King's Research Portal

Document Version
Other version

Political economy of superhuman AI

Mehmet S. Ismail[*]

Thursday 22nd September, 2022

Abstract

In this note, I show that there is a set of assumptions that would render the existence of general 'superhuman' arfiticial intelligence, denoted by AI*, impossible. These assumptions are (i) the 'freedom of the mind,' (ii) open source access to AI*, and (iii) rationality of the representative human agent, who competes against AI*. I prove that under these three assumptions it is impossible that an AI* exists. This result gives rise to two immediate recommendations for public policy. First, 'cloning' digitally the human brain should be strictly regulated, and hypothetical AI*'s access to brain should be banned. Second, AI* research should be made widely, if not publicly, accessible. *JEL*: C70, C80

Keywords: Artificial intelligence, political economy, AGI, game theory, non-zero-sum games

1 Introduction

In 2015, over 150 artificial intelligence (AI) experts signed an open letter calling researchers from disciplines such as economics, law, and philosophy for doing future research on maximizing the societal benefit of AI.[1] The letter, which has now been signed by over 8000 people, comes with an accompanying paper by Russell, Dewey, and Tegmark (2015) who quote in part the following passage from Horvitz (2014).[2]

---

[*]Department of Political Economy, King's College London, London, UK. E-mail: mehmet.ismail@kcl.ac.uk

[1]https://futureoflife.org/2015/10/27/ai-open-letter/

[2]Whether human-level AI will be achieved or not has long been discussed; for a literature review, see, e.g., Everitt, Lea, and Hutter (2018), and the references therein. For potential economic and political harms of AI technologies in the short-term, see Acemoglu (2021).

> ...we could one day lose control of AI systems via the rise of superintel-
> ligences that do not act in accordance with human wishes—and that such
> powerful systems would threaten humanity... Are such dystopic outcomes
> possible? If so, how might these situations arise? What are the paths to these
> feared outcomes? What might we do proactively to effectively address or lower
> the likelihood of such outcomes, and thus reduce these concerns? What kind
> of research would help us to better understand and to address concerns about
> the rise of a dangerous superintelligence or the occurrence of an "intelligence
> explosion"?

In this note, I study institutions and game theoretic assumptions that would prevent the emergence of superhuman-level arfiticial intelligence. The first assumption is called 'Freedom of the Mind,' which essentially prohibits 'cloning' digitally the human brain. Thus, AI* is not allowed to program humans' brain. The second assumption is called 'Access' which gives the representative human agent the permission to access AI*'s source code so as to take AI*'s strategy as given. The third assumption is 'Rationality' which means that the human agent chooses the strategy that maximizes their payoff, given the strategy of AI*. I prove that under these three assumptions it is impossible that an AI* exists. This result gives rise to two immediate policy recommendations. First, 'cloning' digitally the human brain should be strictly regulated, and potential AI*'s access to brain should be banned. Second, AI* research should be made widely, if not publicly, accessible.

2 The setup

2.1 Two-person (general-sum) perfect information games

Let $G = (N, X, I, u, S)$ be an extensive form game with perfect information and perfect recall. Let $N = \{1, 2\}$ denote the set of players, $x \in X$ a node in game tree X with finitely many elements, x_0 the root of the game tree, $z \in Z$ a terminal node, $I : X \setminus Z \to N$ the player function which gives the active player $I(x)$ at every non-terminal node x, and u the profile of utility functions. Let $A_i(x)$ be the set of pure actions of player i at x and $A_i = \bigcup_{x|I(x)=i} A_i(x)$ be player i's set of all pure actions. For every $i \in \{1, 2\}$, A_i has finitely many elements. Let $S'_i = \bigtimes_{x|I(x)=i} A_i(x)$ denote the set of all pure strategies of i. A pure strategy, $s'_i \in S'_i$, of player i is a function, $s'_i : X_i \to A_i$ such that for all $x \in X_i$, $s'_i(x) \in A_i(x)$. Let $S' = \bigtimes_{i \in N} S'_i$ and $s' \in S'$ denote a pure strategy profile. Let $S_i = \Delta(S'_i)$ denote the set of probability distributions over S'_i, $s_i \in S_i$ a mixed strategy

of player i, and $s \in S$ a mixed strategy profile. Let $s_i(x)(a_i)$ denote the probability with which player i chooses action a_i at node x. For every $i \in \{1, 2\}$, every $s_i \in S_i$ is assumed to be programmable in an agreed upon programming language. For a given strategy profile s, let s_{-i} denote the strategy of player in $j \neq i$. Let $u_i : S \to \mathbb{R}$ be player i's von Neumann-Morgenstern expected utility function.

