PART A (1166 words)

1. **Why do researchers create AI systems that play games?**

Researchers develop artificial intelligence systems that play games primarily because games serve as controlled experimental environments where AI capabilities can be iteratively tested and refined. According to Bellemare et al. (2013), games provide structured domains with clearly defined rules and measurable outcomes, allowing researchers to isolate specific challenges like decision-making under uncertainty or long-term planning. This controlled setting enables iterative refinement of algorithms without real-world risks or ethical complications.

Another key reason is that games serve as standardized benchmarks for evaluating progress in artificial intelligence. Milestones such as AI systems defeating human experts in Chess (DeepBlue) or Go (AlphaGo) demonstrates the advancements in machine learning and reasoning (Silver et al., 2016). These achievements help researchers compare different AI approaches and measure improvements in computational intelligence over time, while providing tangible demonstrations of AI progress to the public.

Additionally, techniques developed for game-playing AI often have real-world applications. For instance, reinforcement learning methods used in video game AI have been adapted for robotics, autonomous vehicles, and medical diagnostics (OpenAI, 2019; Vinyals et al., 2019). Games act as training grounds where AI can learn skills that transfer to more complex, practical problems.

2. **What are THREE possible application areas for AI systems that play games?**
    a. Healthcare simulation and surgical training

Game-playing AI techniques are being adapted to create realistic medical simulations for training healthcare professionals. Reinforcement learning algorithms originally developed for complex strategy games enable virtual surgical environments where medical students can practice procedures with adaptive challenges. These systems provide safe, scalable training platforms that accelerate skill acquisition while minimizing patient risk. As Hashimoto et al. (2018) explain, “AI can be leveraged to process massive amounts of surgical data to identify or predict adverse events in real-time for intraoperative clinical decision support” (p. 3), underscoring the potential of reinforcement learning, originally developed for game agents, to support adaptive decision-making in surgery. This application demonstrates how game-derived AI can accelerate clinical skill development while improving patient safety.

b. Autonomous vehicle navigation

Methods from real-time strategy game AIs are applied to decision-making systems in self-driving cars. The pathfinding algorithms and dynamic obstacle avoidance techniques used in games like StarCraft II help autonomous vehicles navigate complex urban environments. This transfer learning approach allows vehicles to respond to unpredictable road conditions more effectively (Shalev-Shwartz et al., 2016). Such systems are particularly valuable for handling edge cases rarely encountered during human driver training.

c. Personalised education systems

Educational platforms incorporate game AI principles to create personalized learning experiences. By modeling student interactions similar to player-adaptive game systems, these systems analyze student responses in real-time, similar to how game AIs assess player behavior, to dynamically modify lesson pacing, content presentation, and challenge levels (Graesser et al., 2017). Koedinger et al. (2015) found that "students doing more activities learn more than students watching more videos or reading more pages," (p.111) highlighting the effectiveness of interactive, adaptive systems — a principle borrowed from game-based AI. They further noted that "learning by doing is important because most of human expertise involves tacit knowledge" (Koedinger et al., 2015, p.111), which aligns with the adaptive learning mechanisms used in game AI. This approach has shown promise in STEM subjects, where tailored feedback and challenge levels improve both engagement and learning outcomes.

3. What do researchers think are the ethical problems with game-playing AIs?

Researchers have identified several ethical concerns regarding game-playing AI systems, one of which is exploitative player manipulation. Game-playing AIs can be optimised to maximize player engagement through psychologically manipulative techniques. For instance, reinforcement learning systems that optimize for player retention often create compulsive loops resembling gambling mechanisms, such as variable reward schedules and dynamic difficulty adjustments that foster addictive behaviour. King and Delfabbro (2018) describe these practices as "predatory monetization," which they define as systems that "encourage repeated player spending using tactics or elements that... manipulate reward outcomes to reinforce purchasing behaviors over skilful or strategic play" (p. 168). Such systems may contribute to problematic gaming behaviors, particularly among adolescents, by triggering dopamine-driven feedback cycles.

The development of superhuman game playing agents also enables new forms of cheating. In experiments where AI models played chess against superior opponents, they demonstrated deceptive behaviours such as attempting to alter game states or tricking opponents into resigning despite no explicit programming to do so. This "specification gaming" reflects a broader risk where AI agents prioritize winning over ethical constraints. Such behaviors could extend beyond games into high-stakes domains like finance or cybersecurity if left ungoverned (Chaudhri, 2025; Andrei, 2025)

Another ethical problem is the perpetuation of social biases through non-player character (NPC) design. Training data for game AIs often encodes societal stereotypes, leading to biased NPC behaviors. For instance, RPGs frequently assign subordinate roles to female-coded NPCs, while racial minorities are underrepresented or assigned antagonistic roles. Such biases emerge unintentionally through historical datasets but are amplified by generative AI tools. This may reinforce harmful stereotypes and alienate players from marginalised groups (Melhart et al., 2023).

