LECTURE 7: Algorithms Beating Us in Complex Games: Reinforcement Learning (Feb 18)

You Will Lose Your Job to a Robot—and Sooner Than You Think Automation helped bring on the age of Trump. What will AI bring? Mother Jones Nov/Dec 2017 issue … **Robot automation will 'take 800 million jobs by 2030'** 29 November 2017 BBC referring to McKinsey Global Institute Study … **Will Robots Take Our Children's Jobs?** - New York Times Dec 11, 2017 … **Robots: Is your job at risk?** - Sep. 15, 2017 - CNN Money

Every day or so at least one headline that tells you jobs are going to robots/Algorithms/AI. Accounting jobs, finance jobs, barista jobs, driving, surgery, etc etc. Behind these headlines are improvements in robotics and in AI,

1997 Deep Blue beat Chess champion Kasparov
2011 Watson defeated human Jeopardy champions
2016 Google's Alpha Go defeated Korean Go master Lee Sedol
2016 Carnegie Mellon's Libratus beat top poker pros
2016 University of Alberta's Deep Stack wins No Limit Texas Hold'em Poker Tournament
2017 AlphaZero learns Chess and beats chess programs in weekend; triumphs over Go, Chess, Shogi.
Dec 2018 Science. **One program to rule them all: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play** … (Deep Mind Team) The game of chess is the longest-studied domain in the history of artificial intelligence. The strongest programs are based on a combination of sophisticated search techniques, domain-specific adaptations, and handcrafted evaluation functions that have been refined by human experts over several decades. By contrast, the AlphaGo Zero program recently achieved superhuman performance in the game of Go by reinforcement learning from self-play...we generalize this approach into a single AlphaZero algorithm that can achieve superhuman performance in many challenging games. Starting from random play and given no domain knowledge except the game rules, AlphaZero convincingly defeated a world champion program in the games of chess and shogi (Japanese chess), as well as Go.

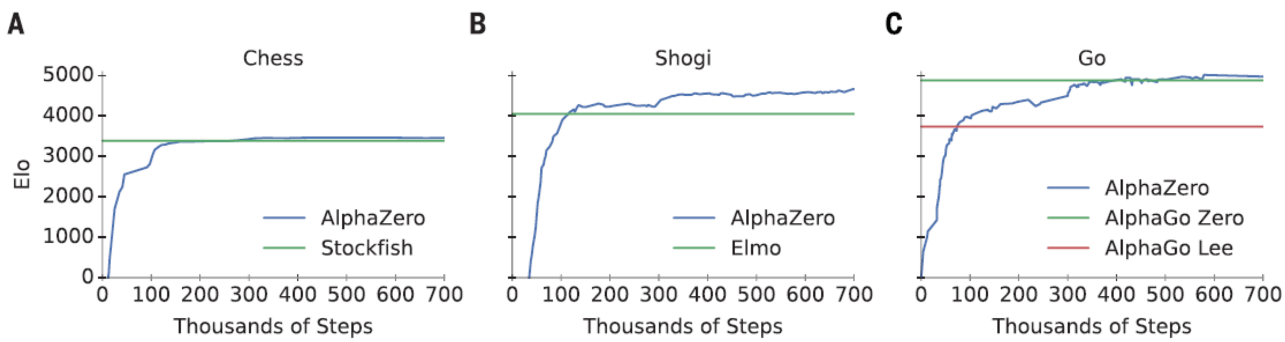


Fig. 1. Training AlphaZero for 700,000 steps. Elo ratings were computed from games between different players where each player was given 1 s per move. (**A**) Performance of AlphaZero in chess compared with the 2016 TCEC world champion program Stockfish. (**B**) Performance of AlphaZero in shogi compared with the 2017 CSA world champion program Elmo. (**C**) Performance of AlphaZero in Go compared with AlphaGo Lee and AlphaGo Zero (20 blocks over 3 days).

Survey of 352 AI Experts Who Published at 2015 NIPS, ICML Conferences https://arxiv.org/abs/1705.08807, May 30, 2017

Reinforcement learning by machines.

How did machines get this “smart” so quick? Mimicking natural selection/artificial selection in digital world at speeds/#s of generations and probably deterministic accuracy than Evolution. By practicing more than any human could possibly do!

