Reinforcement Learning in Public Policy

Peter Henderson
Stanford Computer Science & Stanford Law School

Reinforcement Learning in Public Policy

Peter Henderson
Stanford Computer Science & Stanford Law School

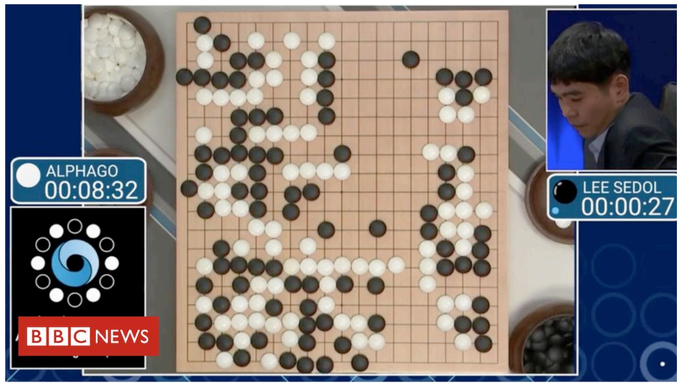
Presentation based loosely on work with a number of wonderful collaborators:

Henderson*, Chugg*, Anderson, and Ho. "Beyond Ads: Sequential Decision-Making Algorithms in Public Policy." *arXiv preprint arXiv:2112.06833* (2021).

And also forthcoming work with Chugg, Anderson, Altenburger, Turk, Guyton, Goldin, and Ho.

All views and opinions expressed in this presentation are my own and not of any of my co-authors, nor of the Internal Revenue Service or any other company or government entity.

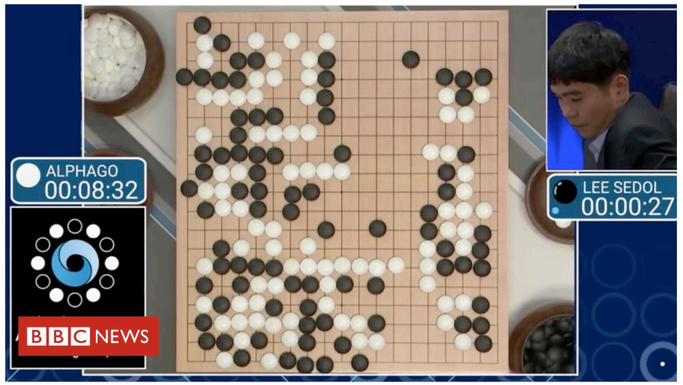
Reinforcement Learning Successes



https://www.bbc.com/news/technology-35785875

Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961

Reinforcement Learning Successes



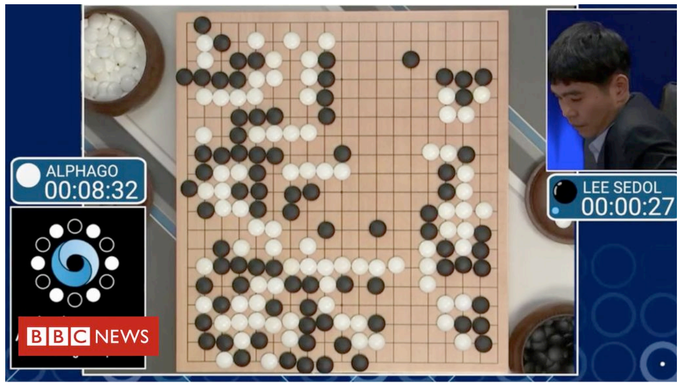
https://www.bbc.com/news/technology-35785875

Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961

https://www.wired.com/story/new-ai-based-navigation-helps-loons-balloons-hover-in-place/

Bellemare, M.G., Candido, S., Castro, P.S. et al. Autonomous navigation of stratospheric balloons using reinforcement learning. Nature 588, 77–82 (2020). https://doi.org/10.1038/s41586-020-2939-8

