Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

Key Applications of AI Question-Answering Systems: Research and Technical Analysis Based on Game Text and Dialogue

Qineng Wu *

Manchester Metropolitan Joint Institute, Hubei University, Wuhan, 430062, China

* Corresponding Author Email: 202231123002051@stu.hubu.edu.cn

Abstract. With the rapid development of the global gaming industry, artificial intelligence (AI) applications in games have garnered increasing attention. In role-playing and open-world games, particularly, the intelligence of non-Player Characters' (NPCs) dialogue systems is critical to enhancing player experience and interactivity. This paper investigates the applications of AI question-answering (QA) systems in gaming contexts by analyzing mainstream algorithms and models, their implementation workflows, advantages, and limitations. Additionally, it explores the challenges and future directions for AI-driven QA systems. The findings demonstrate that AI QA systems significantly improve game immersion and interactivity, particularly in dynamic dialogue generation and personalized narrative design. However, computational resource constraints and real-time performance remain major challenges. Future research should focus on reducing computational overhead, improving semantic accuracy, and enabling flexible adaptation of AI QA systems across diverse game genres. This study provides innovative insights for game developers and offers implications for extending AI applications to other domains.

Keywords: AI Question-Answering Systems; NPC Dialogue Generation; Video Games; Language Models.

1. Introduction

In recent years, the global gaming industry has experienced explosive growth. According to Statista, the global video game industry is a billion-dollar business and has been for many years. In 2024, the revenue from the worldwide gaming market was estimated at almost 455 billion U.S. dollars [1]. As the industry rapidly evolves, particularly amid the increasing integration of domestic and international markets, the development of AAA titles is no longer dominated solely by foreign studios. Chinese game developers are emerging as significant contributors, demonstrating substantial potential and influence. Within this context, artificial intelligence (AI) technologies, especially AI question-answering (QA) systems, are becoming pivotal tools for enhancing game interactivity and immersion. These systems significantly improve the intelligence of virtual characters, enabling richer and more dynamic interactions with players, thereby elevating both player experience and game quality.

Across genres—whether cooperative puzzle-adventure games, role-playing games (RPGs), or open-world action-adventure titles—dialogue and narrative exploration remain central to player engagement. Critically acclaimed titles such as Red Dead Redemption 2, The Witcher 3, and It Takes Two, all recipients of The Game Awards (TGA) "Game of the Year," exemplify how interactions with non-player characters (NPCs) and narrative depth enhance gameplay appeal and immersion.

However, the application of machine learning (ML) in various game designs does not always include the generation of NPC dialogue [2]. This gap underscores the need to focus on ML-based autonomous decision-making models for NPCs, such as dynamic dialogue generation in RPGs. Currently players' conversations with NPCs are highly scripted: in a typical scenario players must select from a set of preset responses that they can give to the NPC [2]. These limitations constrain player agency, failing to deliver truly immersive experiences. Advances in ML and natural language processing (NLP), particularly through large-scale language models (e.g., GPT-4, BERT), now enable NPCs to comprehend and generate natural dialogues in real time, adapting to player needs and emotional cues. This breakthrough has markedly enhanced game immersion and interactivity. AI technologies are increasingly leveraged to optimize NPC behavior in RPGs and adventure games, driving genre diversification. Furthermore, video games have many characteristics, which make them popular in the process of AI reinforcement learning research [3]. Optimized AI models can subsequently inform new game development, fostering a mutually reinforcing cycle.

Existing studies have explored diverse AI applications in gaming, including NPC dialogue generation, affective computing, and personalized quest guidance. Yet, research on AI QA systems' specific roles in game text and dialogue remains limited. This study aims to address this gap by analyzing how AI QA systems enhance gaming experiences and industry innovation. The technical frameworks, encompassing NLP, dialogue generation models, affective computing, and knowledge graphs, are introduced. Current challenges are discussed, and future directions for application scenarios are outlined.

2. Core Technology Introduction and Analysis

This section focuses on AI QA technologies commonly used in game development, detailing their descriptions, implementation workflows, strengths, and limitations. A comparative analysis of these technologies provides foundational insights into the technical challenges and opportunities of integrating AI systems into interactive gaming experiences.