The representative human agent is denoted by H and the AI agent is denoted by AI. Let $G|x$ denote the subgame of G whose game tree starts at a non-terminal node $x \in X$ and contains all successor nodes in X. Similarly, let $(s|x)$ be the strategy profile restricted to the subgame $G|x$. Let $s_i^* \in BR_i(s_j)$ denote a best-response of player i to player j's strategy s_j, i.e., $s_i^* \arg\max_{s_i' \in S_i} u_i(s_i', s_j)$.

2.2 Concepts

The following two definitions are standard in game theory.

Definition 1 (Nash equilibrium). A strategy profile $s \in S$ is called a *Nash equilibrium* if for every player i and for every $s_i' \in S_i$, $u_i(s) \geq u_i(s_i', s_{-i})$ (Nash, 1951).

Definition 2 (Subgame-perfect equilibrium). A strategy profile $s \in S$ is called a *subgame perfect Nash equilibrium* (SPNE) if for every player i and for every non-terminal $x \in X$ where $i = I(x)$, $u_i(s|x) \geq u_i(s_i', s_{-i}|x)$ for every $s_i'|x \in S_i|x$ (Selten, 1965).

In other words, s is a subgame perfect equilibrium if it constitutes a Nash equilibrium in every subgame of G.

Definition 3 (Repeated contest). Let G_1 denote a game of G in which player 1 is H and player 2 is AI, and G_2 denote a game of G in which player 1 is AI and player 2 is H. Let $G_{1,2}^k$, $k \geq 1$, denote the *repeated contest* game in which each stage game consists of two games, G_1 and G_2, and each stage game is repeated k times.

In simple words, the repeated contest between H and AI is defined as the repeated game in which each stage game consists of two games in each of which the roles of the players are swapped. The reason is that game G may be biased towards one player. To given an example, Chess World Champhionship is a repeated contest in which each player plays with white pieces equal number of times.

Definition 4 (Outperformance). Let G be a two-person perfect information game, $G_{1,2}^k$ the repeated contest, and s be the players' strategy profile in $G_{1,2}^k$. Player $i \in \{H, AI\}$ is said to *outperform* player $j \neq i$ if for any $k \in \{1, 2, ...\}$, $u_i(s) > u_j(s)$.

In plain words, player i outperforms player j in game G if no matter how many times the contest is repeated player i's expected payoff is strictly greater than player j's. Of course, the number of repetitions needed to determine the "better" player in practice may depend on game G. To continue with the chess championship example, 20 repetitions may suffice in a match between H vs AI. In the backgammon championship, however, the contest must be repeated more times to be able to accurately choose the "better" player.

3 Assumptions

3.1 Superhuman AGI

I denote a general-purpose 'superhuman' artificial intelligence, if any, by AI*. It could also be called superhuman artificial general intelligence (AGI).

AI* is a computer program, written in some computer language, and it is equipped with finite but significant computing power—e.g., AI* may come with the most powerful supercomputer available at the time of its use. AI* can take any game G as an input and outputs a *solution*—i.e., a mixed strategy profile—based on its source code and computational power.[3] As the game proceeds, it can update its solution. This is similar to AI engines in chess.

Deciding whether a machine is 'human-like' or 'superhuman' inevitably involves subjective human judgements such as the well-known Turing test (Turing, 1950). There will also be a degree of subjectivity in the definition of AI* in part because it is hypothetical, it is compared with humans, and it is not always clear what 'superhuman' would mean in every context such as in non-zero-sum games. In chess, for example, it is now widely accepted that many top engines are superhuman; though, this has not always been the case.

Consider a game G, a set of human 'experts' of G, and a population of humans who play G. An average strategy of a player is the one that receives the average payoff in a tournament of human players whose format is agreed by the 'experts.' The 'experts' could be actual players who play G at the 'top' level or they could be judges (e.g., a boxing judge) who does not necessarily play the game at the top level but can 'differentiate' between above average, average, and below average strategies.

