TEO, ISSN 2247-4382
102 (1), pp. 68-86, 2025

Chess and Artificial Intelligence: A Cognitive Laboratory and an Ethical Challenge

Ciprian Iulian TOROCZKAI

Ciprian Iulian TOROCZKAI
"Lucian Blaga" University of Sibiu, Romania
Email: ciprian.toroczkai@ulbsibiu.ro

Abstract
Chess has long served as a privileged model for studying human reasoning, offering insights into memory, strategy, and decision-making (the sport of the mind). From Alfred Binet's pioneering psychological studies to Alan Turing's exploration of machine intelligence, chess has mirrored humanity's quest to understand cognition. The historic defeat of Garry Kasparov by IBM's Deep Blue in 1997 marked a turning point, demonstrating that brute computational force could outmatch human intuition. However, it was with DeepMind's AlphaZero (2017) that a profound paradigm shift occurred: a system capable of autonomously learning chess strategy, without reliance on human databases, achieving a style both dynamic and creative. AlphaZero's success raises critical epistemological and ethical questions. It challenges traditional concepts of understanding, showing that expert-level competence can emerge without consciousness. Ethically, the evolution from human-machine rivalry to collaboration – epitomized by freestyle chess – suggests that synergy between humans and AI yields superior outcomes. Nevertheless, issues such as explainability, responsibility, and the preservation of human dignity remain pressing concerns, especially as AI systems extend into domains like medicine, law, and scientific research.
This study argues that chess continues to serve as a prototype for broader societal reflections: how we define intelligence, how we govern technological power, and how we ensure that AI remains a tool for amplifying, rather than diminishing, human creativity and ethical responsibility. In the evolving landscape of AI, the greatest

challenge is not technological, but moral: shaping a future where human values remain at the center of innovation.

Keywords
Chess, Artificial Intelligence, AlphaZero, Cognitive Models, AI Ethics

Motivation

This work stems from a constant passion: chess. Although I am not a professional player, my daily practice of this game has taught me essential lessons about patience, strategy, risk-taking, and accepting failure as part of the learning process.

In recent years, as I observed the spectacular developments of artificial intelligence in the field of chess, I felt the need to reflect more deeply on these changes. The emergence of systems such as AlphaZero has not only transformed the way chess is played at the highest level but has also raised fundamental questions about the nature of reasoning, knowledge, and responsibility in the contemporary world.

This personal motivation – a combination of fascination with chess and concern for the ethical implications of technology – underpins the present study. I have sought to explore chess not merely as a game, but as an experimental ground for the major dilemmas of the digital age: the relationship between man and machine, the redefinition of expertise, and the role of ethics in technological development.

Through this study, I hope to offer a balanced perspective, informed both by respect for the human tradition of chess and by the awareness of the responsibility we bear toward emerging forms of artificial intelligence.

Ultimately, I believe that chess teaches us the most important lesson: true progress is not about domination, but about understanding and collaboration. This subtle and demanding lesson remains relevant not only on the chessboard but also in all the major decisions that shape our future[1].

[1] I must mention here a science fiction book I read in my youth that left a deep impression on me: Adrian ROGOZ, *Prețul secant al genunii*, Albatros, București, 1974. The chess-playing planet is not humanoid. It does not speak, build, or display traditional technological progress. And yet, it exhibits a form of thought that is unmistakably

I. Introduction

Chess, a subtle art of anticipation, strategy, and decision-making under time pressure, has traversed the history of humanity not merely as an intellectual competition but as a genuine school of rational thought. Over the centuries, its allure has fascinated thinkers from diverse fields – philosophers, psychologists, mathematicians, logicians, engineers – who have seen in the structure of the game a faithful analogy of human cognitive processes. Distinguished among games by its depth and complexity[2], chess

