What is AI?

A panel discussion on the opportunities and challenges presented by artificial intelligence

22 January 2019
Discussion Points

• CNA’s “AI with AI” podcast is now well into its 2\textsuperscript{nd} season
  – 62 episodes (as of 18 Jan) and counting; some reflections and major themes
• A bit of history: AI – machine learning – narrow AI – general AI – super AI
• Squashing myths and hype: expectations vs. reality vs. buzzwords
  – What AI is \textit{emphatically not}: an “off the shelf, ready-to-use” application
• Applications – \textit{hits, misses}, and the never-ending \textit{hype}
• Challenges – \textit{why “AI” is far from a panacea}
• Growing divide between the technically \textit{literate} and technically \textit{“informed”}
  – Increasing availability of AI development toolkits (\textit{Facebook}, \textit{Google}, \textit{Microsoft},...) and of open datasets, source code, benchmarks, and metrics
• Shifting timescales underlying basic research and defense acquisition
• Ethical and policy dimensions – AI as \textit{good, bad}, and \textit{indifferent}
This is not a formal presentation – only visual backdrops
What is AI? – *broad classic definitions, little consensus*

"AI is the field that studies the synthesis and analysis of computational agents that act intelligently"

[AI is] “that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment”

AI are systems that...

<table>
<thead>
<tr>
<th>Think like humans</th>
<th>Think rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Act like humans</td>
<td>Act rationally</td>
</tr>
</tbody>
</table>

- "The exciting new effort to make computers think . . . machines with minds, in the full and literal sense"
- “Automation of activities associated with human thinking, such as decision-making, problem solving, learning”
- “Study of mental faculties through the use of computational models”
- “Study of the computations that make it possible to perceive, reason, and act”
- “The art of creating machines that perform functions that require intelligence when performed by people”
- “Study of how to make computers do things that, at the moment, people are better”
- “A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes”
- “Branch of computer science concerned with the automation of intelligent behavior”
- .....
The Five “Tribes” of AI


Symbolists
- Symbolic rules used to draw inferences
- "Good Old Fashioned AI"
- IBM’s Deep Blue

Bayesians
- Probabilistic inference via likelihood estimates
- Bayesian Belief Networks
- If $a$ AND $b$ THEN $g$

Evolutionists
- Intelligence emerges from evolution of genetic variability
- Genetic Algorithms (GA)

Connectionists
- Intelligence derives from highly interconnected parts
- Deep Learning Neural Networks

Analogizers
- Function optimization, Nonlinear classification
- Support Vector Machines (SVM)

Deep Learning ⊂ Machine learning ⊂ AI
Narrow AI ← Perception, Learning, Abstraction, Reasoning → General AI

- Narrow AI – a system designed to handle a single task; or to deal with a narrowly-defined “problem”
- Designed to “learn” from data to improve performance on a task
  - Supervised: labeled data
  - Unsupervised: unlabeled data
  - Reinforcement: evaluative feedback
  - Deep learning: use of multiple hidden layers between input and output layers to find abstract representations of patterns in data
- Characteristics
  - Map fairly simple inputs to outputs (images, video, …)
  - Require huge datasets for training (e.g., Image classification: 1.2 million images)
  - Limitations and basic challenges (see dedicated slides)
  - Hidden/latent patterns that may bias data
- Applications
  - (Typically) do well on problems that humans are good at in fairly short time scales
  - Image recognition (medical diagnoses), speech recognition (language translation), game playing (though not all!), …
Google Trends


Web Search

News Search

Note

Note

Note
"We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions, and concepts, solve kinds of problems now reserved for humans, and improve themselves."