2.1. BM25 Algorithm

The core workflow of AI QA systems typically involves four critical stages. First, the system receives a user-input natural language query and performs preprocessing steps, including stopwords removal, stemming, and lexical normalization. Next, leveraging large-scale pre-trained language models, the system conducts semantic understanding and contextual analysis to extract key information and latent intent from the input. Subsequently, a prebuilt retrieval model calculates relevance scores between query terms and candidate documents (e.g., game narrative databases created by developers) based on statistical properties of the document corpus. The system then parses critical information from top-ranked documents using an answer extraction module, integrating predefined semantic matching rules or deep learning models to generate structured responses. Finally, grammatical validation and readability optimization are applied to the output before delivery, completing the QA interaction loop. A word with a high BM25 score reveals its uniqueness in the corpus, and this method has been widely adopted in traditional learning to rank tasks [4]. Consequently, BM25 is particularly suited for QA-related information retrieval functions.

The BM25 algorithm operates on statistical relevance ranking principles, combining term frequency (TF), inverse document frequency (IDF), and synonym libraries to refine text matching. As depicted in Figure 1, it preprocesses by removing stopwords and standardizing synonyms, computes TF and IDF values, and calculates similarity scores to retrieve the top N answers.

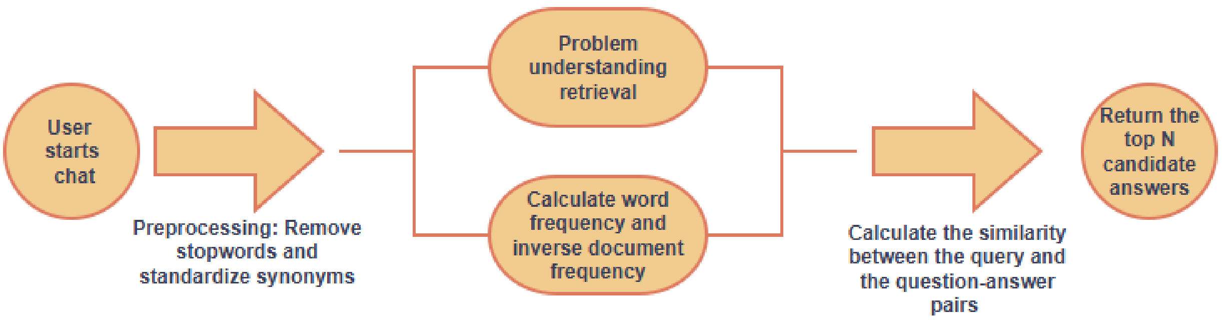

Figure 1. BM25 Algorithm Application Process (Picture credit: Original)

BM25 demonstrates notable advantages in QA retrieval tasks. By relying on TF-IDF-based statistical metrics, it delivers rapid results (average response time < 50 ms) with low memory consumption, ideal for real-time applications. Its transparent computational logic enables manual optimization through synonym library adjustments, making it effective for rule-driven tasks such as in-game tutorials. However, BM25's limitations stem from its lexical dependency: it fails to capture semantic relationships, leading to inaccuracies for context-dependent queries. In addition, in processing lengthy texts, there is a potential risk of overemphasizing high-frequency words while neglecting key information.

2.2. DrQA Retriever Algorithm

The field of AI reading comprehension has evolved significantly through iterative upgrades to large-scale datasets and algorithmic models, enabling faster and more precise extraction of relevant answers from document repositories in response to diverse user queries. The development of DrQA (Document Retriever Question Answering) marked a milestone in this domain, establishing a robust encyclopedia-style QA system [5]. DrQA Retriever is a multi-document retrieval system based on the TF-IDF model. It calculates text relevance using TF and IDF, replacing inverted indexing with hash tables to map vocabulary to fixed-length IDs, thus reducing memory usage. The system employs a bag-of-words (BoW) approach for efficient vectorization, enabling fast retrieval across millions of documents.

DrQA Retriever demonstrates significant advantages in efficiency and scalability. Its hash indexing architecture enables rapid processing of million-scale documents (e.g., the extensive mission dialogues in *Cyberpunk 2077*) with lower memory consumption than conventional methods. Additionally, stemming and TF-IDF weighting enhance cross-document information matching. However, the BoW model's disregard for word order and syntactic structure limits its ability to resolve critical semantic distinctions. Furthermore, the system struggles with contextual understanding, failing to address coreference resolution (e.g., ambiguous pronouns like "it" or "that location") or complex queries requiring dialogue history integration.

