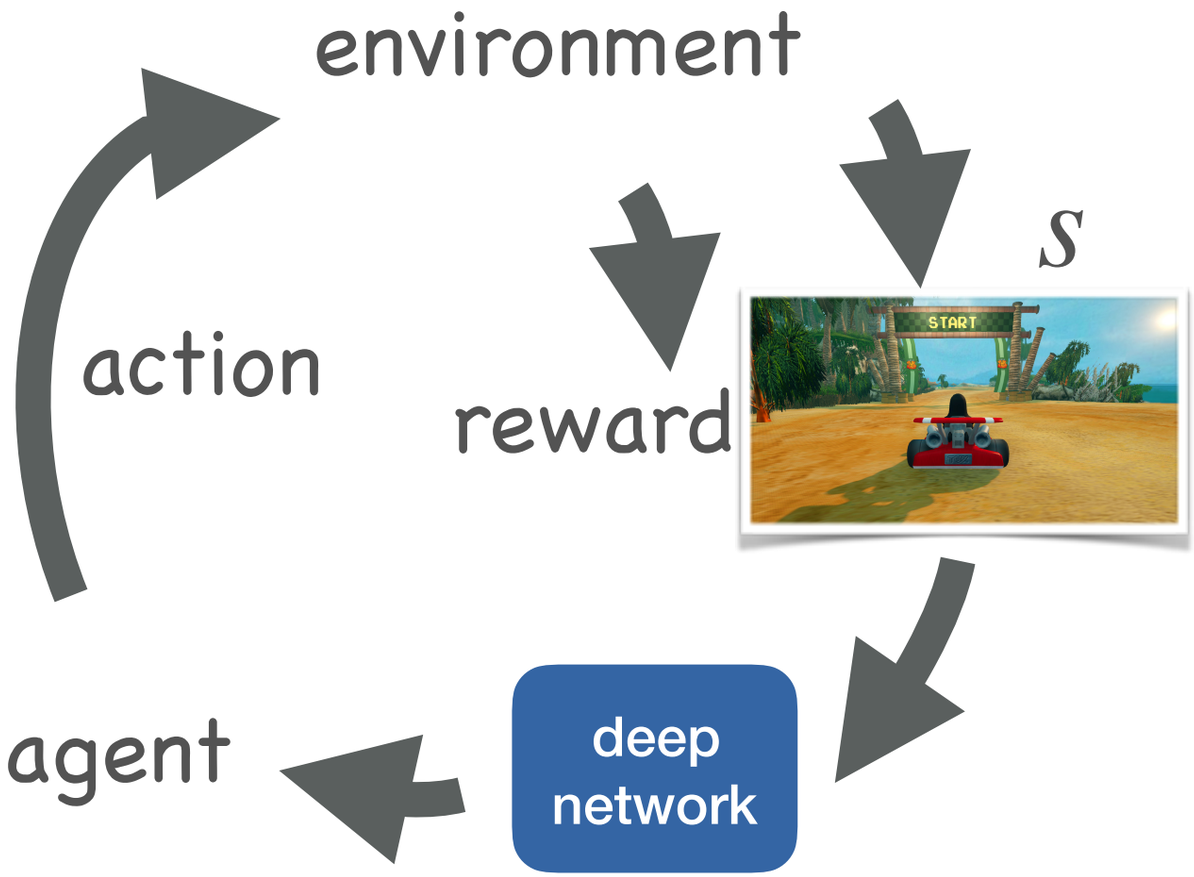


Natural language processing

- Word based models
- Applications
- Translation
- Sentiment analysis
- Topic modelling
- …

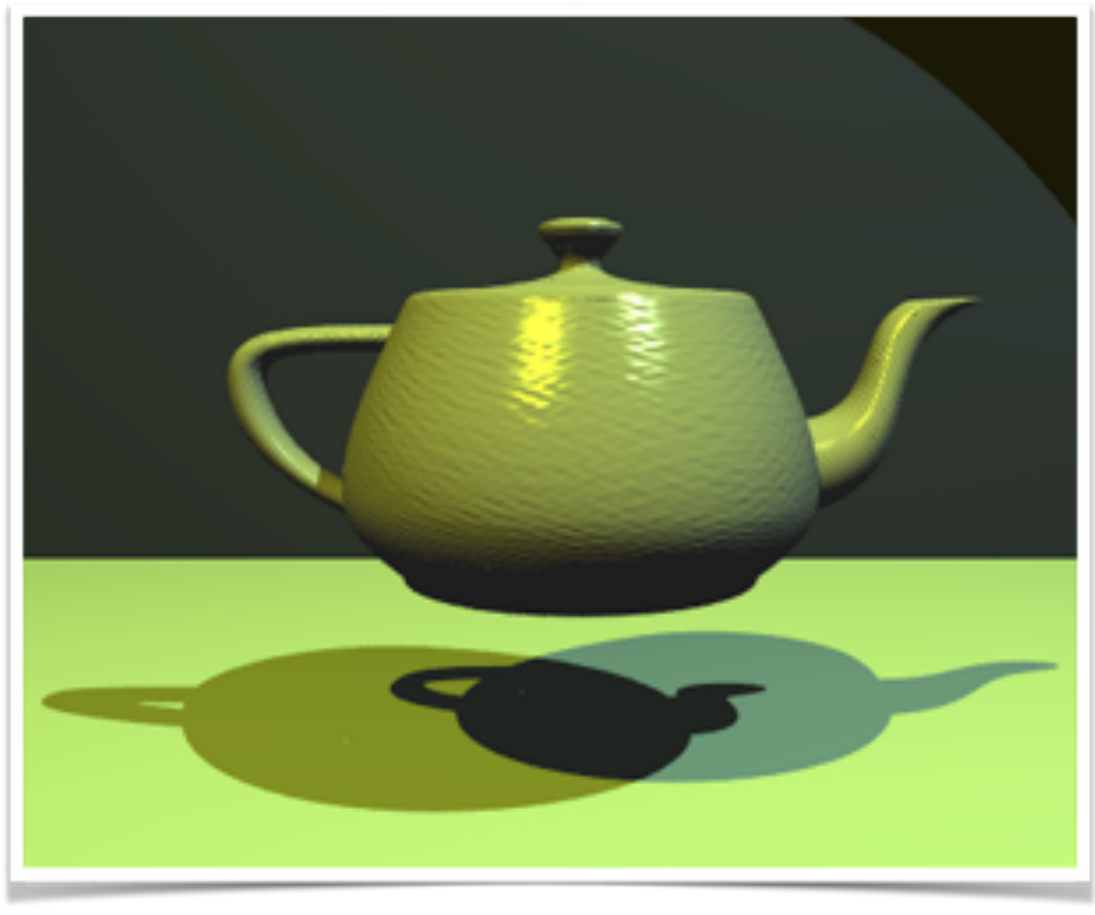
Reinforcement learning

- Q-learning
- Policy gradient++
- Applications
  - Robotics
  - Meta-learning
  - …



Compute graphics

- Generative models
- Applications
  - Matting
  - Image editing
  - Physical simulation
  - …



Deep learning hardware and architecture

- How do we implement any of this efficiently?
  - Fast matrix multiplications
  - Hardware support
  - …