Overview

• Explaining recent AI achievements: two case studies:
  • Alpha* (Go playing systems)
  • Arcade Learning Environment (Atari playing systems)
• The challenge of forecasting
Recent Developments in AI

- Self-driving cars
- AI in education
- AI in gaming
- AI in healthcare
Better Performance, But at What Cost?

- Examples of costs incurred by developers, adapters, or users of AI systems in order to achieve a given level of performance include (Martinez-Plumed et al., 2018):
  - Data
  - Knowledge
  - Software
  - Hardware
  - Manipulation
  - Computation
  - Networking
  - Time
Two Case Studies

- Alpha*
- Arcade Learning Environment (ALE)
Alpha*

- Surpassed human performance in 2016 and continued improvement thereafter

Silver et al., 2017
• What accounts for this progress?
• Some algorithmic improvements
• Also: submitting one kind of cost (computing power) for another (human data)

Amodei and Hernandez, 2018
Arcade Learning Environment (ALE)

- Steady algorithmic progress over time
- Higher (better) performance curves are generally more recent

Hessel et al., 2017

Figure 1: **Median human-normalized performance** across 57 Atari games. We compare our integrated agent (rainbow-colored) to DQN (grey) and six published baselines. Note that we match DQN’s best performance after 7M frames, surpass any baseline within 44M frames, and reach substantially improved final performance. Curves are smoothed with a moving average over 5 points.
ALE (Atari) circa early this year

- Algorithmic changes have boosted performance, but...

Martinez-Plumed et al., 2018
ALE (Atari) today

Ape-X DQfD - much more compute and leveraging of (a few) human demonstrations

X-axis is very approximate; adapted from Martinez-Plumed et al., 2018
The Challenge of Forecasting

- Key inputs aren’t always reported, making principled trend extrapolation difficult
- Falsifiable predictions are rarely made
  - We don’t know who knows what, if anything
- Expert opinion is all over the place
The Challenge of Forecasting

- Key inputs aren’t always reported

Martinez-Plumed et al., 2018
The Challenge of Forecasting

- Falsifiable forecasts are difficult:
  - Evaluation standards and challenge tasks are constantly changing
  - Do you control for compute, data, etc. or not?
- And rare
The Challenge of Forecasting

Grace et al., 2017
What should we expect in the near future?

• Peak performance will depend in large part on continued hardware advances and algorithmic advances that can leverage these effectively
  • Otherwise financial costs will grow greater over time to achieve blockbuster results

• Broad societal deployment will depend on:
  • Reducing hardware/data costs
  • Increasing robustness (reducing need for human oversight, another form of cost)

• Greater impacts may be had in domains where key inputs are cheap (e.g. good simulators, labeled data, human demonstrations)
Thanks!

• Miles.Brundage@philosophy.ox.ac.uk