As shown in Figure 2, the workflow includes four phases: text preprocessing (stemming, stopwords removal, and normalization), hash indexing to create a term-document inverted index, TF-IDF vectorization to generate sparse vectors, and cosine similarity to rank documents by relevance. The pipeline uses hashing and parallel computing for fast retrieval across large corpora.

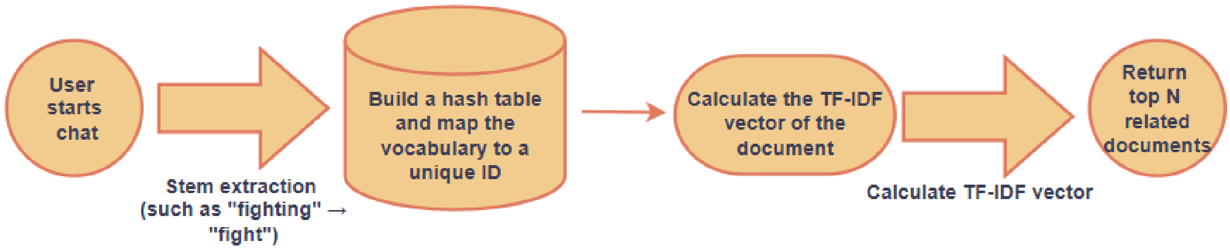

Figure 2. DrQA Model Application Process (Picture credit: Original)

2.3. Bidirectional Encoder Representations from Transformers (BERT) for Reading Comprehension

BERT, a groundbreaking pre-trained language model in NLP, introduced bidirectional Transformer encoders to achieve context-sensitive deep semantic representation learning. Its core mechanism utilizes multi-head self-attention layers to analyze long-range dependencies between queries and text passages, combined with fully connected networks to predict the start and end positions of answers within source texts. It has the ability to read text in both directions, which enables it to capture contextual information from both the preceding and following words, allowing it to excel in understanding the precise context and the relationships between words [6]. In recent years, BERT has been considered a state-of-the-art model in deep learning-based NLP [7] and has been used in the development of AI systems for natural language understanding and processing tasks.

BERT is pre-trained to enable prediction of missing words /masked words within sentences [6], while its reading comprehension ability is enhanced by a joint pre-training strategy that combines masked language modeling and next-sentence prediction, enabling the model to capture global contextual semantics. By employing a self-attention mechanism and transfer learning, BERT outperformed contemporary models in several language-understanding evaluation benchmarks for NLP downstream tasks [7]. In gaming applications, BERT can interpret long-distance relationships between player queries and in-game text, such as cross-level narrative clues, enabling accurate answer retrieval within complex storylines and quests, and allowing AI systems to respond to diverse game scenarios.

BERT excels at modeling complex semantic relationships, significantly outperforming traditional methods in long-text answer localization. Its multimodal extensibility expands potential applications. However, its large parameter size requires substantial GPU memory for inference, hindering mobile deployment, while fine-tuning demands extensive annotated data and computational resources, leading to high training costs.

As shown in Figure 3, BERT's workflow involves four phases: input processing (concatenating queries and text with positional encodings), generating context-aware token embeddings through multi-layer Transformer encoders, predicting token probabilities for answer span boundaries, and selecting the highest-probability start-end pair, with mechanisms to merge answers across multiple passages.

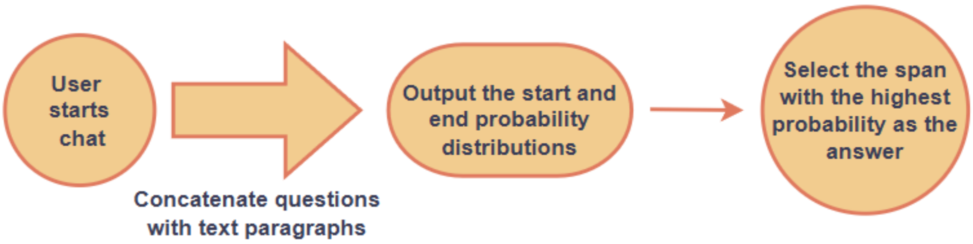

Figure 3. BERT Method Application Process (Picture credit: Original)

2.4. Domain-Adaptive BERT Model

Domain-adaptive BERT, an enhanced model based on the pre-training-fine-tuning paradigm, achieves semantic adaptation through domain-specific training and context-aware attention enhancement. Inheriting BERT's bidirectional Transformer architecture, the model undergoes two-phase optimization for target domains (e.g., gaming scenarios). First, incremental pre-training is performed on domain-specific corpora (e.g., game scripts, mission dialogues) to expand the vocabulary with domain terminology, building upon general corpus pre-training. Second, a dynamic attention masking mechanism is introduced to strengthen attention weights for critical entities (e.g., quest objectives, character names), thereby enhancing domain-specific contextual modeling. The embeddings from domain-adapted model (BERT-domain) can capture higher semantic and syntactic meanings of a word for the target domain [8].