Reinforcement Learning Successes

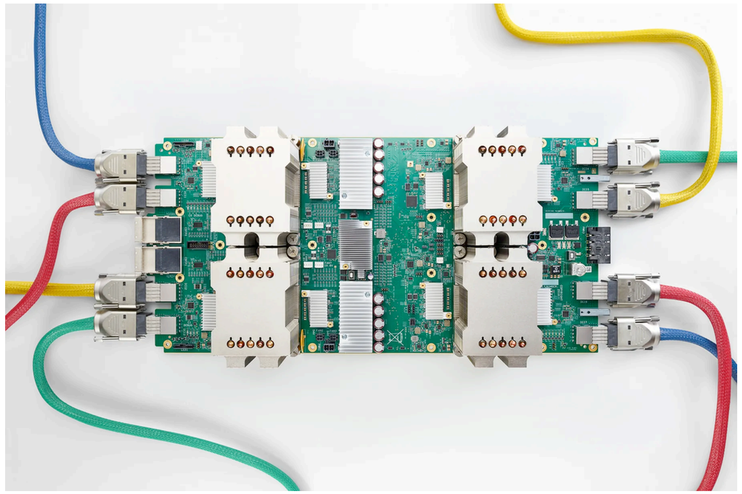


https://www.bbc.com/news/technology-35785875

Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961

https://www.wired.com/story/new-ai-based-navigation-helps-loons-balloons-hover-in-place/

Bellemare, M.G., Candido, S., Castro, P.S. et al. Autonomous navigation of stratospheric balloons using reinforcement learning. Nature 588, 77–82 (2020). https://doi.org/10.1038/s41586-020-2939-8



https://www.theverge.com/2021/6/10/22527476/google-machine-learning-chip-design-tpu-floorplanning

Mirhoseini, A., Goldie, A., Yazgan, M. et al. A graph placement methodology for fast chip design. Nature 594, 207–212 (2021). https://doi.org/10.1038/s41586-021-03544-w

Reinforcement Learning Successes

Military AI vanquishes human fighter pilot in F-16 simulation. How scared should we be?

Artificial intelligence can master difficult combat skills at warp speed, but the Pentagon's futurists must remain mindful of its limitations and risks.

> All the more remarkable, Heron's AI pilot was self-taught using deep reinforcement learning, a method in which an AI runs a combat simulation over and over again and is "rewarded" for rapidly successful behaviors and "punished" for failure. Initially, the AI agent is simply learning not to fly its aircraft into the ground. But after 4 billion iterations, Heron seems to have mastered the art of executing energy-efficient air combat maneuvers.

https://www.nbcnews.com/think/opinion/military-ai-vanquishes-human-fighter-pilot-f-16-simulation-how-ncna1238773

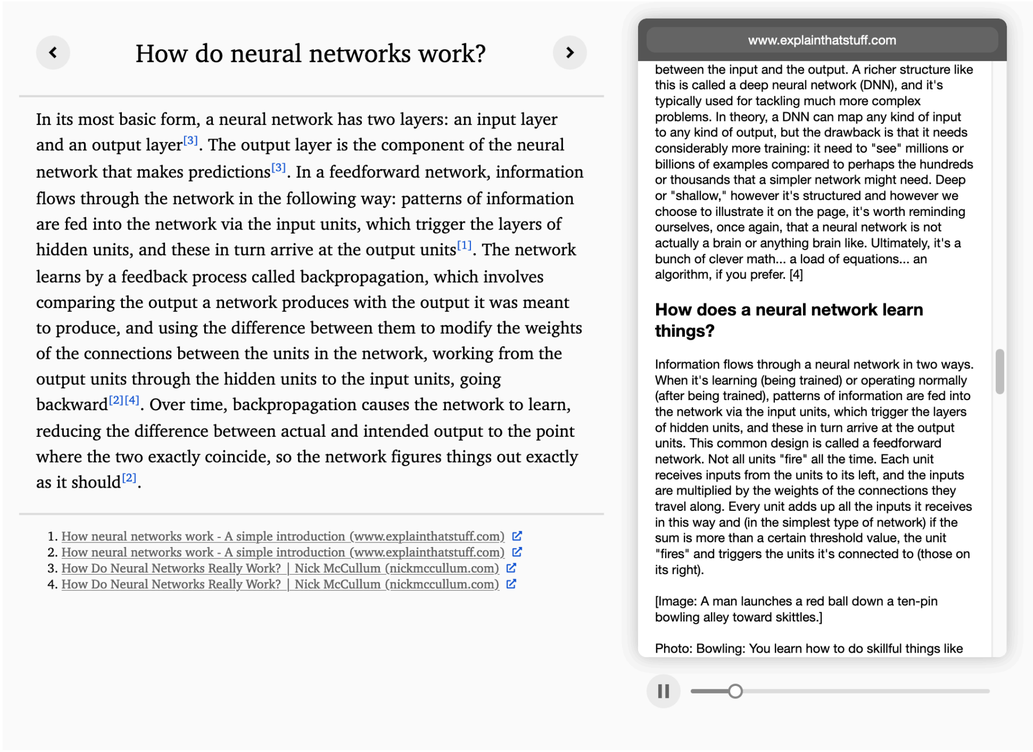
Reinforcement Learning Successes

Military AI vanquishes human fighter pilot in F-16 simulation. How scared should we be?