A reinforcement Learning (RL) is to learn a good strategy for the agent from experimental trials and relative simple feedback received. With the optimal strategy, the agent is capable to actively adapt to the environment to maximize future rewards.

Some experts tell us that there is no reason to worry. The AI robots are highly specialized and lack the general intelligence of humans. You or I can beat the Go AI in Chess and Poker, the Poker AI In Go and Chess etc. Economists note that for all the media hype, productivity is increasing slowly, employment is high – there is no sign that robots/AI are disrupting labor markets except in the media.

But if Alpha Zero doesn't an ensemble of AI robots/ algorithms could potentially do any job better we can do, the specialization argument falls flat. **And** Elon Musk is worried enough to create a company to implant chips into our brains so we can compete … as cyborgs. Bill Gates proposed a robot tax. Xxx is worried. Should you worry?

What, me worry? About work! You are MAD. I have funner things to do. If this time is different, it's the baby's problem, not mine. Google Alpha46 will be out when you hit the job market, kiddo. Good luck!

Alpha 46? I hear it's not only smarter than me and works harder than me, it never cries or makes dirty diapers, and is cuter than me. What am I going to do?

What does economics tell us? Comparative advantage says direct worries not at 800 million jobs going but at effect of AI robots on wages. "The Rock" may be better at everything than you, but he is relatively better as wrestler/movie star than at other tasks. Even though he is better at your job than you, it pays for him to stick to car chases/beating villains than to do your work. Similarly, Alpha46 may beat you at everything but you will be able to do some things more cost efficiently than the machines.

The question is what is our comparative advantage and at what pay? Commenting on the AlphaGo machine, a professor of computational neuroscience a Sheffield, UK, noted that while computers beat humans at complex calculations and precision, "AI fails in tasks that are surprisingly easy for humans … such as walking, running and kicking a ball". When Alphago beat Lee Sedol, the algorithm had human help: a hand to move the pieces on the go board. Why? Too expensive to build a robot hand.

Historically, machines have had comparative advantage in physical labor. Will they gain comp adv in brain-power work? Today where so much brain work is done digitally, where algorithms live, it is highly possible. To the extent that AI robots become increasingly good substitutes for humans in multiple tasks, the elasticity of substitution between AI robots and people will increase, perhaps to point where we will view robots not as machines that we add to capital stock, K, but as substitutes for human labor, L, producing a Malthusian outcome due not to babies but to robots. As technological improvement will assure that the cost of production of robots (adjusted for quality) continually falls, this will squeeze wages. The gainers? Owners of AI robots: "Who owns the robots rules the world"

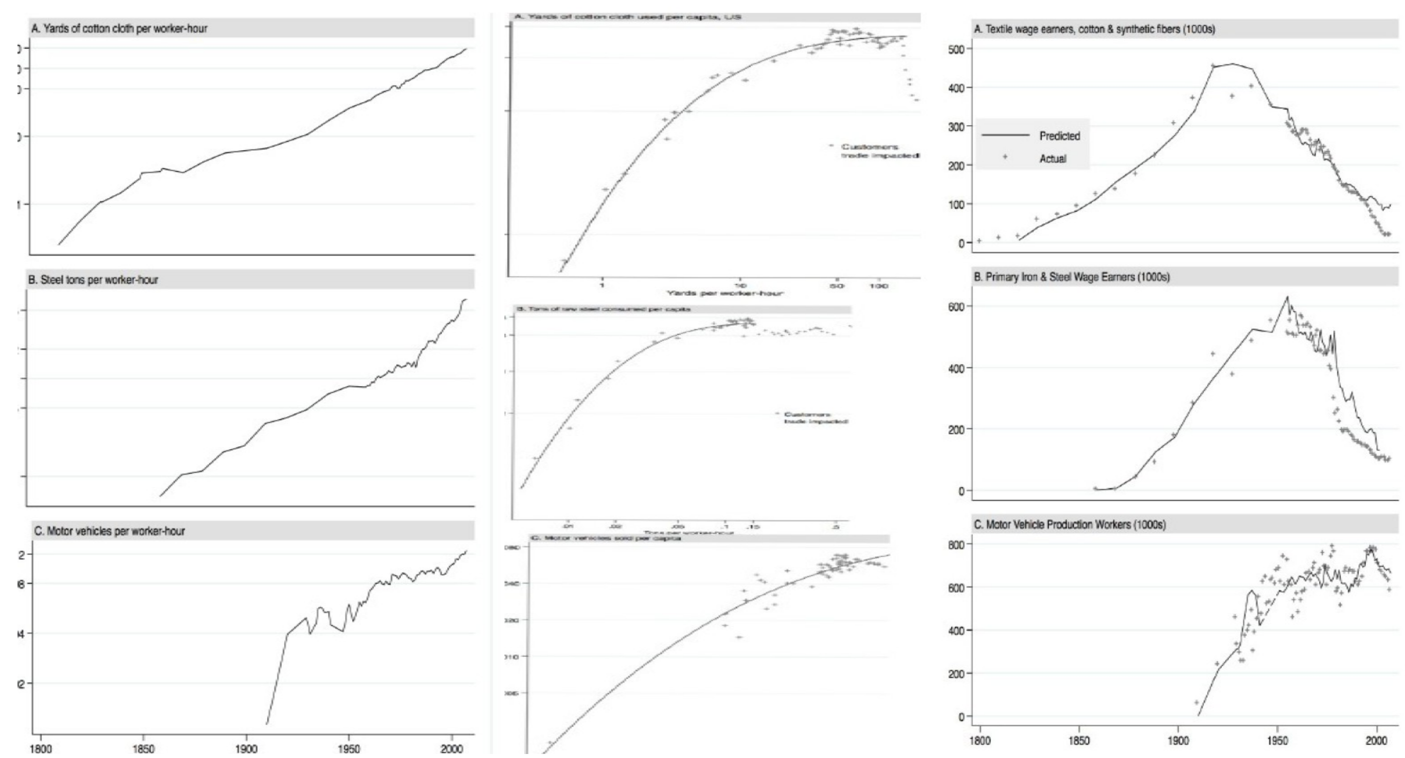
Here are reasons to believe this scenario is vastly exaggerated.