rational and responsive. The actions of the planet – shifting tectonic patterns, subtle changes in weather, electromagnetic signals – begin to form a coherent system of responses to the crew's interventions. Gradually, the human visitors understand that they are engaged in a match: each move they make is countered by the planet, as if by an unseen opponent. It is as though the entire planet is playing chess – not with pieces and boards, but with land, climate, and time. This turns chess into something more than a game. It becomes a language. Rogoz invites us to confront the possibility of non-anthropomorphic intelligence. The planet is not merely alive in a biological sense – it is cognizant. But unlike the classic portrayals of aliens in humanoid or insectoid forms, this intelligence is planetary: diffuse, ecological, deeply integrated into the geological and energetic processes of the world. In this scenario, the humans become the "alien" element – their inability to comprehend the rules of engagement, or even the motives of their opponent, reveals the limitations of human epistemology. Rationality, the story suggests, may exist in forms far removed from our own. There is no war, no conquest. The planet does not attack, but it responds – thoughtfully, intelligently, even playfully. The story thus avoids the tropes of violent confrontation or domination. Instead, it explores a more refined and unsettling theme: humility before the unknown. The humans are forced to recognize that their definitions of intelligence, consciousness, and communication are not universal. The final revelation – that they have been playing a game all along – is both awe-inspiring and humbling. They are no longer explorers charting an inert universe, but participants in a dialogue with a superior form of awareness. In this remarkable story, chess transcends its function as a game. It becomes a symbol of structured thought, of turn-based rationality, of engagement between two minds. But it also reflects the fragility of human assumptions – the idea that intelligence must look like us, think like us, and follow our rules. Adrian Rogoz's story is not about extraterrestrial invasion. It is about recognizing intelligence in forms we are not prepared to understand, and about the possibility that the universe is not silent, but thinking – and playing.

[2] See *The Legend of Chess and the Grains of Wheat*: Long ago, in ancient India, there lived a wise and thoughtful man named Sissa ben Dahir. Seeking to teach his king a lesson in strategy, balance, and foresight, Sissa invented a game – one that reflected the complexity of life, the unpredictability of war, and the need for planning ahead. That

has been aptly compared to the fruit fly *Drosophila melanogaster* – the model organism of genetics – thus becoming a "Drosophila of reasoning". Through chess, researchers such as Alfred Binet, Alan Turing, and Norbert Wiener sought to uncover the mysteries of reasoning, memory, anticipation, and decision-making, believing that performance in such a controlled environment reflects the fundamental workings of the mind.

In the modern era, chess has assumed a new role: that of a battlefield between humans and artificial intelligence. The culmination of this confrontation occurred in 1997, when Garry Kasparov, considered by many the greatest chess player in history, lost a famous match against IBM's supercomputer Deep Blue. This symbolic victory of machine over man profoundly changed perceptions of technological capabilities.

However, progress did not stop at demonstrating brute computational force. With the development of deep neural networks and machine learning algorithms, such as DeepMind's AlphaZero, chess has become a veritable laboratory for investigating critical contemporary themes: the nature of knowledge, the limits of machine learning, the explainability of algorithmic decisions, and, above all, the ethical implications of machine autonomy.

Thus, this study aims to systematically and critically analyze the impact of recent developments in AI as applied to chess, emphasizing both the cognitive and epistemological dimensions and, most importantly, the ethical challenges this rapid evolution brings to our attention.

---

game was chess. When Sissa presented the game to King Sheram, the monarch was so impressed that he insisted on rewarding the inventor with anything he desired. Sissa made what seemed like a humble request: "My king, place a single grain of wheat on the first square of the chessboard. Then place two on the second square, four on the third, and so on – doubling the amount of wheat on each of the 64 squares". The king laughed. What a modest and clever man, he thought. Surely, this was a small reward for such a grand invention. But when the court mathematicians attempted to fulfill the request, they were stunned. By the 20th square, the amount of wheat exceeded a million grains. By the 40th square, it surpassed a billion. By the 64th square, the final total amounted to 18,446,744,073,709,551,615 grains of wheat – more than all the grain that existed in the kingdom, in the world, or even that had ever been harvested by humankind. The king, humbled and astonished, realized he had been outwitted – not through deceit, but through wisdom. Sissa's request was not just a demonstration of mathematical brilliance, but a timeless lesson in exponential growth, humility, and the power of knowledge.

As machines surpass human expertise not only through calculation but also by discovering new strategies, the question of reconfiguring the relationship between humans and their tools becomes increasingly acute.

II. Chess – The Drosophila of Reasoning

II.1. From Binet to Turing: A History of Cognitive Fascination

Interest in chess as a cognitive model began in the second half of the 19th century when the French psychologist Alfred Binet conducted the first systematic studies on the mental abilities of chess players. In his fundamental work *Psychologie des grands calculateurs et joueurs d'échecs*, Binet argued that success in chess does not depend on an absolute mechanical memory or superhuman calculation capacities, but rather on the development of superior skills in pattern recognition and intuitive anticipation[3]. "The great chess player does not simply see more moves ahead, but senses the coherence of the position", Binet noted with remarkable intuition[4].

