AlphaZero: the Downfall and Salvation of Chess

A Research Paper Submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree Bachelor of Science, School of Engineering

William Bradford Spring 2025

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Advisor Kent Wayland, Department of Engineering and Society

Introduction

Traditionally, chess was considered a human affair. Only humans were able to perform the abstract reasoning and creative thinking needed to play chess, making it a 'human' activity. As computing power grew in the 20th and 21st centuries, computers became dramatically better at chess through machine learning — even eclipsing peak human performance — which leads some to argue that chess has become 'less human,' so to speak. Because chess is considered a form of creative expression, this then decreases the value of human input and expression in chess, and by extension devalues human creative expression on the whole.

According to some chess professionals, machine learning detracts from the human expression of playing chess and makes move selection less diverse, thereby making chess less enjoyable to watch and play. Opponents of this argument claim that machine learning helps players understand which moves caused them to lose and how they can perform better next time and promotes personal growth. These two viewpoints are not mutually exclusive, and this paper aims to demonstrate how the problems mentioned by the first viewpoint can be solved by the benefits of the second.

In today's world, computers have become so much better at chess than humans that even the best humans have very little hope of winning against the best machine learning models. These models, which we will refer to as Alpha Zero models because they are originally based on the Alpha Zero machine learning framework, no longer rely on human input data to train themselves to play chess. Because these models were trained without human input, some argue that they reduce human agency in optimal chess move selection. Others argue that they merely change the expression of human agency, and others still argue that machine learning does not affect human move selection. Accordingly, this paper attempts to address the question, "How can the use of Alpha Zero models be evaluated to assess whether machine learning has made chess less human?"

Background and Context

Historically, chess has been a heavily-modified game. From its origins in chaturanga and later shatranj from Indian and Muslim communities respectively, the Euro-centric adaptation into chess has centered around 'optimizing' the game to be as simple and creatively challenging as possible according to European standards (Larson, 2018). Because it was designed around human perception of creativity, it was seen as a human affair (Ensmenger, 2012). Naturally, by virtue of its history in European spheres, chess began to be associated with intelligent thought, which later evolved into using chess as a metric for machine intelligence (Chowdhary et al., 2023). The first recorded instance of a machine 'playing' chess occurred in the mid-1800′s, when a device called a "Mechanical Turk" was used to fool crowds of people into believing that a machine could play chess by concealing a human within the machine (Bloomfield & Vurdubakis, 1997). At that time, the agency of decision making was still considered human, despite the moves being executed 'by a machine' (Bloomfield & Vurdubakis, 1997). In contrast, in modern times, the decision-making component of machine chess is seen as increasingly mechanical and 'inhuman' affair (Bloomfield & Vurdubakis, 1997).

Now, in the world of computer science, chess has become a cliché (Ensmenger, 2022). Since the conception of artificial intelligence, scientists have aimed to create the best chess algorithm possible, which can be generally referred to as a 'chess engine'. For a long time, these engines were comparatively worse at playing chess than the best humans (Ensmenger, 2012). Initial engines were based on human perspectives, providing values to the chess pieces corresponding to human piece valuations. Chess world champions would regularly compete against the best chess engines, until a chess computer called 'Deep Blue' won against the reigning chess world champion in 1997 for the first time (Larson, 2018). Since then, computers have become expected to win chess matches against humans, but the general public had a much different reaction at that time. Articles published then used phrases like "inscrutable conqueror" to describe the cold, emotionless, calculating nature of the machine that 'beat the world champion' (Larson, 2018). Chess engines have only improved since then. New techniques for training machine learning models, referred to as machine learning frameworks, have removed humans from the process entirely, allowing machines to play chess against themselves and each other and learn how to play truly optimal chess, with minimal outside influence (Tomašev et al., 2022). In the space of computer science, these frameworks are used as groundbreaking new techniques that allow researchers to refine their craft (Larson, 2018), whereas chess enthusiasts are more concerned with the model produced by the framework, and what that model accomplishes for chess (Rozman, 2022). Top level chess players such as Levi Rozman analyze and discuss the merits of moves made by chess models regularly and upload them to YouTube for the general public to view (Rozman, 2022). Furthermore, in an effort to improve the viewing and gameplay experience of chess, researchers have used the AlphaZero machine learning framework to qualitatively evaluate the effectiveness of modifications to the rule set of chess (Tomašev et al., 2022).