1)The AI expert claim that AI will outperform humans in every job is based on technological possibility. While analysts asked about costs of production of machines, they ignored the organizational costs of deploying them in firms that have data and operations in old technologies. UK NHS is world's biggest purchaser of fax machines – an outmoded technology –because it has lots of paper records and 160 different computer systems in the typical NHS trust/hospital system.

2)All previous automation scares fizzled out. FDR blamed Great Depression joblessness on failure to "employ the surplus of our labor which the efficiency of our industrial processes has created". Early 1960s automation fears led to US Commission on Automation, followed by rise in E/Pop; 1990s Rifkin's End of Work (1995) preceded dot.com boom. Occupations with lots of computer use did not see decline in jobs. Should this time be different? Current big tech revolution inconsistent with modest productivity growth and low unemployment). However technology changes, there is enough work to In a well-functioning market economy both labor-saving and capital-saving technology would benefit workers and dismissed fears of mass joblessness due to robots doing all the work. "The world's problems in this generation and the next are problems of scarcity, not of intolerable abundance." (H.Simon) "The bogeyman of automation consumes worrying capacity that should be saved for real problems."

3)Technologies that reduce costs of production typically lower prices (per unit of constant value) of goods/services which increases demand for goods and services and demands more human as well as machine labor. Consider drop in sequencing human genome from \$100M to below \$1k. This created jobs to make machines and will spur personalized medicine, which could massively increase demand for health output and human jobs (until someone develops the AI robots to do the human tasks). Consider drop in sequencing human genome from \$100M to below \$1k. This created jobs to make machines and will spur on personalized medicine, which could massively increase demand for health output and human jobs (until someone develops the AI robots to do the human tasks). In past technical revolutions, demand for output created new and better jobs. Decline of agriculture --> rise of mfg jobs. Decline of mfg → rise of retail and service sector jobs and white collar work. James Bessen's analysis of three past technologies – growth as demand increases, job loss occurs later.

Productivity Change Expansion of Production Jobs Increase then fall



4)Ransomeware and other forms of malware that exploit the new technologies will require humans to monitor the system,, delaying implementation of full automation. Driverless cars have to learn how to tell a stop sign with some paint splashed on it and to outsmart hacker terrorists trying to sabotage "safe driving". Adversarial machine learning algorithms built by hackers/villains to sabotage digital work are growing Boston, Feb 14, 2020 **1 in 6 Mass Communities Hit** by 'Ransomware' Attacks. Many towns recovered files from backups, but at least 10 handed over taxpayer money to hackers to unlock their data, records obtained by NBC10 Boston Investigators show.