This idea would strongly resonate in the 20th century, when Alan Turing, one of the fathers of modern computer science, identified chess as an ideal ground for testing a machine's ability to reason. In his famous 1950 article *Computing Machinery and Intelligence*, Turing explicitly proposed using chess as a preliminary test for the concept of artificial "intelligence", famously formulating the question: "Can machines think?"[5]

Norbert Wiener, the founder of cybernetics, reinforced this perspective, arguing that the ability to play chess well requires feedback processes, prediction, and adaptation – defining elements for any intelligent system, whether biological or artificial[6].

---

[3] Alfred Binet, *Psychologie des grands calculateurs et joueurs d'échecs*, Hachette, Paris, 1894, pp. 22-25.

[4] Alfred Binet, *Psychologie des grands calculateurs et joueurs d'échecs*, pp. 31-37.

[5] Alan Turing, "Computing Machinery and Intelligence", in: *Mind* 59, no. 236 (1950), pp. 433-460.

[6] Norbert Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, Hermann et Cie, Paris, 1948, p. 132.

Thus, from an early stage, chess became an essential platform for exploring the boundaries between human reasoning and the emerging capabilities of machines.

II.2. Kasparov and Deep Blue: The End of Innocence

The match between Garry Kasparov and Deep Blue, held in 1997, marked a turning point not only in the history of chess but also in the cultural relationship between humans and machines. Kasparov himself later confessed that, before that match, he firmly believed that the superiority of human intuition and creativity could not be defeated by pure calculation[7].

Yet Deep Blue proved otherwise. Armed with immense processing power and sophisticated positional evaluation algorithms, the supercomputer was able to identify tactical combinations and plan complex strategies without needing human-like consciousness or intuition. This victory sent a shockwave across the world: for the first time, a machine demonstrated that it could defeat human genius in a domain considered the ultimate bastion of intelligence[8].

Kasparov would later describe this experience as "the end of innocence", a painful realization that brute computational force, combined with well-designed algorithms, could substitute what we had long considered the supreme expression of the human mind.

II.3. AlphaZero: The New Paradigm of Cognitive Autonomy

In 2017, DeepMind, a research company owned by Alphabet (Google), introduced AlphaZero to the world – a revolutionary algorithm capable of learning chess on its own, without using human databases or preprogrammed openings. AlphaZero started with only the basic rules of the game and, through intense self-play, developed strategies superior to those of the strongest existing chess engines, such as Stockfish. "In just a few hours, AlphaZero rediscovered – and reinvented – hundreds of years of

[7] Garry Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, PublicAffairs, New York, 2017, pp. 45-47.

[8] Murray Campbell, A. Joseph Hoane Jr. and Feng-hsiung Hsu, "Deep Blue", in: *Artificial Intelligence* 134, no. 1–2 (2002), pp. 57-83.

chess theory", observe commentators Sadler and Regan with admiration[9].

Even more impressive is its style of play: instead of seeking material advantage or neutralizing the opponent through sterile positional play, AlphaZero promotes dynamic chess, favoring initiative, creative sacrifices, and direct attacks.

As Garry Kasparov noted: "Its style strikingly resembles what great human masters consider beautiful chess"[10].

This achievement not only demonstrated the power of machine learning but also raised fundamental questions about the nature of knowledge, creativity, and even "understanding" within an algorithmic universe.

III. The Evolution of Artificial Intelligence in Chess

III.1. Early Attempts: Between Aspiration and Algorithmic Limitations

Attempts to create programs capable of playing chess date back to the 1950s, a period when computer science itself was in its infancy. The first programs developed by pioneers such as Claude Shannon and Alan Turing were profoundly experimental in nature: they sought to replicate how humans analyzed chess positions, albeit with extremely limited processing resources[11].

Claude Shannon, in his landmark 1950 article *Programming a Computer for Playing Chess*, proposed two types of strategies: "strategy A", based on exhaustive exploration of all possible moves, and "strategy B", based on selective evaluations of the most promising moves[12]. This dichotomy laid the foundation for all subsequent developments, balancing brute force against selective intelligence.