For example, a chess strategy that instructs capturing the opponent's piece whenever possible would clearly be a below average strategy. However, it is not usually clear how

---

[3]If the game is very large AI* may not be able to precisely find an 'optimal' solution, but even in that case AI* analyzes the game tree and comes up with a solution.

to define average strategy in other (non-zero-sum) games. Thus, there is an inevitable subjective element in the definition of 'average strategy.' For the sake of this paper, it suffices to assume that (below) average strategy in a game G is based on well-established empirical studies on G. For more dicussion, see section 6.

Definition 5 (Superhuman AGI). An AGI is called *superhuman* if (i) there exists $G \in \Gamma$ such that AGI outperforms H in G, (ii) for every $G' \in \Gamma$, AGI is not outperformed by H in G', and (iii) for every $G' \in \Gamma$, the strategy of AGI in G' is not below average as judged by experts in G'.

In plain words, a superhuman AGI must outperform player H in some games and must never be outperformed by H. Condition (iii) includes a subjective aspect of defining "superhuman" AGI, which says that human experts should judge whether AGI's strategy is 'below average' in a population of humans that potentially include many types of players and not necessarily the "rational" types. Arguably, in general-sum games, a superhuman AGI should not only outperform the representative rational agent H but also do 'reasonably well' among a population of humans of different types. Now, we are ready to state the first assumption.

Assumption 1 (SAGI). *There exists a superhuman AGI denoted by AI*.*

3.2 Access

H has the permission to access AI* and take AI*'s strategy as given. This assumption is stated as follows.

Assumption 2 (Access). *For every game $G \in \Gamma$ and at any point during game G, H takes AI*'s strategy $s_{AI} \in S_{AI}$ as given.*

More specifically, in every game $G \in \Gamma$ and at any point during game G, H is granted 'read' and 'copy' permissions in the sense that they can read the source code of AI*, copy its strategy so as to give a response to this strategy. H is not granted any other permissions. In particular, they cannot modify the source code or the computing power of AI* in any way.

3.3 Rationality

H is rational in the sense that H chooses the strategy that maximizes their utility given AI*'s strategy, which is feasible under the Access assumption.

Assumption 3 (Rationality). *Player $i = H$ is rational if in every game $G \in \Gamma$, for every strategy s_j of AI*, where $j \neq i$, H chooses a strategy*

$$s_H^* \in \arg\max_{s_i' \in S_i} u_i(s_i', s_j). \tag{3.1}$$

3.4 Freedom of the Mind

Player H has the *Freedom of the Mind* if H, and only H, has the total control of their own brain: no other entity has control or access in any way to H's brain. For the sake of this note, it suffices to assume that AI* is not allowed to 'digitally clone' H's brain—i.e., program H's brain in a way that H's decisions can be predicted either deterministically or non-deterministically.

Assumption 4 (FoM). *Player H has the Freedom of the Mind.*

The essence of this assumption is that if a human player has the freedom of the mind, then AI* cannot always predict H's strategy, i.e., whatever the code of AI* is regarding the strategy of H, H can change their strategy in the way they wish.[4]

Having access to the source code of AI*, player H should have the freedom to act in the way they want and be potentially unpredictable by the program of AI*. Player H can never be coerced to follow any particular course of action assumed by AI*.

Note that this is a 'mild' assumption in the sense that a human, who has access to the source code of AI*, can always change the strategy that AI* assumes for the human. Thus, this assumption excludes the possibility that AI* 'controls' or 'reproduces' the brain of the human.

4 Centipede game

I next define the centipede game of Rosenthal (1981), which will be used in the proof of the main result.[5] It is a two-person perfect information game in which each player has two actions continue (C) or stop (S) at each decision node. There are several variations of this game game, but the some of the main charecteristics of a standard centipede game

---

[4]Ismail (2022) shows that 'mutual knowledge of rationality' and 'mutual knowledge of correct beliefs' do not hold in general in n-person games, including two-person games. This result roughly means the following in the context of this paper: it is impossible that both players are rational and have correct prediction about the choice of the other player. Since player H has access to AI*, they can predict its strategy, so if AI* is rational it would actually be impossible for AI* to predict H's choice.