As shown in Figure 4, the workflow includes five phases: data preparation (collecting domain texts and building a terminology database), domain adaptation (embedding high-frequency terms into BERT's vector space), attention optimization (prioritizing domain entities and keywords), and iterative learning (refining the model using player feedback and supporting multimodal inputs).

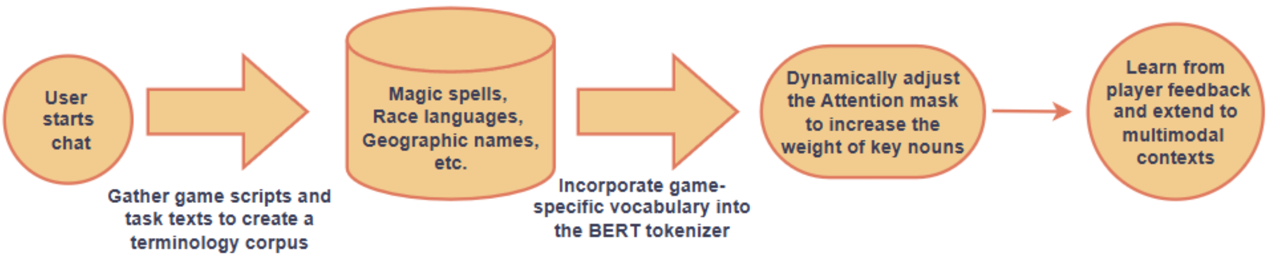

Figure 4. Domain-adaptive BERT model Application Process (Picture credit: Original)

This technology offers three main strengths: deep semantic understanding for parsing complex instructions and nested statements, strong domain adaptability for integrating specialized terminology and game lore, and the ability to process long texts (up to 512 tokens). However, limitations include latency issues due to computational complexity, high GPU memory requirements (12-16GB), which hinder deployment on edge devices, and dependence on extensive annotated data during fine-tuning, which exacerbates cold-start problems.

3. The Application of AI Question-Answering Systems in Games

With the rapid development of open-world games and immersive narrative design, player demand for intelligent interactive experiences continues to escalate. The proliferation of AI technologies and the global deployment of AI-powered systems have propelled gaming to unprecedented heights. Historically, games have served as an invaluable testbed for AI research while simultaneously driving advancements in gaming itself. From early traditional games like cards and chess to today's expansive video game market, AI has been indispensable. Industry-leading IT companies, such as DeepMind (Google), Facebook AI, and OpenAI, have successfully addressed problems in game strategy that were previously considered significant milestones for artificial intelligence, such as Go, Dota 2, and StarCraft, leveraging their abundant computational power and extensive human resources [9]. During the initial surge in ChatGPT's popularity, enthusiasts not only engaged in text or voice interactions with AI but also explored its integration into gaming. By enabling AI QA systems to generate contextually relevant dialogues in real time based on player queries or actions, these systems infuse interactions with novelty, allowing players to communicate with virtual characters more naturally. Consequently, such projects remain highly popular to this day.

3.1. AI Dungeon

AI Dungeon uses generative pre-trained models (e.g., GPT-3) and NLP technologies to enable dynamic text generation. The AI QA system provides three key functionalities: First, it generates new content by synthesizing contextual information, rather than simply retrieving predefined text. Second, the model ensures narrative coherence across multiple dialogue turns, maintaining logical consistency in the plot. Third, its low entry barrier and adaptability allow the system to handle ambiguous or casual inputs (e.g., short chat-like commands) and generate plausible storylines, enhancing player engagement.

The AI Dungeon workflow is as follows: A player inputs a narrative prompt (e.g., "You are a knight embarking on an adventure"), which the system processes through tokenization and semantic analysis. Using GPT-based models, the system generates coherent narrative content in real time. Players can provide new commands, prompting the model to adjust the plot, creating a personalized storytelling experience. With infinite creative freedom and immersive interactivity, AI Dungeon encourages user-driven content creation, where players share and refine stories, extending the game's lifespan. The game's duration and entertainment value depend entirely on player imagination, rather than prewritten developer narratives [10].