However, in the early years, the computational capacity of computers was far too limited to efficiently explore the game's decision tree. As

[9] Matthew SADLER and Natasha REGAN, *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*, New In Chess, London, 2019, p. 21.
[10] G. KASPAROV, *Deep Thinking...*, p. 214.
[11] Claude SHANNON, "Programming a Computer for Playing Chess", in: *Philosophical Magazine* 41, no. 314 (1950), p. 256-275.
[12] Claude SHANNON, "Programming a Computer for Playing Chess", p. 261.

Norbert Wiener observed, "A machine can simulate simple decisions, but in its incipient form, it cannot replicate the complexity of human anticipation"[13].

These limitations led, in the following decades, to the search for algorithmic methods capable of reducing the search space without sacrificing decision quality.

III.2. Minimax and Alpha-Beta Pruning: A Silent Revolution

One of the most important innovations in computational chess was the development of the minimax algorithm, which enabled position evaluation based on the assumption that both players would act optimally for their own interests. Later, the alpha-beta pruning technique refined this method, eliminating branches in the decision tree that could not influence the final outcome[14].

This approach exponentially increased the efficiency of chess programs, allowing them to evaluate millions of positions in a reasonable time frame. Nevertheless, the algorithms remained fundamentally devoid of strategic "understanding". As Hubert Dreyfus pointedly remarked in his critical reflections: "Machines do not understand chess; they compute"[15]. This fundamental difference between human reasoning and machine processing would persist until the early 21st century.

III.3. IBM Deep Blue: The Triumph of Brute Force

In the 1990s, IBM invested heavily in the development of a supercomputer dedicated to chess: Deep Blue. Through a combination of ultra-powerful hardware and efficient search algorithms, Deep Blue managed to analyze approximately 200 million moves per second – a computational capacity incomparably superior to any other system of the time[16].

---

[13] Norbert Wiener, *Cybernetics...,* pp. 150-152.

[14] Donald Michie, "Game-Playing and Game-Learning Automata", in: R. W. Stevens (ed.), *Digital Computers and Their Application*, Princeton University Press, Princeton, 1962, p. 215.

[15] Hubert Dreyfus, *What Computers Still Can't Do*, MIT Press, Cambridge, 1992, pp. 67-70.

[16] Murray Campbell, A. Joseph Hoane Jr. and Feng-hsiung Hsu, "Deep Blue", pp. 57–83.

In the famous 1997 match, Deep Blue defeated Garry Kasparov, marking the first victory of a machine in a chess world championship setting. Kasparov later remarked that facing the machine was "like playing against a galaxy of possibilities"[17].

This victory, although impressive, did not bring a deeper understanding of reasoning mechanisms. Deep Blue was a titan of calculations, but essentially remained an "idiot savant": powerful, yet lacking adaptability and intuition.

III.4. The Era of Stockfish and Fritz: Perfection without Creativity

After the Deep Blue era, the development of chess engines continued through the refinement of evaluation methods and the expansion of databases. Engines such as Stockfish and Fritz became benchmarks of excellence, dominating computer chess competitions. Their main characteristics included: 1) The use of enormous opening and endgame databases. 2) Sophisticated positional evaluation algorithms, manually encoded by experts. 3) A constant increase in processing power, paralleling hardware advances.

Yet, as Matthew Sadler observed, "games between two engines often became sterile, dominated by draws and an obsession with risk avoidance"[18]. Creativity, unpredictability, and the beauty of chess seemed to have become collateral victims of algorithmic perfection.

III.5. AlphaZero: The Emergence of a New Form of Intelligence

In 2017, DeepMind announced the launch of AlphaZero, a revolutionary algorithm that did not rely on databases or preprogrammed rules but instead on autonomous learning. AlphaZero was trained solely with the fundamental rules of chess. Without any supplementary information, it began playing against itself, learning from each victory and defeat. In just a few hours, AlphaZero surpassed the performance of Stockfish, recording remarkable results: 28 wins, 72 draws, and zero losses in 100 games[19].

---

[17] Garry Kasparov, *Deep Thinking...,* p. 89.
[18] Matthew Sadler and Natasha Regan, *Game Changer...*, pp. 60–61.
[19] David Silver *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play", in: *Science* 362, no. 6419 (2018), pp. 1140-1144.