[5]This section is mostly adapted from Ismail (2019); see also Ismail (2020).
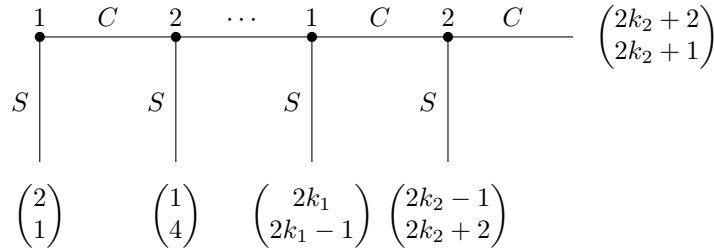
Figure 1: Payoff function of an increasing-sum centipede game.

include (i) the size of the 'pie' increases as the game proceeds, (ii) if player i chooses C at a node, then the payoff of player $j \neq i$ increases, and (iii) the unique subgame-perfect equilibrium is to choose S at every node. To give an example, suppose that there are $m \geq 2$ (even) decision nodes and let $k_i \in \{1, 2, ..., \frac{m}{2}\}$ be the node such that player i is active. Figure 1 illustrates the payoff structure of an increasing-sum centipede game due to Aumann (1998).

Starting from McKelvey and Palfrey (1992), there have been many experimental studies of the centipede game and its variations; see, e.g., Fey, McKelvey, and Palfrey (1996), Nagel and Tang (1998), Rubinstein (2007), Levitt, List, and Sadoff (2011), and Krockow, Colman, and Pulford (2016), which is a meta-study of almost all published centipede experiments. Perhaps the most replicated finding is that human subjects overwhelmingly do not choose S in the first opportunity in increasing-sum centipede games. By contrast, they mostly choose S in the first opportunity in constant-sum centipede games. In addition, the longer the game, the later subjects choose S in increasing-sum centipede games (see, e.g. McKelvey and Palfrey, 1992).

I use the terminology of 'long' centipede game (say, $m \geq 10$) to describe the class of increasing-sum centipede games in which human subjects overwhelmingly choose C in their first opportunity, consistent with the replicated experimental findings.

5 Main result

The following theorem shows that it is impossible that there exists a superhuman AGI if assumptions 2–4 hold.

Theorem 1 (Impossibility of AI*). *Freedom of the Mind, Access, and Rationality imply that AI* does not exist.*

Proof. To reach a contradiction, assume that FoM, Access, and Rationality hold, and AI*

7

exists. By Definition 5, there must be no game in which AI* is outperformed by H. Next, I show in several steps that a long centipede game, denoted by G, is a counter-example in the sense that FoM, Access, and SAGI imply that H outperforms AI* in this game.

First, let $s \in S$ be AI*'s solution of game G characterized by the payoff functions illustrated in Figure 1. Suppose that for every i and for every subgame g of G, $s_i|g \in BR_j(s_j|g)$, i.e., AI*'s solution involves assigning best-responses to each player at every decision node. Then, the solution must be the unique subgame perfect Nash equilibrium in G because otherwise it would assign a non-best-response at least to one of the players at one of the nodes. This is easy to see because in the last node AI* must assign S to the player, who might be itself or H, and given that it must assign S to the previous player and so on. But this solution implies that AI* cannot be superhuman, which contradicts to SAGI assumption. This because such a strategy would be below average: well-replicated experiments show that subjects receive significantly more payoff by choosing C until later stages in the game as discussed in section 4. Now, suppose that $s_{AI}(x_0)(S) > 0.75$, i.e., AI*'s solution s in G assigns more than 0.75 probability to choosing S at the root of the game. Then, the maximum payoff AI* can receive in the best-case scenario is less than $2 \times 0.75 + (m + 2) \times 0.25$, where m is the number of decision nodes in G. Clearly, this payoff is less than $m/2$ for $m \geq 8$, so s_{AI} would be considered below average in G. As a result, $s_{AI}(x_0)(C) \geq 0.25$, i.e., AI* chooses C at the root with probability more than 0.25. This is a lower bound for a strategy to be considered below average in G, but for the remainder of the proof it suffices that $s_{AI}(x_0)(C) \geq 0.25$.