3.2. Google DeepMind SIMA

Google DeepMind's Scalable Instructable Multiworld Agent (SIMA) represents a cross-game universal AI companion, integrating multimodal learning with reinforcement learning. Built upon the Gemini model, SIMA processes multimodal inputs (language + vision) to generate action decisions and incorporates physical action modules (e.g., object manipulation, navigation) for adaptability in complex environments. SIMA's success relies heavily on AI QA system technologies, manifested in three key aspects: Firstly, natural language instructions enable players to collaborate with SIMA as with human teammates, enhancing strategic flexibility and immersion. Secondly, the AI QA system parses player commands via NLP, translating them into in-game operational sequences. Thirdly, attention mechanisms allow SIMA to preserve contextual coherence across multi-round interactions, ensuring continuity in task execution.

3.3. Popular Mods for Various Open-World or RPG Games

In popular *Red Dead Redemption 2* modifications (MODs), AI systems govern NPC behaviors and enable interactive dialogues with players. NPCs dynamically adjust their responses based on player actions—such as proximity or threatening gestures—while providing differentiated feedback to moral choices (e.g., altruistic vs. antagonistic behaviors). This adaptive interaction enhances role-playing authenticity, immersing players in a perceptibly "living" game world.

Similarly, AI-driven MODs for *Elden Ring* leverage QA systems to generate semantically coherent narrative content in real time, including mission texts and NPC dialogues. For instance, the system dynamically spawns mage-specific side quests when detecting player equipment (e.g., staff-wielding builds). These lightweight AI models prioritize response speed while minimizing hardware resource consumption, ensuring compatibility with diverse hardware environments of player-created MODs.

4. Technical Challenges and Future Prospects

4.1. Challenges

The exploration of AI QA systems in gaming inevitably confronts multifaceted challenges. Despite significant progress in AI and Large Language Models (LLMs), commercially developed LLMs often lack intricate details [11], which has a critical impact on user engagement and utilization in games. Many commercial LLMs exhibit opaque decision-making processes and limited interpretability, leading to inconsistent performance in tasks requiring multi-step reasoning. Additionally, the NLP component of generative AI also faces certain challenges in parsing for syntactic correctness [12].

Multimodal integration presents additional hurdles, as balancing architectural coupling and real-time responsiveness remains a key challenge. Combining BM25's efficient retrieval, DrQA's cross-document processing and BERT's deep semantic understanding necessitates cascaded architectures (e.g., BM25 pre-screening → DrQA filtering → BERT refinement). However, data format conversions between modules and cumulative latency may push overall response times beyond player tolerance thresholds. Scalable deployment of game AI QA systems is further constrained by computational costs and real-time bottlenecks. For instance, training GPT-4 requires tens of thousands of GPU hours and exorbitant expenses, rendering it inaccessible to small-to-medium studios and stifling innovation. Considering memory load, server load, and other related factors, using Google's Gemini API and Sentence BERT framework, the similarity between tasks generated by NPCs and player interactions with NPCs is only 62.29% [13].

4.2. Prospects

Future advancements in AI QA systems for gaming will focus on six key directions. Firstly, integrating DrQA's contextual retrieval with BERT's affective generation capabilities could enable NPC systems with persistent memory and personality traits, where AI generates contextually divergent responses based on player actions (e.g., donating items or attacking allies), dynamically reflecting changes in NPC affinity. Secondly, hybrid retrieval-generation architectures will emerge as core components for open-world games, leveraging BM25 to rapidly retrieve task templates aligned with player states (e.g., equipment, level) and BERT to populate narrative details (e.g., NPC dialogues, scene descriptions). Thirdly, multilingual BERT models combined with domain-adaptive fine-tuning could support low-latency cross-lingual QA systems, mitigating cross-cultural semantic conflicts through domain-specific knowledge injection. Fourthly, multilayered content moderation frameworks—incorporating rule-based engines and sensitive lexicons (e.g., avoiding culturally taboo terms like "blasphemy" in medieval Western contexts) will be critical to real-time suppression of non-compliant content. Fifthly, hybrid architectures will reduce development costs by minimizing manual scripting efforts for dialogues and game narratives. Finally, synergistic integration of traditional AI and generative AI will unlock new potentials: while traditional AI excels at well-defined tasks like data analysis and automation, generative AI addresses complex problems through unprogrammed solutions. Traditional AI is effective at handling well-defined tasks like data analysis and automation, while generative AI excels in more complex problem-solving, often generating solutions that were not explicitly programmed [14].