Its games were distinguished by an extremely active style, with material sacrifices in favor of initiative, spectacular attacks, and a clear preference for dynamic dominance over static advantages. As Garry Kasparov noted: "AlphaZero does not play chess the way a human or a classic computer would. It plays a new kind of chess, created by itself"[20].

This achievement marked a paradigm shift: for the first time, an algorithm excelled not just through calculation but also exhibited characteristics of emergent creativity.

IV. The Cognitive and Epistemological Implications of Artificial Intelligence in Chess

IV.1 What Does It Mean to "Understand" in an Algorithmic Context?

Until the emergence of deep learning systems, chess programs operated on a strictly procedural logic: they processed millions of positions, applied predefined evaluation rules, and chose the moves with the highest score. This method, while efficient in terms of performance, did not involve actual understanding. As philosopher Hubert Dreyfus insightfully remarked: "A machine that calculates rapidly does not understand the game; it merely simulates apparent expertise"[21].

AlphaZero fundamentally changes this paradigm: it does not apply rules coded by programmers but creates its own strategies through self-experimentation. As David Silver and his colleagues emphasize: "AlphaZero learns by direct experience, without preexisting human knowledge, reformulating the fundamental concepts of chess"[22].

Thus, a natural question arises: Can we consider that AlphaZero "understands" chess if it plays better than any human or traditional program, even though it has no access to accumulated human knowledge?

In contemporary epistemology, this problem fits within the debate between *functionalism* and *interpretativism*: Is understanding simply the

[20] G. Kasparov, *Deep Thinking...*, 214.
[21] Hubert Dreyfus, *What Computers Still Can't Do…*, pp. 102-105.
[22] David Silver *et al.*, "A general reinforcement...", pp. 1140-1144.

ability to operate effectively, or does it necessarily require consciousness, intentionality and meaning?

Through its results, AlphaZero demonstrates that operational understanding – defined as the ability to act appropriately in complex contexts – can exist even in the absence of consciousness.

IV.2 Reevaluating Fundamental Values: From Material Gain to Activity

By analyzing AlphaZero's games, many commentators have noticed a profound shift in strategic priorities. Unlike the classical tradition, which favored material accumulation (winning pawns, pieces), AlphaZero privileges initiative, piece activity, and dynamic pressure over the opponent. As Sadler and Regan note: "AlphaZero is willing to sacrifice a pawn, an exchange, or even an entire piece to achieve superior piece activity"[23].

This approach suggests that notions once considered fundamental in chess – such as material advantage – are, in fact, relative and contextual. In AlphaZero's chess, a position rich in active possibilities, with continuous initiative, is worth more than static material gains.

Illustrative examples include games where AlphaZero deliberately sacrifices a rook or several pawns to maintain long-term pressure – a strategy that, in classical human chess, would have been considered risky or even erroneous.

Thus, it becomes clear that machine learning can not only imitate existing rules but also reformulate the basic principles of a domain, offering a new vision of what it means to act "correctly" or "optimally."

IV.3. Autonomous Knowledge Generation: A Philosophical Challenge

One of the most profound epistemological implications of AlphaZero's performance is the phenomenon of autonomous knowledge generation. Without access to human databases or established theories, AlphaZero independently reinvented dozens of essential chess concepts – control of

[23] Matthew Sadler and Natasha Regan, *Game Changer...*, pp. 89-91.

the center, wing attacks, pawn sacrifices for initiative. This capacity raises a fundamental question: If a non-human entity can reach valid knowledge through self-learning, how should we redefine the status and nature of knowledge?

Luciano Floridi emphasizes in *The Ethics of Information* that: "In the digital age, artificial systems can generate emergent semantic knowledge"[24]. Thus, chess offers us a paradigmatic example of epistemology without a subject: knowledge becomes an emergent product of performance optimization within a decision space, without requiring consciousness or reflective intentionality.

This perspective profoundly modifies the traditional understanding of the relationship between intelligence, knowledge, and agency.

IV.4. Chess as an Epistemological Laboratory: Broader Implications

Beyond its spectacular game achievements, chess in the AlphaZero era has become an epistemological laboratory for studying how knowledge can be generated, refined, and applied in artificial systems. Three major dimensions can be identified: 1) *Adaptive Heuristics*: AlphaZero proves that heuristics (rules of thumb) can be autonomously discovered without human cultural transmission. 2) *Pragmatic Validation*: In the absence of explicit theories, performance becomes the ultimate criterion for validating knowledge. 3) *Methodological Transferability*: The same self-learning principles applied in chess can be extended to other domains – predictive medicine, logistics optimization, scientific research. As Yoshiyasu Takefuji concludes: "Modern machine learning is not just a tool; it is an autonomous source of additive and shareable intelligence"[25]. Thus, chess, a discipline centuries old, remains in the algorithmic era an essential model for understanding the future of knowledge.