— Dartmouth AI Project Proposal, John McCarthy, 1955
AI Milestones: 1997 - 2019

**Era 1:** Machine learning
- 2009: ImageNet is Created (Fei Fei Li, 2007)
- 2010: Self-driving car prototype (Google: 2010)
- 2011: Google Brain (Ng, Dean: 2011)
- 2012: Watson defeats Jeopardy Champions (IBM: 2011)

**Era 2:** Machine learning
- 2009: GPU-accelerated learning (Google: 2009)
- 2015: CNNs dominate ImageNet (Ng, Dean: 2012)

**Era 3:** Contextual adaptation
- 2016: AutoML Introduced (Google: 2017)

**Just in 2017-19**
- 2019: Learnability can be undecidable (Ben-David: Jan 2019)

**2015**
- 2015: Project Maven (April 2017)
- 2015: AI Community calls ban of all offensive AWs (July 2015)
- 2015: Google employees resign over Maven (May 2018)
- 2015: Google decides not to renew Maven (June 2018)
- 2016: Joint AI Center (JAIC) (June 2016)
- 2016: AlphaGo beats Lee Sedol (DeepMind: March 2016)

**2017**
- 2017: AutoML Introduced (Google: 2017)

**2019**
- 2019: Learnability can be undecidable (Ben-David: Jan 2019)

**DNN Dark Period**
AI Milestones: expectations and computational backdrop

Perceptron / Rosenblatt

The New York Times
July 8, 1958

NEW NAVY DEVICE LEARNES BY DOING

Psychologist Shows Embry of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) — The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

9,000 IBM POWER9 CPUs and 27,000 NVIDIA Tesla V100 Tensor Core GPU

Exaflop performance ($10^{18}$ floating point operations / sec) achieved in 2018!

https://upload.wikimedia.org/wikipedia/commons/d/df/PPTExponentialGrowthof_Computing.jpg
AI Computation

• (OpenAI) Since 2012, the amount of computation used in the largest AI training runs has doubled every 3.5 months!
  – Metric = petaflop/sec-days
  – Equal to $10^{15}$ neural net operations per second for one day, or a total of about $10^{20}$ total operations
• By comparison, Moore’s Law has an 18-month doubling period

*AlexNet was a landmark 8-layer CNN (developed by Alex Krizhevsky) that was champion of the ImageNet Large Scale Visual Recognition Challenge in 2012
Some AI / ML “Hits” – during 2018-2019

- Entire human chess knowledge learned and surpassed by AlphaZero in 4 hours
- AI “defeats” humans on a Stanford University reading comprehension test
- Deep Voice/Baidu develops algorithm to mimic voice with just snippets of audio
- Photorealistic speech-driven facial reenactment
- AI system finds correct sequence of steps to synthesize complex organic molecules (a task much more complex than the game of Go)
- Deep learning convolutional NN outperforms human cardiologists in a task involving heart scans
- “Cocktail problem” – AI learns to pick out individual voices in noisy crowd
- ML replicates chaotic attractors and calculates Lyapunov exponents from data
- AI system is trained to “see” in extreme low-light conditions
- AI learns to sense people’s postures and movement through walls with WiFi
- ML recreates periodic table of elements in hours
- One-shot self-supervised imitation learning achieves human-level play (on difficult exploration games Montezuma’s Revenge, Pitfall!, and Private Eye)
- 99.95th percentile ranked human Dota2 team lost 2/3 matches to OpenAI Five
Some AI / ML “Misses” – during 2018-2019

- AI learns to associate colon cancer patients with specific clinics to which they were sent rather than the actual cancer (i.e., bias in electronic medical records)
- Less than stellar performance by an AI-controlled drone pitted against human pilot
- Tesla on autopilot crashes into a Laguna Beach police patrol vehicle
- AI system to predict outcomes of chemical reactions falls short of 90% accuracy goal (achieved 80% even for small “proof-of-concept” caliber size of atoms set)
- Tesla ‘on Autopilot’ slams into parked fire truck on California freeway
- Facial recognition software wrongly identifies 28 lawmakers as crime suspects
- Google’s “Talk to Books” semantic-search offers little improvement over keywords
- ML methods are dominated by traditional statistical ones on a comparative test of forecasting performance (using standard time series benchmarks)
- The six most accurate methods are basic statistical methods, not ML
- IBM Watson reportedly recommended “unsafe and incorrect” cancer treatments
- AI “fails” to predict winner in FIFA World Cup 2018 (an example of negative hype?)
- Driverless Tesla kills autonomous robot
- OpenAI’s “not quite complete win” over Dota2
AI / ML – persistent challenges

“I’m trying to draw a distinction between a machine learning system that’s a black box and an entire field that’s become a black box. Without deep understanding of the basic tools needed to build and train new algorithms, researchers creating AIs resort to hearsay, like medieval alchemists.”