The AlphaZero framework was designed to be adaptable and learn how to play many different types of games. Repeated modifications to the framework have been proposed, which allow the framework to be more applicable to any game (Tomašev et al., 2022). To quantitatively evaluate the effectiveness of the framework and its ability to adapt to different game rulesets, and to qualitatively evaluate the variation of optimal move selection of different game rulesets, Tomašev et al. trained several machine learning agents to play different variants of chess. These agents then played against each other, and their matches were observed qualitatively and quantitatively to determine which formats encourage varied move selection at an expert level (Tomašev et al., 2022). Here, we see clearly that the design of these frameworks is influenced by both machine learning researchers and chess professionals, and promotes a dialogue between both parties in which the sociotechnical imaginaries are both expressed in the framework itself. In STS terms, this is referred to as a “boundary object.”

A boundary object is a technology or tool that is multifunctional enough to adjust to the requirements of several groups that employ it, but also hardy enough to retain its identity between all settings in which it is employed and facilitate discussion between social groups as a product of their independent pursuit of sociotechnical imaginaries. (Matthies et al., 2023). Boundary objects were conceptualized to help examine the dynamics of two or more groups focusing on distinct goals in which the context of the object is relatively standard (Trompette & Vinck, 2009). They are usually considered 'desirable' by all participating groups and are very prominent in the analysis of "computer-supported cooperative work" (Trompette & Vinck, 2009).

In the context of modern software development, one example of a boundary object is technical documentation (Matthies et al., 2023). Technical documentation helps those who are not specialized in a given field interact with and employ tools produced by specialists of the field, which is one of the same dynamcis we observe in machine learning research. In the field of machine learning, one particularly prominent boundary object found in prior works is the "structural causal model." This is a field-specific textual and graphical work that enumerates the environmental factors that may influence a prediction given by any statistical model (Heyn & Knauss, 2022). When a researcher in a specific field can provide a structural causal model, this allows machine learning researchers to validate the associated machine learning model by providing a type of guide for the expected output of a correctly functioning model, providing a stronger verification of accuracy for machine learning specialists. In this manner, structural causal models act as boundary objects between machine learning researchers and experts in any field (Heyn & Knauss, 2022).

In much the same way, we can observe machine learning models functioning as boundary objects between machine learning research and the chess industry. Between the contexts of machine learning research and chess, models retain the same core identity. In both contexts, they are statistical models that provide inferences based on data obtained from the state of a chess game. Despite this, their precise uses differ dramatically. In the field of machine learning, they are used to determine which methods of training produce the most accurate or beneficial output. In the field of chess, they are used as advisors or high-level consultants for how to improve at chess and improve the rule set of the game itself. In this manner, they function as boundary objects between machine learning and chess and enable a dialogue between professionals in these fields. As we will discuss later in this paper, this professional dialogue consists of machine learning researchers asking — or hearing from without asking — chess professionals which varieties of AI tools might be valuable in the field of chess. These tools are then developed by machine learning researchers, tested by chess professionals, and then iterated upon based on the feedback those professionals provide.

Applied Theory

This paper proposes that there are two main contexts in which machine learning models function as boundary objects. First and foremost, they function as feedback mechanisms for chess players to assess the quality of their moves or learn better moves for a given scenario. This is the most obvious and most common application of these models. Second, they take the form of a professional consult regarding variations of chess and revisions to the competitive rule set. In this context, they are used to assess what peak gameplay might look like in a variation of chess to glean more information about how interesting the variant would be to play and watch.

As discussed in the results section, chess professionals like Levi Rozman review AI chess matches to gain a greater understanding of the game and gather new ideas for future chess matches. Other players, both professional and amateur, use AI tools to review their chess matches and improve their gameplay. This is the most common and proliferated usage of the technology and the most straightforward implementation of machine learning models as boundary objects. Here, we see that models are trained by researchers to improve their ability to play the game, then repurposed by chess enthusiasts to provide feedback on chess games and moves. Accordingly, both parties have their shared interests expressed in the model.