Artificial intelligence can master difficult combat skills at warp speed, but the Pentagon's futurists must remain mindful of its limitations and risks.

All the more remarkable, Heron's AI pilot was self-taught using deep reinforcement learning, a method in which an AI runs a combat simulation over and over again and is "rewarded" for rapidly successful behaviors and "punished" for failure. Initially, the AI agent is simply learning not to fly its aircraft into the ground. But after 4 billion iterations, Heron seems to have mastered the art of executing energy-efficient air combat maneuvers.

https://www.nbcnews.com/think/opinion/military-ai-vanquishes-human-fighter-pilot-f-16-simulation-how-ncna1238773



Reinforcement learning on 175B parameter models
https://openai.com/blog/webgpt/

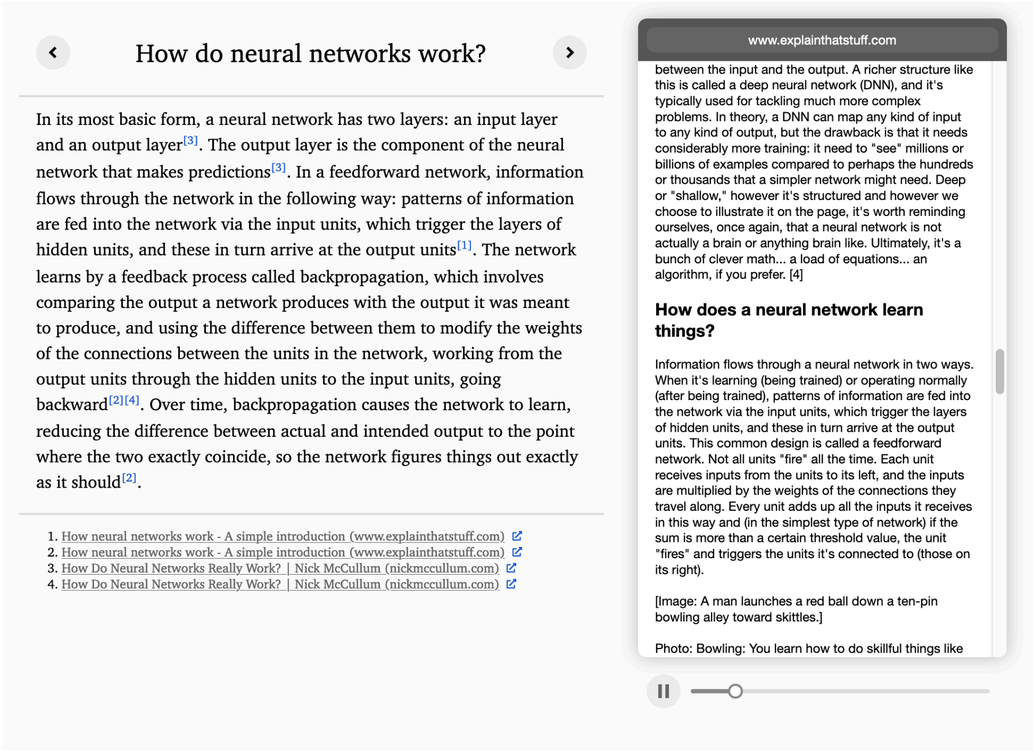
Reinforcement Learning Successes

Military AI vanquishes human fighter pilot in F-16 simulation. How scared should we be?

Artificial intelligence can master difficult combat skills at warp speed, but the Pentagon's futurists must remain mindful of its limitations and risks.

All the more remarkable, Heron's AI pilot was self-taught using deep reinforcement learning, a method in which an AI runs a combat simulation over and over again and is "rewarded" for rapidly successful behaviors and "punished" for failure. Initially, the AI agent is simply learning not to fly its aircraft into the ground. But after 4 billion iterations, Heron seems to have mastered the art of executing energy-efficient air combat maneuvers.

https://www.nbcnews.com/think/opinion/military-ai-vanquishes-human-fighter-pilot-f-16-simulation-how-ncna1238773



Reinforcement learning on 175B parameter models
https://openai.com/blog/webgpt/

We'll come back to this one.