---

[24] Luciano FLORIDI, *The Ethics of Information*, Oxford University Press, Oxford, 2013, pp. 157-159.

[25] Yoshiyasu TAKEFUJI, "Machine learning intelligence is addable and shareable", in: *Science* (eLetter, 6 December 2018).

V. The Ethical Implications of Artificial Intelligence in Chess and Society

V.1 From Competition to Collaboration: Reconfiguring the Human–Machine Relationship

The victory of Deep Blue over Garry Kasparov in 1997 was, for many, a wake-up call regarding the future relationship between humans and artificial intelligence. Initially, the dominant reaction was one of antagonism: the machine was seen as a dangerous rival, capable of threatening human supremacy in domains previously considered unreachable by algorithms[26].

Reflecting on that experience, Kasparov later confessed: "I realized too late that it was not a battle between man and machine, but an opportunity for collaboration"[27].

In the following years, this vision of collaboration crystallized into the concept of freestyle chess (or centaur chess), where teams composed of humans and computers work together, combining human intuition with the computational power of algorithms.

The results were remarkable: hybrid teams managed to outperform both the strongest human players and the most powerful autonomous chess engines. This experience suggests a fundamental ethical lesson: true strength lies not in mutual exclusion, but in synergy between the complementary abilities of human beings and machines.

V.2. The Explainability of Algorithmic Decisions: A Major Ethical Challenge

One of the foundational principles of ethics in artificial intelligence is the requirement for explainability. In chess, if a chess engine recommends a move, it can be later analyzed and evaluated. However, in domains such as medicine, justice, or finance, algorithmic decisions directly impact people's lives.

The essential problem is that many modern deep learning systems – including AlphaZero – are essentially opaque: they arrive at decisions

[26] Garry KASPAROV, *Deep Thinking...*, pp. 45-47.
[27] Garry KASPAROV, *Deep Thinking...*, pp. 192-195.

through highly complex internal structures, impossible to interpret in simple human terms[28]. As Finale Doshi-Velez and Been Kim highlight: "Deep learning systems operate as black boxes: they produce correct results but without the ability to explain their internal processes transparently"[29].

This lack of transparency raises serious issues regarding decision-making responsibility: Who is accountable if an algorithm makes a mistake? How can a decision be contested if we do not understand the internal mechanisms that led to it?

Even in chess, where consequences are limited to the loss of a game, this opacity generates discomfort. In critical fields, the need for *Explainable AI (XAI)* becomes imperative.

V.3. Algorithmic Autonomy and the Problem of Moral Responsibility

With the development of systems capable of self-learning and autonomous adaptation, a crucial ethical question arises: Can moral responsibility be delegated to a machine?

In the case of chess, the responsibility of a program like AlphaZero is simple: optimize moves to win the game. But what happens when similar algorithms are deployed in decision-making systems regarding bank loans, medical diagnoses, or even the use of lethal force in armed conflicts?[30]

Cathy O'Neil warns in *Weapons of Math Destruction*: "Algorithmic models can amplify existing injustices, and their lack of transparency makes them nearly impossible to correct"[31]. From an ethical point of view, responsibility must firmly remain in the hands of the human actors who design, implement, and supervise algorithmic systems. Assigning

[28] Steven Cramton, "Freestyle Chess: Human-Computer Collaboration in Competitive Play", in: *AI & Society* 33, no. 2 (2018), pp. 311–320.
[29] Finale Doshi-Velez and Been Kim, "Towards A Rigorous Science of Interpretable Machine Learning", *arXiv preprint* arXiv:1702.08608 (2017).
[30] See Tim Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences", in: *Artificial Intelligence* 267 (2019), pp. 1-38; Finale Doshi-Velez and Been Kim, "Towards A Rigorous Science of Interpretable Machine Learning", *arXiv preprint* arXiv:1702.08608 (2017).
[31] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown, New York, 2016, pp. 97-101.

responsibility to non-conscious entities, like neural networks, is not only erroneous but also dangerously convenient.