Second, by Access assumption H takes the strategy s_{AI} of AI* as given. And, Rationality assumption implies that H chooses $s_H^* \in \arg\max_{s_i' \in S_i} u_i(s_i', s_{AI})$ (see 3.1), i.e., H best-responds to the strategy of AI*. In addition, FoM assumption implies that H's strategy cannot be predicted by AI*, so $s_{AI} \notin BR_{AI}(s_H^*)$, i.e., AI*'s strategy cannot be a best-response to H's strategy. This is because H is already best-responding to AI*, and if both players are best-responding to each other, then the only possible outcome is to choose S at the first node, which already leads to a contradiction as shown above.

As a result, it implies that H outperforms AI* in the repeated contest $G_{1,2}^k$ for any k because in both G_1 (the game in which H is player 1) and G_2 (the game in which H is player 2), $u_H(s_H^*, s_{AI}) > u_{AI}(s_H^*, s_{AI})$. In other words, H's payoff must be strictly greater than AI*'s payoff in the repeated contest. This is because (i) H best-responds to AI* in both G_1 and G_2 and (ii) $s_{AI}(x_0)(S) \leq 0.75$, i.e., AI* does not choose S with high enough probability at the root. Notice that the payoff function of G is such that the best-responding player receives a greater payoff unless player 1 chooses S with high

enough probability at the root of the game. As desired, H outperforms AI* in the repeated contest, which contradicts to the supposition that AI* is superhuman. □

6   Discussion of the assumptions

Access

Assume that FoM and Rationality hold but Access does not hold. In that case, H would best-respond to some belief about AI*'s strategy. There would be no guarantee that H's belief will be correct. Thus, H would not necessarily be able to outperform H. This implies that AI* *may* be superhuman.

Rationality

Assume that FoM and Access hold but Rationality does not hold. Then, this would *not* contradict to the assumption that AI* is superhuman. The reason is that if H is not rational then they may pick a strategy such that H is outperformed by AI*.

Freedom of the mind

Assume that Access and Rationality hold, but FoM does not. So, AI* might be able to program H's brain and predict precisely what H will choose and can best-respond. AI*'s solution would take into account H's strategy in every game. Ismail (2022) shows that in general players cannot be both rational and predict the others' strategies correctly (which applies to centipede games as well). It implies that one of the players could outguess the other player, depending on perhaps the computational power of AI*. As a result, one cannot rule out the scenario that AI* outperforms H in every game, in which case the theorem would not hold.

Superhuman AGI

As defined in section 3, there must be at least one game in which AI* outperforms the representative human agent H and there should be no games in which AI* is outperformed by H. The definition of outperformance is a formal one, and I believe it is also reasonable in the context of non-zero-sum games.

The notion of superhuman AGI also includes a subjective component in which human experts must consider AI*'s strategy to be not 'below average.' The precise form of this

assumption is certainly open to discussion. But the part of this assumption that is needed and used in the proof of the main theorem is based on replicated empirical findings in perhaps one of the most well-studied games in economics. To be sure, centipede game and what should its 'correct solution' be have been discussed in the literature since its introduction by Rosenthal (1981). But there is little discussion, if any, regarding the fact that overwhelming majority of people do not choose to stop at the first node in long incresing-sum centipede games.

Finally, I would like to note that existence of AI* is obviously a theoretical assumption. It does not make any claim whether it is practically possible to build such a machine. Nor does it make any prediction about when, if possible, will AI* be built. That being said, there have been major developments towards AI* in the past three decades. Notably, Deep Blue was the first chess engine that beat a World Chess Champion, Garry Kasparov, in 1997 (Campbell, Hoane Jr, and Hsu, 2002). Earlier, Tesauro (1995) developed TD-Gammon in 1992 at IBM which became the first computer program based on learning from self-play that surpassed 'average' human level in a major board game. It was a strong program but not strong enough to beat the very best players at the time. Starting from late 1990s to early 2000s computer programs such as Backgammon Snowie and GNU Backgammon achieved superhuman play in backgammon by improving upon TD-Gammon's algorithm. More recently, DeepMind's AlphaGo became the first program that beat a top professional player in Go (Silver et al., 2016). Later, Silver et al. (2018) introduced AlphaZero that achieved a superhuman play in not just one game but in three different games: chess, shogi, and Go.