Government has launched many AI investment projects

BRIEFING ROOM

The Biden Administration Launches AI.gov Aimed at Broadening Access to Federal Artificial Intelligence Innovation Efforts, Encouraging Innovators of Tomorrow

MAY 05, 2021 • PRESS RELEASES

Government has launched many AI investment projects

BRIEFING ROOM

The Biden Administration Launches AI.gov Aimed at Broadening Access to Federal Artificial Intelligence Innovation Efforts, Encouraging Innovators of Tomorrow

MAY 05, 2021 • PRESS RELEASES

AI Weekly: U.S. agencies are increasing their AI investments

Government has launched many AI investment projects


BRIEFING ROOM

The Biden Administration Launches AI.gov Aimed at Broadening Access to Federal Artificial Intelligence Innovation Efforts, Encouraging Innovators of Tomorrow

MAY 05, 2021 • PRESS RELEASES

AI Weekly: U.S. agencies are increasing their AI investments

Artificial Intelligence:
An Accountability Framework for Federal Agencies and Other Entities

Government has launched many AI investment projects

BRIEFING ROOM

The Biden Administration Launches AI.gov Aimed at Broadening Access to Federal Artificial Intelligence Innovation Efforts, Encouraging Innovators of Tomorrow

MAY 05, 2021 • PRESS RELEASES

AI Weekly: U.S. agencies are increasing their AI investments

Artificial Intelligence:
An Accountability Framework for Federal Agencies and Other Entities

NETWORKS / CYBER

JAIC developing first-of-its-kind 'integration layer' for AI algorithms

"So if you don't have an integration layer like that, then you have to go find all of the data sources yourself... This is something we don't have yet in the department," Lt. Gen. Michael Groen told Breaking Defense.

By JASPREET GILL on February 09, 2022 at 11:13 AM

Government has launched many AI investment projects

BRIEFING ROOM

The Biden Administration Launches AI.gov Aimed at Broadening Access to Federal Artificial Intelligence Innovation Efforts, Encouraging Innovators of Tomorrow

MAY 05, 2021 • PRESS RELEASES

AI Weekly: U.S. agencies are increasing their AI investments

Artificial Intelligence:
An Accountability Framework for Federal Agencies and Other Entities

How The Federal Government's AI Center Of Excellence Is Impacting Government-Wide Adoption Of AI

Kathleen Walch Contributor
COGNITIVE WORLD Contributor Group
AI

NETWORKS / CYBER

JAIC developing first-of-its-kind 'integration layer' for AI algorithms

"So if you don't have an integration layer like that, then you have to go find all of the data sources yourself... This is something we don't have yet in the department," Lt. Gen. Michael Groen told Breaking Defense.

By JASPREET GILL on February 09, 2022 at 11:13 AM

Where does RL fall into government or public policy deployments?

What makes for a good RL and Public Policy problem?

What are the areas of public policy & government services that can be improved by RL?

What are the unique research challenges encountered in RL for public policy?

Can we verify RL well enough to feel confident in deployments? Is deployment of RL in these settings a net social good?

What makes for a good RL and Public Policy problem?

Recall that RL maximizes future reward as part of a Markov Decision Process.

1. Your environment changes over time.
2. You need to both explore and exploit.
3. You (maybe) need to plan ahead.*

* For the purposes of this talk, we will consider contextual bandits as in-scope as they can be formulated as an RL problem with discount 0.

What are the areas of public policy & government services that can be improved by RL?

Enforcement

Environmental Protection Agency audits Concentrated Animal Feeding Operations for compliance
[Chugg, Anderson, Eicher, Lee, & Ho 2021]

Internal Revenue Service audits taxpayers for compliance
[Henderson et al., 2022]

Brazilian government audits regional governments for corruption
[Ash, Galletta, & Giommoni, 2021]