V.4. Impact on Concepts of Merit and Expertise

AlphaZero has demonstrated that an autonomous algorithm can achieve a level of expertise superior to human capabilities within a well-defined domain. This achievement questions traditional concepts of merit, expertise, and human value: If a system can learn faster, deeper, and more efficiently than a human, what does it mean to be an "expert"? What is the value of human effort, gradual accumulation of knowledge, and experience, in a world where a machine can reconstruct (and surpass) that expertise in a matter of hours?

As Matthew Sadler observes: "AlphaZero did not just defeat Stockfish; it redefined the very nature of learning and progress in chess"[32]. Ethically, it is essential to recognize and preserve the uniqueness of human creativity: not as a quantitative competition of performance, but as an expression of meaning, emotion, and aspiration.

V.5. Ethical Synergy: Man and Machine as Partners

Ultimately, the fundamental ethical lesson offered by the chess experience in the AI era is the necessity of synergistic collaboration between humans and machines. As freestyle chess competitions demonstrate, the combination of human intuition and algorithmic computing power can exceed the performance of any isolated entity. Steven Cramton concludes: "In freestyle chess competitions, centaurs – human-machine combinations – dominate both human players and autonomous computers"[33].

From an ethical perspective, this implies: 1) Maintaining human control over major decisions. 2) Ensuring transparency of algorithmic processes. 3) Respecting human creativity and dignity as non-negotiable values. Thus, the ethical future does not envision the replacement of humans, but rather the amplification of human capabilities through an intelligent partnership with machines.

[32] Matthew SADLER and Natasha REGAN, *Game Changer...*, pp. 143-147.
[33] Steven CRAMTON, "Freestyle Chess...", pp. 314-316.

VI. Case Studies and Reflections on the Future

VI.1. Chess as a Prototype for Other Fields of Artificial Intelligence

The experience of chess as a laboratory for AI development is not unique; rather, it serves as a prototype for many other fields where reasoning, decision-making, and learning are essential. For example, in medicine, deep learning algorithms are already being used to analyze medical images, diagnosing certain types of cancer with an accuracy superior to that of human specialists[34]. Similar to AlphaZero, these algorithms learn from raw data without explicit human guidance.

In the legal field, systems like COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) attempt to predict the risk of criminal recidivism, thereby influencing judicial decisions[35]. Although these applications raise major issues concerning fairness and transparency, the analogy with chess is clear: AI is becoming an active agent in complex decision-making processes.

Another spectacular example is AlphaFold, also developed by DeepMind, which has succeeded in predicting the three-dimensional structures of proteins with unprecedented accuracy – solving a problem that had long resisted traditional scientific methods[36].

These examples show that the model of autonomous learning, successfully experimented in chess, is transferable and scalable, with significant ethical implications for each field of application.

VI.2. Future Models of Ethical Human–Machine Collaboration

Starting from the lessons of freestyle chess, three fundamental models of collaboration between humans and AI can be outlined:

- *The Augmentative Model*. In this model, the human retains full control, using AI as a consulting tool. Systems provide suggestions, but

[34] Andre Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks", in: *Nature* 542 (2017), pp. 115-118.

[35] Julia Angwin *et al*., "Machine Bias", in: *ProPublica* (2016).

[36] John Jumper *et al*., "Highly accurate protein structure prediction with AlphaFold", in: *Nature* 596 (2021), pp. 583-589.

the final decision entirely belongs to the human actor. This is the preferred model in sensitive fields like medicine or justice, where moral responsibility must remain clearly human.

- *The Assisted Autonomy Model*. Here, AI proposes autonomous solutions, while the human intervenes only to supervise or correct critical decisions. This model allows for increased efficiency but requires strict standards of explainability and auditability.
- *The Creative Synergy Model*. The most advanced model envisions a genuine collaboration between human and machine, where each brings unique abilities. Just as "centaurs" dominate freestyle chess competitions, hybrid teams could generate solutions that are impossible to reach by either party alone. As Garry Kasparov notes: "Together, humans and machines can discover patterns and solutions inaccessible to either alone"[37]. From an ethical standpoint, this model offers the greatest promises – but also the greatest challenges regarding responsibility, transparency, and the preservation of human dignity.