In recent years, AI has been used to revise the rule set of chess through integration with the tiebreaker system in tiebreakers (Anbarci & Ismail, 2024). The cliché of modern chess is that far too many games end in ties. The 2018 world championship match ended in 12 consecutive draws, and from 1910 to 2018, 18% of chess world championship matches have ended in ties (Anbarci & Ismail, 2024). Current methods of breaking ties are regarded as unfair due to their asymmetric nature, which causes one predetermined player to win after a certain amount of time passes. To remedy this, researchers have modified AI models to provide precise feedback about the quality of each move made by each player and award a win to the player that generally made better moves throughout the game (Anbarci & Ismail, 2024). This is a direct example of machine learning and chess interfacing. In this instance chess enthusiasts use AI models in the interest of making chess competitions more fair and machine learning researchers use this as an opportunity to produce more accurate models for this specific use-case.

Another way that AI is used in modern chess is to assess the quality of adjustments to the rule set of chess. This context produces an interesting dynamic in which new variants of chess are designed to produce more engaging human gameplay and other new variants of chess emerge to challenge machine learning techniques (Iqbal, 2013). One new variant is "Switch-Side Chain-Chess," which allows players to switch sides under certain conditions. This is one of the few novel chess variants designed specifically for machine learning purposes, with the specific intent of stimulating improvement in machine learning models (Iqbal, 2013). We see machine learning models being used by chess professionals to progress the development of the game, and we see machine learning specialists influencing this development both directly and indirectly. This format of chess is intellectually stimulating for humans as well, providing novel and creatively challenging dynamics that are previously unseen in the chess sphere (Iqbale, 2013). This reinforces the characterization of machine learning models as boundary objects because they also encourage the development of new chess variants, which provides mutual benefit to the chess world.

Methods

This work reviews freely avialable media works produced by chess professionals and related community members to help clarify the role of machine learning in the humanity of chess. This process took place in three stages. First, several works were collected to form an initial collection. This initial collection was comprised of anything and everything that seemed to express a unique opinion regarding machine learning in chess. These works were then individually reviewed to gain an understanding of the precise information they detailed. Works that did not express unique sentiments about the involvement of machine learning tools in chess were pruned from the collection. The sentiment of a work was considered non-unique if the work could be summarized using the same codes as another, higher profile work and did not introduce any nuance on those opinions. This now-curated collection was reviewed again to establish codes — or lists of sentiments — for the third and final stage of the review process. In the third stage, deductive coding was used to evaluate the sentiments expressed in the works. This process culminated in the deductive coding-based review of 8 distinct media works.

Results and Discussion

There are two main parties in the dialogue about machine learning in the chess sphere: those who are for machine learning and those who are against it. In this essay, we will consider those who are neutral about machine learning in chess to be against it because all reviewed evidence indicated that this sentiment was more strongly associated with machine learning being perceived as unhelpful in chess. Using a mix of deductive coding, there were eight main sentiments (or codes) extracted from these professional interviews and works. The codes that expressed dissenting opinions on the use of AI in chess were 'AI is unhelpful,' 'AI is misleading,' 'AI has solved chess', and 'AI stifles chess creativity'. Codes that expressed a favorable view of AI in chess were as follows: 'AI chess promotes unorthodox plays,' 'AI can be used to improve chess', 'AI removes barriers to entry in chess', and 'AI generates public attention for chess.'

How AI Ruined Chess

As with many technologies, not all chess professionals and enthusiasts are in support of machine learning or its involvement in chess. Magnus Carlsen, one of the world's leading chess world champions, expressed in an interview that machine learning chess bots are both unhelpful to analyze and pointless to play against because they do not provide human-actionable feedback, and there is no hope to learn from your mistakes against chess engines (Lex Fridman, 2022). In short, he believes that AI has no serious place in chess. In a separate interview with Lex Fridman

(2022), Chess Grandmaster Hikaru Nakaura suggested that he believes that chess might eventually be "solved" by machines, and that certain starting moves might always result in a draw in machine matches, but he did not believe that humans would be able to implement machine strategies (Lex Fridman, 2022). Both of these interviews indicated some degree of disapproval regarding AI in chess, which seemed to indicate that both parties were against the influence of AI in chess.