Schaeffer et al. (2007) went further than these studies in that they showed that the outcome of checkers, a major boardgame, is draw if both players play optimally. That being said, it is unreasonable to expect a complete solution in the near future for major games such as chess and Go in part because of their enormous size. However, there are analytical solutions for some large combinatorial games such as Nim. Hex is another well-known game that has not yet been solved analytically but it can be shown by a strategy-stealing argument that the first player wins by optimal play.[6]

---

[6]Catch-Up is a less well-known two-person combinatorial game whose simple rules lead to a complicated gameplay (Isaksen, Ismail, Brams, and Nealen, 2015). Brute-force backward induction solutions of the small versions of the game and empirical evidence suggest that its optimal outcome may be a draw whenever a draw is possible, as conjectured by Isaksen et al. (2015). Brams (2022) argues that it may be a difficult game for AI to perform well against a randomizing opponent if backward induction is excluded from AI's toolbox.

References

Acemoglu, D. (2021). Harms of AI. *Oxford Handbook of AI Governance*, forthcoming.

Aumann, R. J. (1998). On the centipede game. *Games and Economic Behavior 23*(1), 97–105.

Brams, S. J. (2022). A game that stymies AI. *Plus Magazine*.

Campbell, M., A. J. Hoane Jr, and F.-h. Hsu (2002). Deep Blue. *Artificial Intelligence 134*(1-2), 57–83.

Everitt, T., G. Lea, and M. Hutter (2018). AGI Safety Literature Review. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, pp. 5441–5449. AAAI Press.

Fey, M., R. D. McKelvey, and T. R. Palfrey (1996). An experimental study of constant-sum centipede games. *International Journal of Game Theory 25*(3), 269–287.

Horvitz, E. (2014). One hundred year study on artificial intelligence: reflections and framing. White paper, Stanford University, Stanford, CA (ai100.stanford.edu).

Isaksen, A., M. Ismail, S. J. Brams, and A. Nealen (2015). Catch-Up: A game in which the lead alternates. *Game & Puzzle Design 1*(2), 38–49.

Ismail, M. S. (2019). One for All, All for One—Von Neumann, Wald, Rawls, and Pareto. *arXiv:1912.00211*.

Ismail, M. S. (2020, July). One for All, All for One—Von Neumann, Wald, Rawls, and Pareto. In *Proceedings of the 21st ACM Conference on Economics and Computation*, Virtual Event Hungary, pp. 763–764.

Ismail, M. S. (2022). Mutual knowledge of rationality and correct beliefs in n-person games: An impossibility theorem. *arXiv:2209.09847*.

Krockow, E. M., A. M. Colman, and B. D. Pulford (2016). Cooperation in repeated interactions: A systematic review of centipede game experiments, 1992–2016. *European Review of Social Psychology 27*(1), 231–282.

Levitt, S. D., J. A. List, and S. E. Sadoff (2011, April). Checkmate: Exploring backward induction among chess players. *American Economic Review 101*(2), 975–90.

McKelvey, R. D. and T. R. Palfrey (1992). An experimental study of the centipede game. *Econometrica: Journal of the Econometric Society*, 803–836.

Nagel, R. and F. F. Tang (1998). Experimental results on the centipede game in normal form: an investigation on learning. *Journal of Mathematical Psychology 42*(2-3), 356–384.

Nash, J. (1951). Non-Cooperative Games. *The Annals of Mathematics 54*(2), 286–295.

Rosenthal, R. W. (1981). Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory 25*(1), 92–100.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *The Economic Journal 117*(523), 1243–1259.

Russell, S., D. Dewey, and M. Tegmark (2015, Dec.). Research priorities for robust and beneficial artificial intelligence. *AI Magazine 36*(4), 105–114.

Schaeffer, J., N. Burch, Y. Bjornsson, A. Kishimoto, M. Muller, R. Lake, P. Lu, and S. Sutphen (2007). Checkers is solved. *Science 317*(5844), 1518–1522.

Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nach-frageträgheit. *Zeitschrift für die gesamte Staatswissenschaft*.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature 529*(7587), 484–489.

Silver, D., T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science 362*(6419), 1140–1144.

Tesauro, G. (1995). Temporal Difference Learning and TD-Gammon. *Communications of the ACM 38*(3), 58–68.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind 59*(236), 433–460.