Advances in Artificial Intelligence and the Impact on Test and Evaluation (T&E)
## Artificial Intelligence Taxonomy

### Reasoning
- Programmed
  - Planning/Logic/Expert Systems
  - Optimization
  - Agent-Based Systems

### Learning
- Supervised
  - Bayesian Networks
  - Regression
  - Neural Networks

- Reinforcement
  - Genetic Algorithms
  - Markov Decision Processes

- Unsupervised
  - Clustering
  - Principal Component Analysis

### Deep Learning
- Convolutional Neural Networks
- Recurrent Neural Networks
- Deep Q Learning
- Autoencoders
- Generative Adversarial Nets
Deep Learning revolution fueled by competition – ImageNet

ImageNet crowd-sourced database has grown to include 14M images, hand-annotated with 20K categories

In 2012, AlexNet was a breakout winner and used an 8-Layer CNNs, with CNNs being feasible for the first time due to GPU resources
ImageNet

GoogLeNet, 2014 Winner – 22 Layers

ResNet (Microsoft), 2015 Winner – 152 Layers
Deep Reinforcement Learning

- Use DNN to learn optimal policy for engaging in a stateful environment
- Advances in game play
  - AlphaGo
    - Fully observable, quantized state
    - Learn policy (6 weeks of GPU training)
- Video Games
  - 2015 – Atari Breakout (Google)
  - Today – Complex first person shooters
Generative Adversarial Networks

- Related concepts have existed since 1990, but in limited form
- Ian Goodfellow published current GAN concept in 2014
- Field has taken off

Edmond de Belamy
GAN-generated painting
Sold at auction for $432,500
Deep Fakes

- Ability to leverage latent space to create specific images
- Tie together audio and video synchronously
Tools

- Tools are widely available to develop, train, and execute networks in a plug-and-play way
- Google TensorFlow seamlessly integrates GPU capabilities
- Python integration makes it easy to compose networks
- Right now more “art” than “science” with how networks are constructed to achieve good performance
  - No rigorous methodology for composing layers into networks
  - “Hyperparameter optimization” is a fancy way of saying “brute force” network structure until it works

Deep Learning

- What society thinks I do
- What my friends think I do
- What other computer scientists think I do
- What mathematicians think I do
- What I think I do
- What I actually do

In [1]:
```python
import keras
Using TensorFlow backend.
```
Impact on T&E

AI for T&E

T&E for AI
AI for T&E

• Q: how can we use new AI algorithms to improve the T&E process?
### Synthetic Data Generation

- **Approach**
  - Use small sample of input data to train a GAN
  - Generate large amounts of input data across large latent space

- **Applications**
  - Primary applicable to testing individual units/components
  - Identify characteristics of failure modes
  - (Leverage latent space representation)

### State Space Exploration/Excitation

- **Approach**
  - Create an RL-based agent that seeks to interact with system
  - Create sequences of interactions that drive different state transitions within the system under test

- **Applications**
  - Subsystem and system-level testing
  - Suitable for T&E on stateful systems
  - Identify sequences that result in “bad” states
AI Test Automation
T&E for AI

• Many systems, military and commercial, are incorporating AI algorithms as part of their design and implementation
  • AI for machine perception – DCNNs for sensing the environment, and detecting/identifying objects within it
  • AI for system control – DRLs for understanding the environment and making optimal decisions within the estimated state

• Machine learning algorithms – challenge with repeatability and explainability
  • Under what circumstances does it fail?
  • How robust is it to noise?
  • How robust is it to failures (correlated noise)?
  • How robust is it to adversarial inputs?
Examples

“pig” + 0.005 x = “airliner”
Examples
Examples

Classifier Input

Classifier Output

- banana: 1.0
- slug: 0.0
- snail: 0.0
- orange: 0.0
Examples
Opportunities

**Systematic Testing**

- Leverage AI to test AI
  - Generate large synthetic data sets
  - Use latent space to identify areas where AI model fails
- Tests that characterize many models against range of criteria to identify those with most resilience
  - Noise
  - Distortion

**Formal Verification**

- Black box testing from digital circuits
  - Neural network has similar structure to circuit
  - Systematic exploration of all paths
  - Identify undertrained areas in network that most frequently lead to divergent output, and adjust model parameters to compensate
- Very preliminary research in this area

**Adversarial AI**

Need to develop a threat model for AI models – impact adversary can have depending on access
Training data, model parameters, runtime data, etc