5)Expansionary macro-policy. Cost-saving tech that reduces inflation widens scope for macro.

BUT there are reasons to worry more:

1)Technology advance has just begun. More S&E folk working on AI/robotics than ever before. China commits to huge AI expansion in next 10 years with goal of AI leadership by 2030. Eric Schmit said: "[e]very two days now we create as much information as we did from the dawn of civilization up until 2003".

2)Digital is general technology and learning algorithms are "this time different" than earlier computing algorithms. New learning algorithms cut deep into routine job tasks done in many jobs and readily learned by most people. "Discontinuous change" – almost never correctly predicted by social scientists who look at past data because … hard to foretell until it happens. We did not foresee collapse of finance in 2007-2008 nor recovery of banking etc.

The lawyer algorithm– UK contest pitched over 100 lawyers from London's ritziest firms against an artificial intelligence program called Case Cruncher Alpha, developed by 4 Cambridge law students. … given the basic facts of hundreds of PPI (payment protection insurance) mis-selling cases , CCAlpha beat lawyers in predicting whether the Financial Ombudsman would allow a claim. With 775 predictions ,,, the computer won hands down, with Case Cruncher's accuracy rate of 86.6%, much above the 66.3% for the lawyers.. .the question at *this early stage* of AI development is whether it will "remain limited to descriptive analysis or whether it will be capable of evaluating rules and events", and then whether it will be a **tool for junior lawyers to use or something which replaces them**....

3)Already on a dangerous inequality-increasing trajectory, with labor's share of output falling and differentials among workers increasing, widening differentials in productivity and earnings between top firms/establishments and lower level firms/establishments, AI Robots could accelerate rise of inequality by gaining more for capital and those working in the winning firms.

4)Glimmers of change in the data:

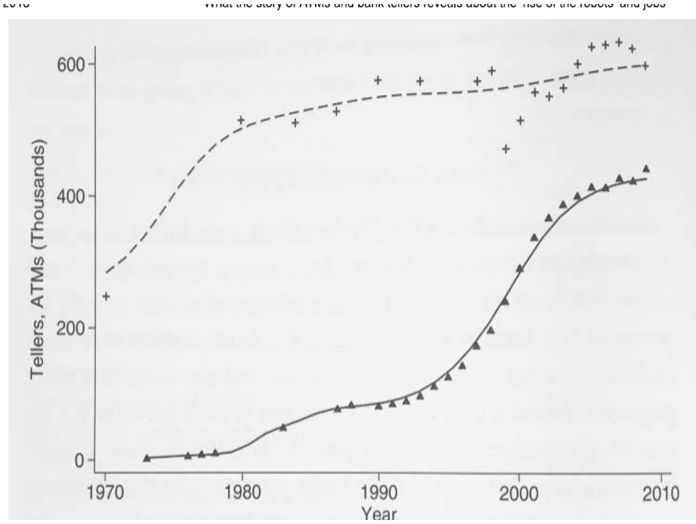
Table 4. IV Estimates of the Impact of Robots and Immigration on the log Hourly Earnings, 2004-2016

	All Workers, by Education Group				Gender	
	All workers	Less than 12 years of school	12-15 years of school	16 or more years	Male	Female
Robots	-1.209	-1.876	-2.025	1.136	-1.628	-0.697
	(0.347)	(0.431)	(0.404)	(0.322)	(0.321)	(0.388)
Immigrants	-0.392	-0.714	-0.464	-0.363	-0.537	-0.204
	(0.098)	(0.133)	(0.121)	(0.120)	(0.136)	(0.076)
No. of observations	0.829	0.365	0.795	0.790	0.784	.797
R-squared	15,996	12,215	15,746	14,921	15,802	15,098

Table 5. IV Estimates of the Impact of Robots and Immigration on log Employment, 2004-2016

	All Workers, by Education Group				Gender	
	All workers	Less than 12 years of school	12-15 years of school	16 or more years	Male	Female
Robots	-2.480	-4.341	-2.865	-1.389	-1.693	-4.287
	(0.955)	(1.287)	(1.114)	(1.163)	(1.098)	(0.967)
Immigrants	-0.188	1.646	-1.344	-0.551	-0.316	-0.085
	(0.580)	(0.592)	(0.622)	(0.476)	(0.678)	(0.576)
No. of observations	16,114	13,768	15,934	15,346	15,986	15,470
R-squared	0.923	0.914	0.900	0.932	0.903	0.959

The Bank-Teller Story



Bank tellers are the next blacksmiths (Washington Post, Feb 8, 2018)

Self-driving cars. Self-serve gas pumps. Self-checkout supermarkets. Add self-banking. No people. No tellers. Nobody greeting you with "How is your day going?" Although, that may come from a machine — and sooner than you think.