Optimal Policymaking

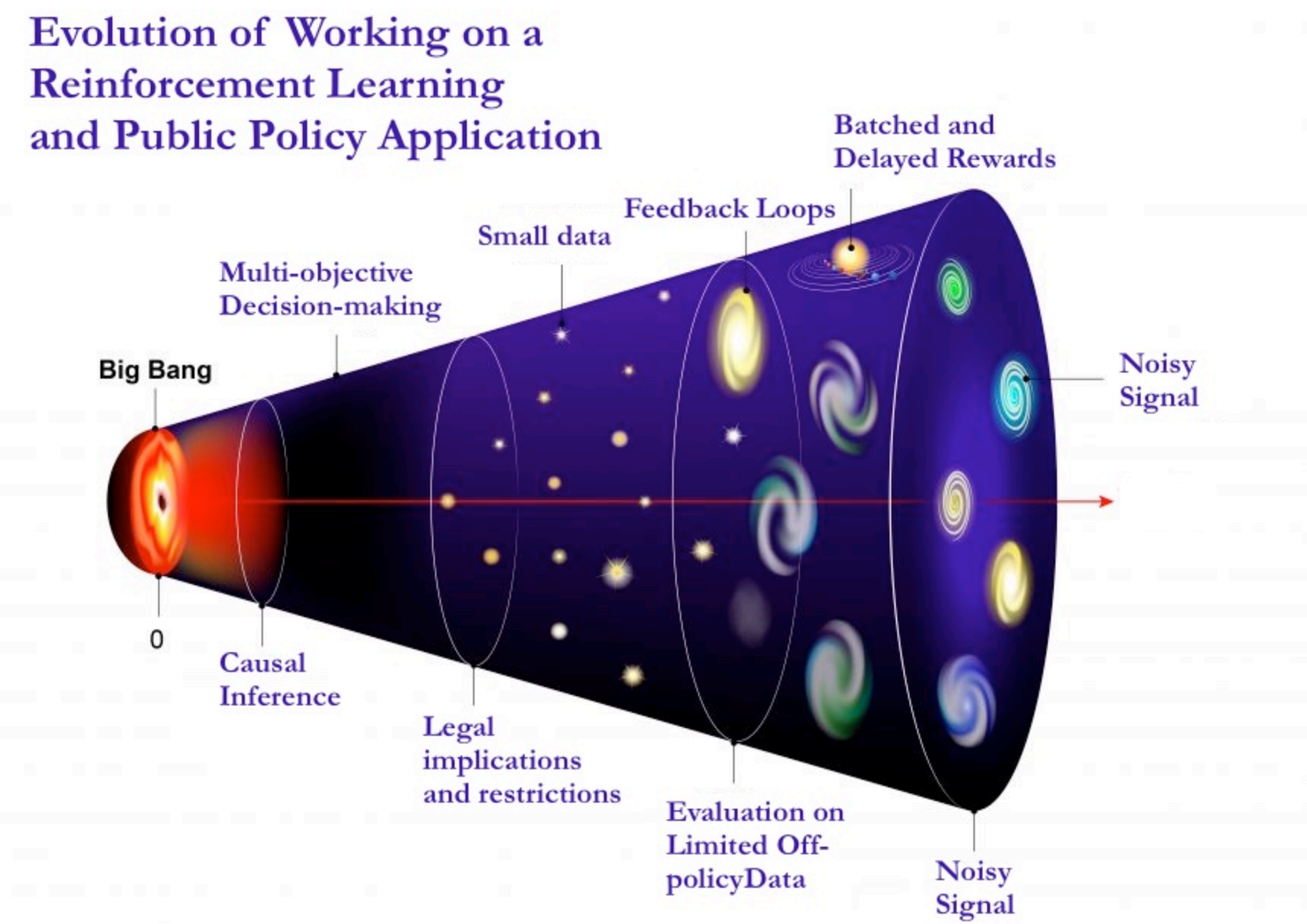
Optimal Tax Policy
[Zheng & Trott et al., 2020]

Resource Allocation

Covid-19 test allocation
[Chugg et al., 2021; Bastani et al., 2021]

Delaware Department of Transportation plans to test RL for traffic control
[Gettman, 2019]

What are the difficult research challenges encountered in RL for public policy?



What are the difficult research challenges encountered in RL for public policy?

Multi-objective decision-making

In enforcement, many agencies must both prioritize where to target enforcement to maximize misreporting (reward) and they are required by law to give statistically valid estimates of overall misreporting (estimation).