VI.3. Future Ethical Challenges: Between Innovation and Responsibility

Looking ahead, the main ethical challenges are: 1) *Ensuring explainability*: Technology must develop Explainable AI to allow auditing and challenging of algorithmic decisions[38]. 2) *Preventing algorithmic discrimination*: AI models must be carefully monitored to avoid reproducing or amplifying existing social inequalities[39]. 3) *Protecting human creativity*: In an increasingly automated world, we must maintain spaces where human creativity and initiative remain central. 4) *Establishing international regulations*: There is a need for the development of global legal norms concerning the use of AI in critical fields, particularly the military domain[40].

---

[37] Garry Kasparov, *Deep Thinking...,* p. 222.
[38] Tim Miller, "Explanation in Artificial Intelligence", in: *Artificial Intelligence* 267 (2019), pp. 1-38.
[39] Cathy O'Neil, *Weapons of Math Destruction,* pp. 113-116.
[40] United Nations, *Report of the 2020 Group of Governmental Experts on Lethal Autonomous Weapons Systems*, Geneva, 2020, https://documents.un.org/doc/undoc/gen/g20/319/98/pdf/g2031998.pdf.

Chess teaches us a valuable lesson: the victory of the machine over the human does not mark the end of human dignity, but the beginning of a new era where ethical collaboration becomes the cornerstone of progress.

VI.4. An Orthodox Perspective

Cristinel Ioja's article "Artificial Intelligence and Human Intelligence: A View on New Technologies from the Perspective of Orthodox Anthropology"[41] discusses the fundamental difference between human and artificial intelligence (AI) through the lens of Orthodox theology. He argues that while AI can simulate human cognitive processes, it cannot replicate the human being's spiritual and ontological depth, which stems from being created in the image of God.

At the dawn of the 21st century, the rise of AI signals a pivotal anthropological shift. Though it promises significant advantages across various domains, it simultaneously fosters post-humanist ideologies that risk severing humanity from its transcendent origin. In this context, Ioja emphasizes that true human identity is not the product of technological evolution but is rooted in Christ, the divine Archetype.

Self-knowledge, according to Orthodox anthropology, is a journey of ascetic effort and spiritual communion. Technologies, focusing solely on material and cognitive dimensions, bypass the ontological mystery of the person. AI, although able to replicate external functions of human intelligence, cannot access the inner spiritual reality that defines human uniqueness.

Human intelligence, in its fullness, encompasses spiritual faculties—such as the heart (*kardia*), spirit (*pneuma*), and mind (*nous*)—that remain unreachable to algorithmic imitation. The human being's longing for transcendence, his thirst for infinite meaning, cannot be encoded or simulated.

In conclusion, Ioja asserts that Orthodox theology preserves an authentic understanding of human nature. Despite the impressive advancements in AI, no artificial construct can capture the profound mystery and divine

[41] Cristinel Ioja, "Artificial Intelligence and Human Intelligence: A View on New Technologies from the Perspective of Orthodox Anthropology", in: *Teologia* 98 (2024), nr. 1, pp. 10-23.

vocation embedded within each person. Human dignity and freedom, grounded in communion with God, infinitely surpass the capacities of technological replication.

VII. Conclusions

Chess, this "Drosophila of reasoning," has been and continues to be a privileged laboratory for understanding the cognitive, epistemological, and ethical implications of artificial intelligence. From early basic algorithms to the emergence of AlphaZero, chess has shown that: 1) AI can learn, create, and even surpass human expertise in complex domains. 2) Algorithmic "understanding" does not require consciousness or human-like meaning. 3) The ethics of AI usage is more crucial than ever for protecting fundamental human values.

Facing increasingly powerful artificial intelligence, the great challenge is not merely technological, but moral: How to use this power to enhance human dignity, creativity, and freedom – not to diminish them. Chess offers us the essential lesson: our real adversary is not the machine, but our own moral and epistemological limitations.

In light of these challenges, the fundamental choice is between fear and collaboration[42]. Chess urges us to choose collaboration: to harness the power of algorithms to amplify human creativity, not to erode it. As Kasparov wisely put it: "The future belongs to those who embrace machines, not those who fear them"[43].

[42] For more details see: Mark Coeckelbergh, *AI Ethics*, The MIT Press, Cambridge, MA, 2020.
[43] Garry Kasparov, *Deep Thinking...*, 248.