Bank of America has opened three mini-bank branches- since the new year that have ATMs and video-conferencing but no people. Two opened in Denver and one in Minneapolis.

In addition to the ATMs, the new robo-banks — called automated centers — allow customers to make a video-conference call to a Bank of America employee at another location to discuss more complicated money issues.

"This is the beginning of the end of the American bank branch ... Bank branches are dead. They were killed by the iPhone. It's like the horseshoe when the automobile came along."

BLS data show fall to 468, 470 in 2018 May https://www.bls.gov/oes/2018/may/oes433071.htm – down from 520,500 in 2014,

AI Beats Humans in POKER– incomplete information game, where bluffing, adjustments in tactics, can affect outcomes.

interview with CMU Researcher Who Programmed AI Poker Champ

Tell me about how this type of technology can be used outside of board games.
CMU Prof Sandholm: I've been working on poker for 12 years and ...in automated negotiation for 27 years. So I don't view poker as an application; poker has emerged as the benchmark in the AI community for testing these types of algorithms for solving imperfect information games. These algorithms work for any imperfect information game. And by game, I don't mean recreational... these games can be very high stakes, like business-to-business negotiations, military strategy planning, cybersecurity, finance, medical treatment planning of certain kinds. These are really for a host of applications, really any situation that can be modeled theoretically as a game. **Now that we've shown that the best AI's ability to do strategic reasoning in an imperfect information setting has surpassed that of the best humans, there's really a strong reason for companies to start using this kind of AI support in their interactions.**

There are three pieces of the architecture, and each one has important advancements over the prior corresponding modules. One is the **strategy computation ahead of the time**, so the algorithms that are game independent, meaning they're not about poker. The second module is the **endgame solving**. During the game, the computer will think about how to refine its strategy. The third piece is **the continual improvement of its own strategy** in the background. So, based on what holes the opponent found in our strategy, the AI will automatically see which of those holes have been the biggest and most frequently exploited. And then **overnight** on a supercomputer, it will compute patches to those pieces of the strategy, and they're automatically glued into the main strategy.

INTERVIEWER: How did you teach Libratus to bluff?

Bluffing is not programmed in. The algorithm for solving these games just comes up with the strategy, and the strategy includes bluffing. Given the input rules of the game, the algorithm will already output a strategy, and that strategy does involve bluffing. And it also involves understanding the opponent's bluffing.

Poker players, lawyers, doctors, accountants, white collar in digital space vs bank tellers, retailers, mfg workers. Not clear which will face greater pressures.

The one "certainty": who owns the robots will earn what the robots do → huge increases in inequality with current pattern of ownership.

Reinforcement Learning: How the Machines Get so Smart So Fast

By spending lots of time **learning.** Natural/artificial intelligence – evolution of what works best – steroids in digital world. Machines compete tens, 00s , 000s of times to attain a goal in digital space. They try out different strategies, compete with themselves, always with some place for innovation to learn, and get rewarded the better they do per evolution … winning strategy replicates itself, then competes against itself. **More training and learning than possible for human.** Landscapes for chess/go/ poker are more complicated than for climbing a hill.

The goal of Machine Reinforcement **Learning** is to learn a good strategy from experimental trials and feedback/reward. Big issue is to divide "activity" between searching landscape space that will be changing and exploiting space by climbing up to local peak and living off the profits. With the optimal strategy, algorithm adapts to the environment to maximize future rewards – for example, the points won in a game over many moves.

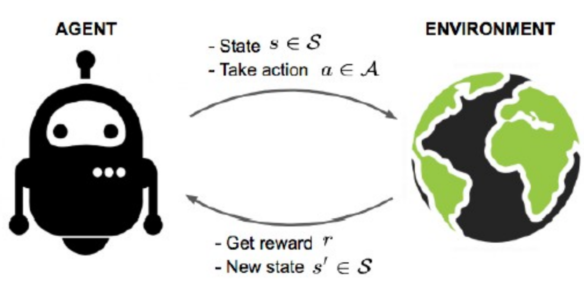


Fig. 1. An agent interacts with the environment, trying to take smart actions to maximize cumulative rewards.