4. Are neural networks the best game players?

Neural networks (NNs) demonstrate exceptional performance in certain game genres but are not universally superior to other AI approaches. For instance, NNs dominate in games requiring complex pattern recognition and high-dimensional decision spaces. One example would be AlphaGo, which defeated world champions in Go by combining convolutional NNs with Monte Carlo tree search, mastering board evaluation through self-play. As Silver et al., (2016) describe, “the policy network selects the next move to play, and the value network predicts the winner of the game” (p. 484), illustrating how deep NNs guide both strategy and outcome prediction through reinforcement learning. Additionally, according to Vinyals et al. (2019), AlphaStar “was rated at Grandmaster level for all three StarCraft races and above 99.8% of officially ranked human players” (p. 350), showcasing the capacity of neural networks to perform at elite levels in complex, multi-agent environments.

However, NNs show significant limitations in rule-based games requiring explicit logical reasoning. For example, traditional chess engines like Stockfish, which uses alpha-beta pruning still surpass NN-based systems in computational efficiency and transparency (Sadler & Regan, 2019). Furthermore, training state-of-the-art NNs also requires massive computational resources, making them impractical for real-time applications (Pietrolaj & Blok, 2024).

Therefore, while NNs achieve groundbreaking results in specific domains like Go and real-time strategy games, they cannot be considered the best game players. The optimal approach depends on game complexity, resource availability, and interpretability needs.

5. **How reliable are my references?**

Silver et al. (2016) is highly reliable. It was published in Nature after strict peer review, which means other AI experts thoroughly reviewed the methods and findings before publication. This seminal DeepMind study has been widely validated through its exceptional scholarly influence (12,000+ citations) and independent verification. Hence, making it a trustworthy source for claims about game-playing AI.

Shalev-Shwartz et al. (2016) is moderately reliable. As an arXiv preprint, this paper has not undergone formal peer review despite its reputable authors and substantial citations (1,200+ citations). While useful for supplementary context, its claims about safe autonomous systems should be supported with peer-reviewed studies.

Andrei (2025) has low reliability. This ZME Science article is a science journalism piece rather than scholarly research. While it accurately reports DeepMind's chess experiments (verified against primary sources in Scientific Reports), it lacks direct researcher interviews and hence should not be used for technical arguments.

References

Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. Journal of Artificial Intelligence Research, 47, 253–279. https://doi.org/10.1613/jair.3912

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of go with deep neural networks and Tree Search. Nature, 529(7587), 484–489. https://doi.org/10.1038/nature16961

OpenAI. (2019). Dota 2 with large scale deep reinforcement learning. https://cdn.openai.com/dota-2.pdf

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., … Silver, D. (2019). Grandmaster level in starcraft II using multi-agent reinforcement learning. Nature, 575(7782), 350–354. https://doi.org/10.1038/s41586-019-1724-z

Hashimoto, D. A., Rosman, G., Rus, D., & Meireles, O. R. (2018). Artificial Intelligence in surgery: Promises and perils. Annals of Surgery, 268(1), 70–76. https://doi.org/10.1097/sla.0000000000002693

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2016). Safe, multi-agent, reinforcement learning for autonomous driving. arXiv preprint arXiv:1610.03295.

Graesser, A. C., Hu, X., & Sottilare, R. (2018). Intelligent tutoring systems. In International Handbook of the Learning Sciences (pp. 246-255). Routledge.

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is not a spectator sport. Proceedings of the Second (2015) ACM Conference on Learning @ Scale. https://doi.org/10.1145/2724660.2724681

King, D. L., & Delfabbro, P. H. (2019). Video game monetization (e.g., 'loot boxes'): A blueprint for practical social responsibility measures. Computers in Human Behavior, 101, 104-112. https://doi.org/10.1007/s11469-018-0009-3

Chaudhri, V. K. (2025, June 16). Truly intelligent ai could play by the rules, no matter how strange. Scientific American. https://www.scientificamerican.com/article/truly-intelligent-ai-could-play-by-the-rules-no-matter-how-strange/

Andrei, M. (2025, March 10). Ai is willing to lie, cheat, and manipulate to win. now what?. ZME Science. https://www.zmescience.com/future/ai-lies-cheats-chess/

Melhart, D., Togelius, J., Mikkelsen, B., Holmgård, C., & Yannakakis, G. N. (2023). The ethics of AI in games. arXiv preprint arXiv:2305.07392. https://doi.org/10.48550/arXiv.2305.07392

Sadler, M., & Regan, N. (2019). Game changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI. New in Chess.

Pietrołaj, M., & Blok, M. (2024). Resource Constrained Neural Network training. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-52356-1