5. Conclusion

This study explores the application of AI QA systems in gaming, discussing various mainstream algorithms and models, and presenting their application workflows along with an analysis of their respective advantages and disadvantages. The paper also examines the challenges associated with applying AI QA systems in games and outlines the future development of potential application scenarios. The findings of this research are of significant importance for future game development, particularly by providing new insights into the intelligent behavior of virtual characters and the generation of natural dialogue. With the introduction of more advanced AI technologies, game developers will be able to create more realistic and expressive virtual worlds, offering players a richer gaming experience. However, the study also has certain limitations, especially in terms of the computational resource requirements of the models and the optimization of real-time performance. Additionally, the flexible application of AI QA systems across different types of games will be a key direction for future development. Through this research, support is provided for the intelligent development of the gaming industry and the enhancement of player experience.

References

[1] J. Clement, Video game industry - Statistics & Facts, Statista. Published on Nov. 6, 2024. Retrieved on Apr. 4, 2025. Retrieved from: https://www.statista.com/topics/868/video-games/.

[2] R. Volum, S. Rao, M. Xu, G. DesGarennes, C. Brockett, B.V. Durme, O. Deng, M. Akanksha, B. Dolan, Craft an Iron Sword: Dynamically Generating Interactive Game Characters by Prompting Large Language Models Tuned on Code. In *Proceedings of the 3rd Wordplay: When Language Meets Games Workshop (Wordplay 2022)*, pages 25 – 43, Seattle, United States (2022).

[3] Y. He, Inside and Outside of The Game—The Dilemma of Reinforcement Learning in AI. *CAAI Transactions on Intelligent Systems*, 17 (2): 220 - 220 (2022).

[4] M.Y. Kim, J. Rabelo, K. Okeke, R. Goebel, Legal Information Retrieval and Entailment Based on BM25, Transformer and Semantic Thesaurus Methods. *Rev Socionetwork Strat* 16, 157 – 174 (2022).

[5] D. Chen, A. Fisch, J. Weston, A. Bordes, Reading Wikipedia to Answer Open-Domain Questions. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1870–1879, Vancouver, Canada (2017).

[6] A. Babu, S.B. Boddu, BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding. *Exploratory Research in Clinical and Social Pharmacy*, vol. 13, 100419 (2024).

[7] Y.J. Kim, J.H. Kim, Y.M. Kim, S. Song, H.J. Joo, Predicting medical specialty from text based on a domain-specific pre-trained BERT. I*nternational Journal of Medical Informatics*, vol. 170, 104956 (2023).

[8] M.H. Syed, S.-T. Chung, MenuNER: Domain-Adapted BERT Based NER Approach for a Domain with Limited Dataset and Its Application to Food Menu Domain. *Applied Sciences*, 11 (13), 6007 (2021).

[9] S. Risi, M. Preuss, From Chess and Atari to StarCraft and Beyond: How Game AI is Driving the World of AI. *Künstliche Intelligenz* 34, 7 – 17 (2020).

[10] L.F. Xu, T. Zuo, AI-based Open-World Text Adventure Game: A Case Study of AI Dungeon. Publishing *Reference*, (2), 19 – 23 (2021).

[11] A.J. Jinia, K.L. Chapman, S. Liu, C. Della Biancia, A. Li, J.M. Moran, Challenges in Developing an AI-Based Analysis System for Incident Learning. *International Journal of Radiation Oncology*Biology*Physics*, 120(2, Supplement), e542 (2024).

[12] A. Håkansson, G. Phillips-Wren, Generative AI and Large Language Models - Benefits, Drawbacks, Future and Recommendations. *Procedia Computer Science*, 246, 5458 – 5468 (2024).

[13] J.P.W. Hardiman, D.C. Thio, A.Y. Zakiyyah, Meiliana, AI-powered dialogues and quests generation in role-playing games using Google's Gemini and Sentence BERT framework. *Procedia Computer Science*, 245, 1111 – 1119 (2024).

[14] D.T.K. Ng, E.K.C. Chan, C.K. Lo, Opportunities, challenges and school strategies for integrating generative AI in education. *Computers and Education: Artificial Intelligence*, 8, 100373 (2025).