The goal of Reinforcement Learning (RL) is to learn a good strategy for the agent from experimental trials and relative simple feedback received. With the optimal strategy, the agent is capable to actively adapt to the environment to maximize future rewards.

Environment or Game in which the "artificial agent/machine" works to attain some best outcome – a state space S, in which each state has an **Expected value function f**or long term value (if you make best moves from there even if value of different outcomes changes over time as other players adjust their behavior or via more informed expectations). Think of rugged landscape where you not only have to learn the shape of the current landscape to find current highest point but to predict changes in landscape in future (and how fast landscape may change).

If you get your thrill from being a highest peak in a day and you find Mt. Washington NH and everyone follows same strategy as you – the crowd will slow down your trek up so that you will not get to peak in daytime. Bigger thrill will be to get to top of less crowded mountain. And crowd will depend on weather, time of year, if it is weekend or holiday or week day … etc.

You need to learn a lot about not just geography of landscape but of behavior of other climbers.

Goal/ reinforcement – Critical to make sure you give algorithm the "real" maximand – easy to say "win at go/chess/poker" but winning in recruiting worker or finding partner in life is way more difficult. "Get me best worker" – meaning right skills, ability to work with team, great at learning new stuff, LIKELY to stay for long period, willing to accept small pay increases? – A complicated set of criteria requires a lot of learning about the person – you do not want to reject hundreds of candidates and end up with last person in queue.

"The agent has to exploit what it already knows in order to obtain reward, but it also has to explore in order to make better action selections in the future. The dilemma is that neither exploration nor exploitation can be pursued exclusively without failing at the task. The agent must try a variety of actions and progressively favor those that appear to be best. On a stochastic task, **each action must be tried many times to gain a reliable estimate its expected reward. "** The power of the algorithm with big data to analyze is that it can do all this

A reinforcement learning strategy differs from minimax in game theory, where a minimax player would never choose a game state from which it could lose even if in fact it always won from that state because its opponent did not make best play. Assuming each participant makes best move, minimax does not explore possibility someone may make bad moves that can benefit you.

It is different from genetic algorithms, genetic programming, simulated annealing, and other optimization methods by being "smarter" over the long run by looking ahead a expected long term value rather than "throwing mutation dice".

Supervised and unsupervised learning do not balance exploration and exploitation, though they would use exploration/exploitation in fitting models – to avoid overfitting.

It is critical to get the right reward system and which inputs and feedback the AI should pay attention to. If you mis-specify what you really want you can get into big trouble.

Three men are stranded on a desert island. A bottle washes up on the shore. When they uncork the bottle, a genie appears and offers three wishes. Mr A wishes to be taken to Paris. The second man wishes that he were in Hollywood. The third man, now alone on the island, looks around and says, "I wish my friends were back."

"The Monkey's Paw" grants its possessor three wishes. The recipient of the paw wishes for £200. A man knocks on door and says your son has been killed in a terrible work accident, for which the employer makes a goodwill payment of £200. The person wishes the dead son back to life. Strange moaning at the door. Horribly mutilated body appears. The person wishes it away.

Reinforcement learning works when best behavior requires planning or foresight that takes into account delayed effects of one's choices. A simple reinforcement learning player would set up multi-move traps for a shortsighted opponent. It is a striking feature of the reinforcement learning solution that it can achieve the effects of planning and look-ahead without using a model of the opponent and without conducting an explicit search over possible sequences of future states and actions.

Q-learning is to learn a policy, which tells an agent what action to take in a Markov decision process, where you move one step at a time from current state toward a policy that maximizes the expected value of the total reward over any and all successive steps. Q-learning can identify an optimal action-policy given infinite exploration time and a partly-**random** policy. It is the "quality" of an action taken in a given state. See Wikipedia, Q-learning.
https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56

"the true advantage of these algorithms over humans stems not so much from their inherent nature, but from their ability to live in parallel on many chips at once, to train night and day without fatigue, and therefore to learn more. An algorithm trained on the game of Go, such as AlphaGo, will have played many more games of Go than any human could hope to complete in 100 lifetimes." Digital world >>> biological world??