– Ali Rahimi, Google AI

Rahimi’s assertion at NIPS 2017 received a 40 sec ovation

- Intensely data hungry
- “Devil in the details” level of development highly nontrivial
- Basic research concerns – e.g., reproducibility
- Inherently opaque – understandability, explainability, emergence
- Not well integrated with prior knowledge
- Limited “understanding” of context (that humans take for granted)
- Limited capacity for transfer (to other problems / domains)
- Does not easily distinguish causation from correlation
- Struggles with open-ended inference
- Difficulty with exploration games w/sparse rewards (RL methods)
- Lives best in static universes
- Only nascent development of meta-learning and lifelong-learning
- Fragility – vulnerable to attack and/or exploitation
- Fundamental limits on ability to anticipate emergent behaviors
- Deeply prone to the “hype machine”
AI / ML – persistent hype

- “Alibaba's AI software surpasses humans in reading test”
  - News Asia (Jan 2018)

- “Computers are better than humans at reading”
  - CNN (Jan 2018)

- Theory of Mind-net (ToM-net)
  - Google, DeepMind (Feb 2018)

- “Pretty sure Google's new talking AI just beat the Turing test”
  - Engadget (May 2018)

- “Maryland researchers say they discovered 'Holy Grail' of machine learning”
  - Washington Times (May 2018)

- “Scientists Have Invented a Software That Can 'See' Several Minutes Into The Future”
  - ScienceAlert (June 2018)

- “A team of AI algorithms just crushed humans in a complex computer game”
  - Technology Review (June 2018)

- “IBM’s AI Wins Debate with Human – twice”
  - Big Think (June 2018)

- “When bots teach themselves to cheat”
  - Wired (August 2018)

- “Robot 'talks' to MPs about future of AI in classroom”
  - BBC News (October 2018)

- “This clever AI hid data from its creators to cheat at its appointed task”
  - TechCrunch (Dec 2018)

- 'Hi-tech robot' on Russian state television turns out to be man in suit
  - Oddity Central (Dec 2018)
Snapshots of AI as a burgeoning “science” (1/3)

Closing the AI Knowledge Gap

Ziv Epstein, Blakeley R. Payne, Judy Harwen Shen, Abhimanyu Dubey, Rjbarke Felho, Matthew Groth, Nick Obadovich, Manuel Cebrian, Iyad Rahwan
Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
Correspondence: [cebrian, irahwan]@mit.edu

Abstract
AI researchers employ not only the scientific method, but also methodology from mathematics and engineering. However, the use of the scientific method – specifically hypothesis testing – in AI is typically conducted in service of engineering objectives. Growing interest in topics such as fairness and algorithmic bias show that engineering-focused questions only comprise a subset of the important questions about AI systems. This results in the AI Knowledge Gap: the number of unique AI systems grows faster than the number of studies that characterize these systems’ behavior. To close this gap, we argue that the study of AI could benefit from the greater inclusion of researchers who are well positioned to formulate and test hypotheses about the behavior of AI systems. We examine the barriers preventing social and behavioral scientists from conducting such studies. Our diagnosis suggests that accelerating the scientific study of AI systems requires new incentives for academia and industry, motivated by new tools and institutions. To address these needs, we propose a two-sided marketplace called Terngloss. On one side, AI contributors upload existing and novel algorithms to be studied scientifically by others. On the other side, AI researchers develop and post machine intelligence tasks designed to evaluate and characterize algorithmic behavior. We discuss this market’s potential to democratize the scientific study of AI behavior, and thus narrow the AI Knowledge Gap.

1 The Many Facets of AI Research
Although AI is a sub-discipline of computer science, AI researchers do not exclusively use the scientific method in their work. For example, the methods used by early AI researchers often drew from logic, a subfield of mathematics, and are distinct from the scientific method we think of today. Indeed AI has adopted many techniques and approaches over time. In this section, we distinguish and explore the history of these.