The Improper Payments Information Act of 2002 (IPIA)
Improper Payments Elimination and Recovery Act of 2010 (IPERA)
Improper Payments Elimination and Recovery Improvement Act of 2012 (IPERIA)

A case study of unique challenges: Internal Revenue Service Audit Selection

The presenter is a detailed IRS employee working with the IRS under a student volunteer agreement. All data work for this project involving confidential taxpayer information was done on IRS computers by IRS employees. At no time was confidential taxpayer data ever outside of the IRS computing environment. The views and opinions presented in this presentation reflect those of the author and do not necessarily reflect the views or the official position of the Internal Revenue Service.

A case study of unique challenges: Internal Revenue Service Audit Selection

Tax gap estimate (difference between paid and true owed taxes) is **$441 billion per year.**

Some estimate that investment into information technology for identifying misreporting could yield **10:1 returns on investment**. *See* Sarin and Summers (2019).

Sources

https://www.irs.gov/newsroom/the-tax-gap

Sarin, Natasha, and Lawrence H. Summers. Shrinking the tax gap: approaches and revenue potential. No. w26475. National Bureau of Economic Research, 2019.

A case study of unique challenges: Internal Revenue Service Audit Selection

Current Process (*simplified*)

National Research Program

Stratified Random Sample
~15k audits per year

Estimates to meet reporting requirements.

Risk Model & other selection methods

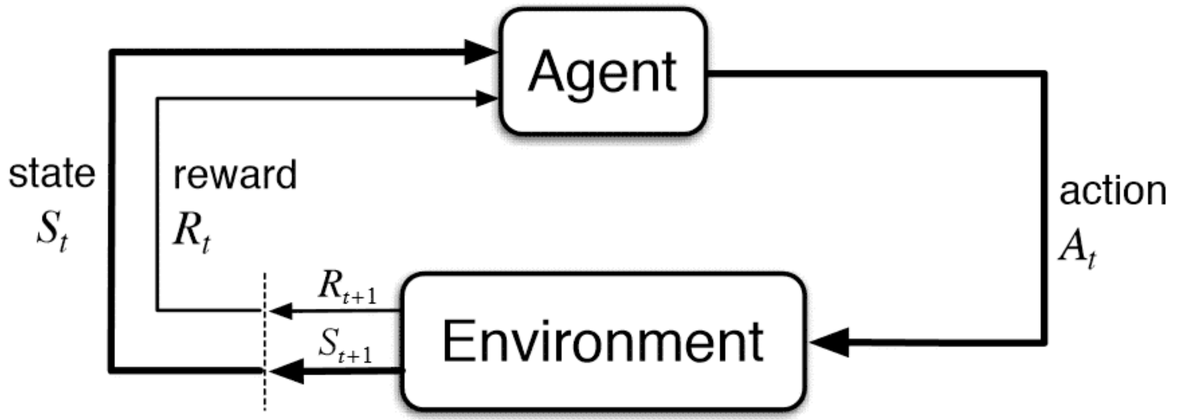
Op Audit Selection

Risk-selected
+ Other Selection Mechanisms

>>100k audits per year

* Note there are a number of other methods that go into current Op Audit selection and tax gap estimation.