Several chess professionals indicated that they feel like many professional matches are decisively won by the player who has studied the most AI chess matches (Vaicher-Lagrave et al., 2020). Interestingly, these players cite Magnus Carlsen as one of the main players who allegedly plays similarly to AI agents like AlphaZero, which is in stark contrast to Carlsen's own interview, in which he claims that he does not use chess AI to review his games and does not review AI chess games (Lex Fridman, 2022). In either case, it is clear that AlphaZero has contributed to a growing sense that chess is a 'solved' game in that high-level chess players are "playing from memory" rather than creatively expressing themselves through chess (Simonite, 2020). This sentiment is reinforced by the fact that some bots use "endgame table bases," which are large databases that store the 'solutions' to chess board states with low numbers of pieces (Perez-Rey, 2023). This gives players the impression that creativity matters less because there is a definitive 'correct answer' to some chess game states.

National Chess Grandmaster Sam Copeland claims that artificial intelligence actually misleads chess players into making worse moves in some of the most highly creative situations (Copeland, 2020). In his words, chess engines are not very useful for identifying "fortresses," which are situations in which chess pieces cannot be safely captured, no matter what the opposing player does (Copeland, 2020). In these situations, chess engines are prone to making moves that provide an immediate advantage, but guarantee a draw later in the game (Copeland, 2020). In these ways, chess engines encourage a culture that focuses on moves that guarantee that the player will not lose, rather than encouraging players to aim to win and thereby stifling the creativity of new chess players and reducing their ability to express themselves through chess. In several professional opinions, the rise of chess AI is the leading cause of the 18% draw rate in professional chess matches mentioned previously (Anbarci & Ismail, 2024). This presents a narrative that AlphaZero has 'ruined' chess.

How AI Will Save Chess

Some sources view AI as a new type of player, which can give advice like any other player. Sources like Levi Rozman seem to anthropomorphize the machines, giving them human characteristics, and using words like "love" to describe how the machine "feels" about making certain types of moves (Rozman, 2022). In this way, he can identify further with the chess AI and help viewers identify gameplay strategies that might benefit them, which might contribute to the diversity of their move selection, thus bolstering their creative expression through chess. This contrasts drastically with the views expressed by Copeland (2020) described previously. The main difference in the way that Rozman views AI is that he highlights moves that are significantly different from those chosen by humans, and makes a special effort to analyze why they work in human terms (Rozman, 2022). Copeland, on the other hand, highlights AI use as a coach, and emphasizes its inability to provide varied feedback since it always advises the use of an 'optimal' move (Copeland, 2020). Copeland's use of AI makes it less of a creative tool, whereas Rozman uses AI as inspiration for future chess moves.

Retired chess grandmaster Vladimir Kramnik believes that modern chess is mostly memorization, and AI has made this problem worse. According to him, chess gameplay has grown increasingly boring over the years and AlphaZero is to blame (Simonite, 2020). Despite this, he still views AlphaZero as something that might save chess. Working with Tomašev (Tomašev et al., 2022), he sought to use AlphaZero to uncover new rule sets for chess that might be more creatively stimulating. Using AlphaZero, he and his research group tested nine different modifications of the standard chess rule set. By the end, he claimed to have found a new method of playing chess which made him feel "like a child" again (Simonite, 2020). In his words, after "three moves" of these new chess variants, "you simply don't know what to do." Using this method, chess enthusiasts may be able to use AI to promote creative expression through revision of the chess ruleset.