ML Arxiv papers per year
(Compiled by Zak Stone, product manager at TensorFlow)

Neural Information Processing Systems (NIPS)
2018 conference sold out in 11 min 38 sec!

https://arxiv.org/abs/1803.07233
Snapshots of AI as a burgeoning “science” (2/3)

- Paper evaluates such performance across multiple forecasting horizons using a large subset of 1045 monthly time series used in M3 Competition
- The six most accurate methods are basic statistical methods, not ML

"It should become clear that ML methods are not a panacea that would automatically improve forecasting accuracy. Their capabilities can easily generate implausible solutions, leading to exaggerated claims of their potentials and must be carefully investigated before any claims can be accepted."

Symmetric Mean Absolute Percentage Error (sMAPE)
Snapshots of AI as a burgeoning “science” (3/3)

- GAN = Generative Adversarial Network, introduced by Ian Goodfellow (et.al.) in 2014
- Consist of two competing networks: a generator (G) and a discriminator (D)
  - G tries to create random synthetic outputs (e.g., images of faces)
  - D tries to tell these apart from real outputs (e.g., a database of celebrities)
- As G and D “compete,” they both get better and better
- The result is a generator network that produces realistic outputs

...the list goes on – and on, and on - for about 250 entries!

Avinash Hindupur, The GAN Zoo, https://deephunt.in/the-gan-zoo-79597dc8c347
When NNs don’t work, they can be unpredictably bad!


When NNs don’t work, they can be *unpredictably bad!*

**Single-Pixel Attacks**

- **Ship** → **Car** (57.7%)
- **Nematode** → **Frog** (99.9%)
- **Gibbon** → **Airplane** (99.3%)


**Adversarial Patch Attacks**

A real-world attack on VGG16, using a physical patch generated by the white-box ensemble method described in Section 3. When a photo of a tabletop with a banana and a notebook (top photograph) is passed through VGG16, the network reports class ‘banana’ with 97% confidence (top plot). If we physically place a sticker targeted to the class “toaster” on the table (bottom photograph), the photograph is classified as a toaster with 99% confidence (bottom plot).


---

**The Elephant in the Room**

When NNs don’t work, they can be *unpredictably bad*!

It’s not just a “single AI solution”

Key functional components and relationships of an autonomous unmanned system, including elements that describe human–machine interaction / collaboration

After figure 3.1 in A. Finn and S. Scheding, Developments and Challenges for Autonomous Unmanned Vehicles, Springer-Verlag, 2010
It’s not just a “single AI solution”

Each component may be associated with (a set of entwined) AI methods

1. Human-Robot Interaction
2. Natural Language Processing
3. Multiagent Systems
4. Internet of Things
5. Social Network Analysis; Crowdsourcing
6. Image Recognition
7. Deep Learning
8. Neural Networks
9. Heuristic Search
10. Machine Perception; Computer Vision
11. Machine Learning
12. Goal-directed Behavior
13. Expert Systems
14. Knowledge Representation; Reasoning
15. Mobile Robotics
16. Robotic Swarms
17. Collaborative Systems

After figure 3.1 in A. Finn and S. Scheding, Developments and Challenges for Autonomous Unmanned Vehicles, Springer-Verlag, 2010
The Manifold Hypothesis

• Natural data forms lower dimensional structures (manifolds) in embedding space
  – *Each manifold represents a different entity*

• Learning (“understanding” data) achieved by separating the manifolds
  – *Easy to do (and visualize) when D = 2*
    (“Stretching and squashing”)
  – *Much harder when D >> 1*
What is AI?

A panel discussion on the opportunities and challenges presented by artificial intelligence

22 January 2019

Questions?
Comical Views of Machine Learning & Super-AI

---

Xkcd Webcomic, https://xkcd.com/1838/; Creative Commons License

Xkcd Webcomic, https://xkcd.com/1450/; Creative Commons License: https://creativecommons.org/licenses/by-nc/3.0/us/

Abtruse Goose, https://abstrusegoose.com/594; Creative Commons License