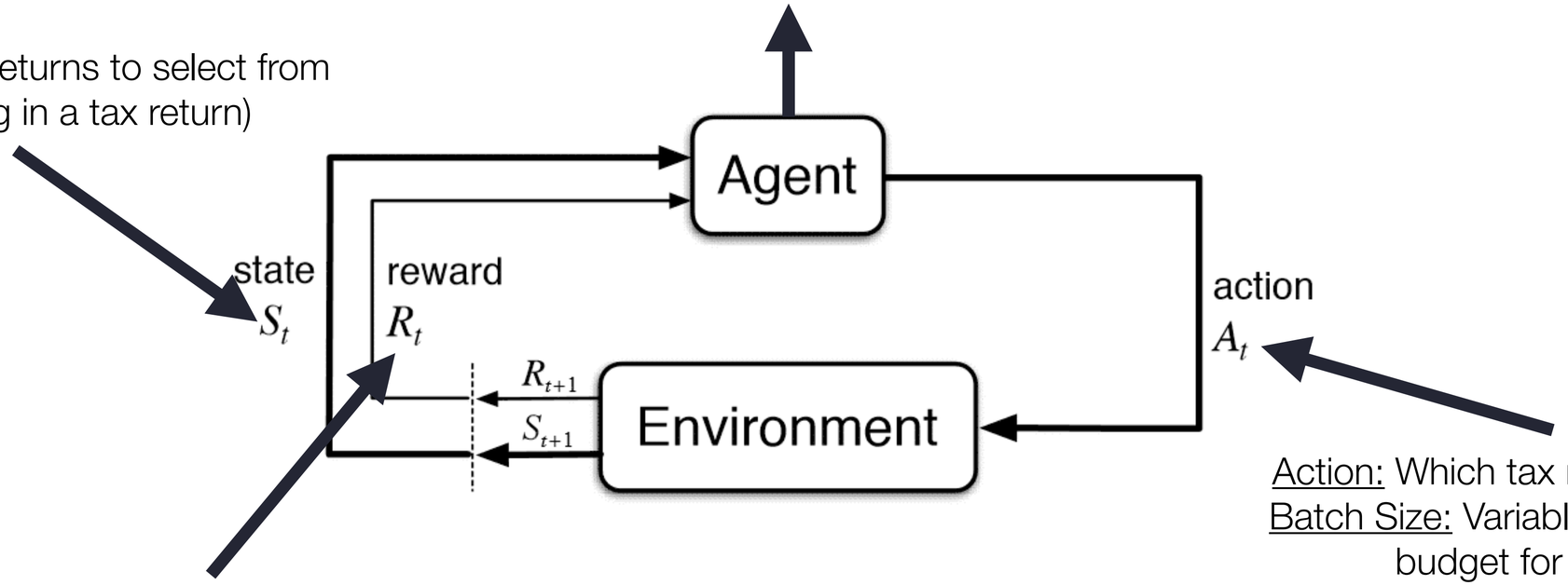
A case study of unique challenges: Internal Revenue Service Audit Selection



A case study of unique challenges: Internal Revenue Service Audit Selection

Output: Misreporting Estimate

State: Population of tax returns to select from
(Features: everything in a tax return)



Action: Which tax returns to audit next
Batch Size: Variable, depends on audit budget for a given year

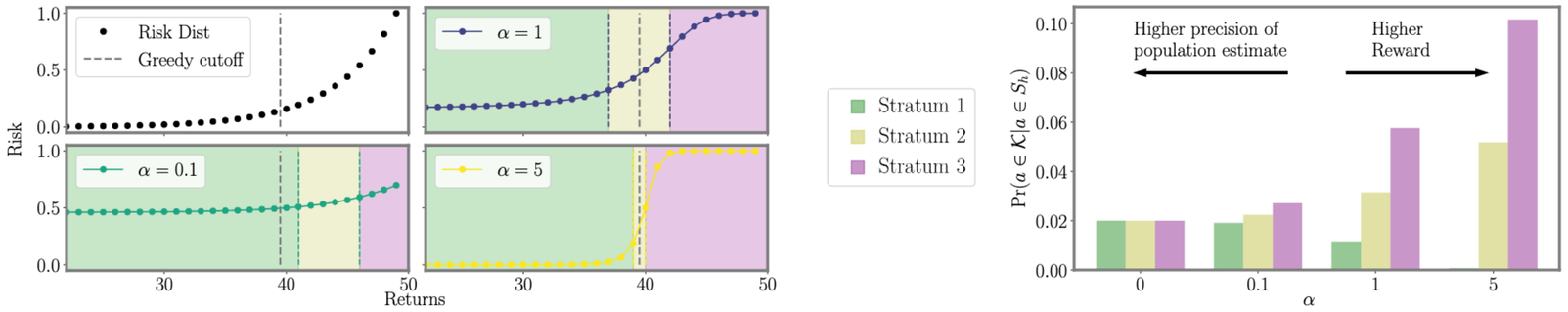
Reward: Amount of adjustment to taxes owed after audit
Delay: Usually 1-2 year delay

Temporal Drift
Tax code changes
Behavioral adjustments (less likely to misreport if someone in network recently audited)
New evasion mechanisms
Covariate Drift (structural changes to the economy)

A case study of unique challenges: Internal Revenue Service Audit Selection

Problem: Population estimation requires reward-sub-optimal exploration. Exacerbates the explore-exploit trade-off problem.

Solution: Planning required to optimally re-use information to achieve robust population estimates.