Grandmaster chess player Maurice Ashley views modern chess AI as an equalizer of chess resources (Sprenger, 2020). He claims that chess AI is “like having a teacher in your pocket,” allowing younger, less privileged chess players to gain the feedback they may need to excel at chess, which significantly lowers the barrier to entry (Sprenger, 2020). Chess was a historically aristocratic game (Januário et al., 2024), and in Ashley’s opinion, this is somewhat alleviated by AI-powered chess coaching. Despite the fact that AI advice may not be the most accurate or applicable to human implementation (Copeland, 2020), having any form of feedback in chess significantly bolsters a new player’s ability to improve. Women’s Chess Grandmaster Andrea Botez claims that chess is being “modernized” by machine learning, and that we are seeing new and previously unused strategies as a direct result of chess computing (Sprenger, 2020). It is inferred from this claim that less experienced chess players gain different takeaways from their beginner matches when they accept feedback from AI, and this causes them to develop previously unseen strategies when they eventually develop their own styles as they build more experience (Sprenger, 2020). Maurice echoes this sentiment, claiming that modern chess engines directly influence players to make more aggressive and outlandish plays in-game (Sprenger, 2020). It was unclear from the interviews whether these professionals believe that the best chess players use AI to gather feedback about their moves.

Some chess professionals claim that, even now, chess engines have not fully ‘mastered’ chess and lack the creative ability to see every possible best move (Vaicher-Lagrave et al., 2020). This would suggest that AI games more of an interesting spectacle than a standard for how the game should be played, which means that chess AI merely increases the entertainment value of the game, attracting new players and generating more attention for chess as an industry, rather than harming the creative expression of chess players (Vaicher-Lagrave et al., 2020).

This presents a clear case that, although some chess professionals perceive that machine learning has reduced the degree of creative expression in chess, many professionals believe that it is profoundly beneficial on the whole. While the early versions of chess AI gave professionals the impression that AI harms the expressiveness of the game, we can clearly see that novel techniques like AlphaZero promote the continued refinement of chess as a means of creative expression, as well as promoting fairer competition by functioning in a tie-breaking capacity in professional settings. It provides greater access to the game by delivering actionable feedback to users that might not have access to coaching, encourages the development of new chess rule sets, helps break ties fairly in professional games, and encourages more aggressive chess play, which breaks the mould of prior chess standards.

Conclusion

Chess is a well-loved, thoroughly studied game that has long been in the spotlight of machine learning. In the modern state of chess, we have seen a rapid increase in player perception that AI is a necessary tool that provides a competitive edge and dinsincentivizes player creativity. This makes chess feel 'less human' through a lack of creative expression. Despite the clear ways in which AI makes chess 'less human,' general professional opinion suggests that it is still beneficial overall. It reduces the barrier to entry for new players, encourages more aggressive chess play, and helps make tournaments more fair. Furthermore, it has helped allow greater creative expression through modifications to the chess rule set that may otherwise have gone untested. To that end, it is clear that, even if machine learning has made chess 'less human,' it has also allowed for significantly more creative expression in chess, thus making chess 'more human' as well. Going forward, it may be fruitful to regard machine learning models as novel actors in the sphere of chess and view chess variants as boundary objects between machine learning and chess, but this paper did not uncover enough evidence to support that claim in good faith. Future efforts may allow chess to become even more creatively expressive and eventually allow for a perfectly fair competition environment in the professional chess scene.

References

Anbarci, N., & Ismail, M. S. (2024). AI-powered mechanisms as judges: Breaking ties in chess. *PLOS ONE*, *19*(11). https://doi.org/10.1371/journal.pone.0305905

Bloomfield, B. P., & Vurdubakis, T. (1997). THE REVENGE OF THE OBJECT? On Artificial Intelligence as a Cultural Enterprise. *Social Analysis: The International Journal of Social and Cultural Practice*, *41*(1), 29–45. http://www.jstor.org/stable/23171730

Chowdhary, S., Iacopini, I., & Battiston, F. (2023, March 23). Quantifying human performance in chess. *Scientific Reports*, *13*(1), 1 - 8.

Ensmenger, N. (2012, February 1). Is chess the drosophila of artificial intelligence? A social history of an algorithm. *Social Studies of Science (Sage Publications, Ltd.)*, *42*(1), 5 - 30.

Heyn, H.-M., & Knauss, E. (2022). Structural causal models as boundary objects in AI system development. *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI.* https://doi.org/10.1145/3522664.3528615

Iqbal, A. (2013). A New Chess Variant for Gaming AI. In: Anacleto, J.C., Clua, E.W.G., da Silva, F.S.C., Fels, S., Yang, H.S. (eds) *Entertainment Computing – ICEC 2013. ICEC 2013. Lecture Notes in Computer Science*, vol 8215. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41106-9_2

Januário, J. d. A., Giglio, S. S., & Amaral, S. C. F. (2024, August 1). Grandmasters, distinct elite: Taste submitted to discussion from the social conditioning factors of the predilection for chess. *International Review for the Sociology of Sport*, *59*(5), 787 - 802.

Larson, M. (2018, January 1). Optimizing Chess: Philology and Algorithmic Culture. *Diacritics*, *46*(1), 30 - 29.

Lea, G. R. (2020). Constructivism and its risks in artificial intelligence. *Prometheus*, *36*(4), 322–346. https://www.jstor.org/stable/10.13169/prometheus.36.4.0322

Luchitskaya, S. (2015). Review of Chess in the Middle Ages and Early Modern Age. *A Fundamental Thought Paradigm of the Premodern World. Fundamentals of Medieval and*

Early Modern Culture, 10, by D. O'Sullivan. Mediaevistik, 28, 387–390. http://www.jstor.org/stable/44163605

Matthies, C., Heinrich, R., & Wohlrab, R. (2023, June). Investigating Software Engineering Artifacts in DevOps Through the Lens of Boundary Objects. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering* (pp. 12-21).

Stewart, H., & Watson, N. (2020). A Sociotechnical History of the Ultralightweight Wheelchair: A Vehicle of Social Change. *Science, Technology, & Human Values*, *45*(6), 1195–1219. https://www.jstor.org/stable/26933350

Tomašev, N., Paquet, U., Hassabis, D., & Kramnik, V. (2022, February 1). Reimagining Chess with AlphaZero: AI is driving the next evolution of chess, giving players a glimpse into the game's future. *Communications of the ACM*, *65*(2), 60 - 66.

Trompette, P., & Vinck, D. (2009). Revisiting the notion of boundary object. *Revue d'anthropologie Des Connaissances.* https://doi.org/10.3917/rac.006.0003

Wilkenfeld, Y. (2019). Can Chess Survive Artificial Intelligence? *The New Atlantis*, 58, 37–45. https://www.jstor.org/stable/26609113

Primary Sources

Copeland, Sam. Chess.com. (2021). *Chess Engines Are Wrong About These Positions.* YouTube. Retrieved March 21, 2025, from https://www.youtube.com/watch?v=TLNtIGuhjaM.

Lex Fridman. (2022). *Magnus Carlsen on AlphaZero: Its willingness to sacrifice pieces is fascinating.* YouTube. Retrieved January 23, 2025, from https://www.youtube.com/watch?v=z_wXRCMY1nE&t=13s.

Nakamura, H., & Fridman, L. (2022, October 18). *Perfect chess opening: If AI solves chess | Hikaru Nakamura and Lex Fridman.* YouTube. https://www.youtube.com/watch?v=BGIoZikJn_w

Perez-Rey, L. (2023). *Why AI Chess Bots Are Virtually Unbeatable.* YouTube. WIRED. Retrieved March 21, 2025, from https://www.youtube.com/watch?v=CdFLEfRr3Qk.

Rozman, L. (2022, January 6). *3600 Elo Stockfish CRUSHED By Leela!.* YouTube. https://www.youtube.com/watch?v=2BndigWZC2A&list=PL9t6iwrQx1DQtxTeklrlCJdSrqrfS9aLi

Simonite, T. (2020, September 10). *AI ruined chess. now, it's making the game beautiful again.* Wired. https://www.wired.com/story/ai-ruined-chess-now-making-game-beautiful/

Sprenger, Richcard. *Have computers killed chess?* (2020). YouTube. Retrieved March 21, 2025, from https://www.youtube.com/watch?v=2bSyq7Z9ErA.

Vachier-Lagrave, M., Gustafsson, J., Shirov, A., Movsesian, S., & Heimann, A. GRENKE Chess. (2020). *Chess Grandmasters on Google Deepmind AlphaZero || Artificial Intelligence in Chess.* Youtube. Retrieved March 21, 2025, from https://www.youtube.com/watch?v=hNw7whJqsJ8.