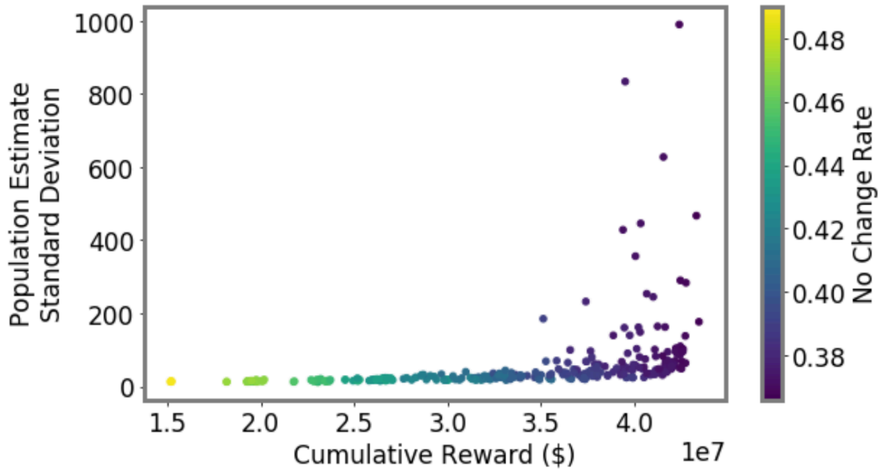
(a) Risk distribution and parameterizations

(b) Probability of sampling an individual

A case study of unique challenges: Internal Revenue Service Audit Selection

We find some **bias-variance-reward trade-off** between estimation and reward maximization objectives.

Future research: reduce the impact of this trade-off.



Done with IRS Case Study and Associated Work

What are the difficult research challenges encountered in RL for public policy?

Public Policy & Government Applications

Interesting Reinforcement Learning Research Challenges

Can we verify RL well enough to feel confident in deployments?

1. Formalization is better than status quo in many cases. Can identify biases/problems and can adjust in iterative fashion.
2. **But** not if evaluation is bad and gives false sense of security or if there are dramatic failure modes.

Can we verify RL well enough to feel confident in deployments?

Famous example is the use of 7 positive labels with leave-one-out validation to train/evaluate a random forest for detecting “couriers” as part of the SKYNET program at the NSA.

https://arstechnica.com/information-technology/2016/02/the-nsas-skynet-program-may-be-killing-thousands-of-innocent-people/

Can we verify RL well enough to feel confident in deployments?

Safety

Safety fallbacks when not functioning within acceptable parameters
[Gramble & Gao, 2018]

Policy performance certificates
[Dann et al., 2018]

Bias/Fairness Audits

Often not thought about in the context of RL, but important for real-world deployment.
[EEOC, https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness]

Realistic Evaluation*

Across distributions of data simulating different populations/conditions
[Henderson & Islam et al., 2017; Bouthillier et al., 2021; Agarwal et al., 2021]

Statistically well-powered evaluation
[Colas et al., 2018; Card et al., 2021]

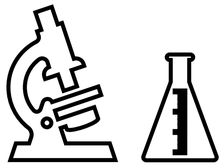
Causality Checks

Is the agent making rational decisions?

* Robust evaluation already a common requirement across some, but not agencies/applications. See, e.g., NIST Facial Recognition Test framework.

Can we verify RL well enough to feel confident in deployments?

Better validation



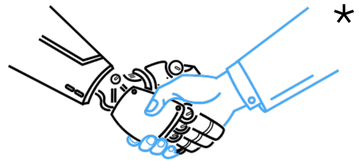
Trust & Confidence

Equity & Safety

True Efficiency

Appropriate Failsafe Mechanisms and Red Lights

Better public policy & government services



* Image from https://www.causalens.com/trustworthy-ai/

BUT have to stop and think about what you're working on.

How can the RL system fail in a way that makes things socially worse than the existing human system?

Military AI vanquishes human fighter pilot in F-16 simulation. How scared should we be?

Artificial intelligence can master difficult combat skills at warp speed, but the Pentagon's futurists must remain mindful of its limitations and risks.

All the more remarkable, Heron's AI pilot was self-taught using deep reinforcement learning, a method in which an AI runs a combat simulation over and over again and is "rewarded" for rapidly successful behaviors and "punished" for failure. Initially, the AI agent is simply learning not to fly its aircraft into the ground. But after 4 billion iterations, Heron seems to have mastered the art of executing energy-efficient air combat maneuvers.

Need to think hard about some deployments where mistakes cost lives.

Thank you!
Feel free to reach out with questions and comments.