

Generative AI Meets Adventure: Elevating Text-Based Games for Engaging Language Learning Experiences

Boaventura DaCosta 

Solers Research Group, Orlando, FL, USA
Email: bdacosta@solersrg.com

How to cite this paper: DaCosta, B. (2025). Generative AI Meets Adventure: Elevating Text-Based Games for Engaging Language Learning Experiences. *Open Journal of Social Sciences*, 13, 601-644.
<https://doi.org/10.4236/jss.2025.134035>

Received: March 16, 2025

Accepted: April 27, 2025

Published: April 30, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This work explores generative AI-enhanced text-based games for language learning. Utilizing ChatGPT, the study modifies an early version of *Colossal Cave Adventure* to generate contextually appropriate responses, making the game more interactive and engaging. The method involved creating a tailored model to emulate the original game, followed by iterative refinements. Play-through-based comparisons revealed that the model successfully reproduced the game's core mechanics while introducing dynamic content and contextual hints, enhancing the overall gaming experience. At the same time, the model's nondeterministic nature showed benefits in the context of language generation while introducing challenges, such as inconsistencies in the progression of language complexity. Altogether, generative AI holds promise for improving educational, text-based games by increasing engagement and adaptability; thus, educators are encouraged to consider AI-enhanced gaming as an innovative tool for language instruction and interactive learning. However, further research is also needed to address AI limitations to ensure learning consistency, alongside the value of exploring broader applications beyond language acquisition.

Keywords

Artificial Intelligence, ChatGPT, Colossal Cave Adventure, Generative AI, Interactive Fiction, Language Learning, Large Language Models, Text-Based Games

1. Introduction

Text-based games, which gained popularity in the 1970s and 1980s with titles like

Colossal Cave Adventure (CCA) (Crowther & Woods, 1975-1977) and *Zork* (Infocom, 1977-1982), rely on written natural-language commands (Ammanabrolu & Riedl, 2019; Côté et al., 2019; Dambekodi et al., 2020; Lyömiö, 2017; Madotto et al., 2020; Sharma, 2008; Yao et al., 2021) (i.e., “actions”; Towle & Zhou, 2023: p. 1) that are entered into a command line interface (Allington, 2015; Côté et al., 2019). Due to the technological limitations of the time (Kim, 2018), they used plain text to describe the environment, with player interaction occurring through simple text commands, such as “go north” (Côté et al., 2019; Nagai & York, 2021; Sargsyan, 2013). Namely, early versions of these games did not use graphics beyond ASCII to depict scenes (Kim, 2018). Thus, it is to be expected that their popularity diminished (Kirill, 2024; Lyömiö, 2017; Sharma, 2008; Wright & Weible, 2024) with the advent of games featuring graphics (Kim, 2018; Kirill, 2024; Lyömiö, 2017; Wright & Weible, 2024), animation, and sound (Kim, 2018).

Notwithstanding their decline, these single-player games (Kostka et al., 2017) continued to be fondly remembered (Jerz, 2007; Kim, 2018) and, in recent years, have experienced a resurgence (Jerz, 2007; Kim, 2018; Kirill, 2024). Among the different reasons offered (Kim, 2018), they have been explored for their potential to help improve linguistic skills (Reinders & Wattana, 2015; Wright & Weible, 2024) by engaging learners through actual text that requires them to use language to interact with the game environment (Kim, 2018; Pereira, 2013). It has been asserted, for instance, that these games can encourage learners to participate in reading activities (Manuaba, 2017; Pereira, 2013), helping with comprehension (Manuaba, 2017; Sargsyan, 2013). This is important, as Marková (2023), citing Harmer (2015), states, as one of the four core skills in second language (L2) acquisition, reading plays a significant role in enhancing different linguistic areas, such as vocabulary, grammar, and spelling.

The use of these games is, in part, predicated on the idea that language is learned through practical use, which Pereira (2013) explains aligns with L2 acquisition theories. Learners can encounter words and phrases during gameplay that are not commonly used in everyday conversation or that they may not be familiar with, broadening their lexical scope (Pereira, 2013). Moreover, they need to grasp the meaning of almost all words to advance, reflecting their understanding (Pereira, 2013) and engagement with the language (Kim, 2018; Pereira, 2013). Basically, because learners interact with these games through text, a solid understanding of language is essential for successful gameplay (Yao et al., 2020). So, the fact that learners can progress serves as evidence of their comprehension (Pereira, 2013).

However, Marková (2023), citing Nassaji (2011), explains that reading is a complex process that involves the “retrieval of different kinds of knowledge from different types of memory, decoding of the visual stimuli, syntactic parsing, semantic processing, making sense of discourse, and incorporating all this acquired information into one’s preexisting knowledge” (p. 13). Namely, reading is not a natural skill learned through exposure (Marková, 2023). This is an argument that has been supported by others, such as Furtado et al. (2018), who affirm that the number of

words learned through exposure may lead to the misconception that inferring the meaning of a word from context is a simple task, instead highlighting that learners frequently struggle to extract contextual information when encountering unfamiliar vocabulary. Citing Hulstijn (1992), Hulstijn et al. (1996), and Laufer and Sim (1985), they add that this includes incorrectly inferring the meaning of words (Furtado et al., 2018).

Marková (2023), citing Pope (2010), does acknowledge that such new mediums seem to be exciting to learners but curbs this enthusiasm by stating that the narratives are often “disjointed and perplexing” (p. 34). They require considerable effort to understand and keep up with, possibly explaining why they have not become mainstream, remaining popular only among dedicated supporters (Marková, 2023). Although Marková offers this in the context of e-books with interactive features, something similar can be said of text-based games, despite claims that they are more straightforward to understand and play (e.g., Pereira, 2013). For example, the objectives in these games are rarely obvious, uncoverable only through exploration (Yuan et al., 2018), including comprehending complex navigation stemming from their expansive, looping layouts (Lyömiö, 2017) and the lack of on-screen markers, in-game maps, and other visual and audio cues commonly found in many modern-day games. Then there is the issue that text-based games have relatively linear plots (Ammanabrolu & Riedl, 2019), and their scripted text is static and, thus, can be predictable (Kirill, 2024), a problem commonly found in early, pure-text versions. Granted, nondeterministic instances of these games exist (Kostka et al., 2017) that support a myriad of paths and endings (Zelinka, 2018), but the same decisions (i.e., identical actions performed in the same situations) in subsequent playthroughs result in the same text descriptions (Kirill, 2024; Zelinka, 2018).

Recent advancements in generative artificial intelligence (AI) might offer solutions to these and similar challenges that impact the efficacy of text-based games in language-learning contexts. Large language models (LLMs) have made significant contributions to the study of natural language processing and human-computer interaction (Lanzi & Loiacono, 2023) as they are able to accept written text input, producing human-like responses (Harshitha, 2022). Lanzi and Loiacono (2023), for instance, explain that in addition to creating random text from high-level instructions, these models can also mix and create varying text from existing content, all the while keeping the meaning clear. In simpler terms, they can generate dynamic and varied responses from identical inputs (Kirill, 2024), affording learners not only the freedom to craft their text responses but the responses that they can receive can be nearly limitless (Harshitha, 2022).

This ability to adaptively create content has opened opportunities to explore personalized learning (Jegade, 2024; Office of Educational Technology, 2023; Terzic et al., 2023) by tailoring instruction to individual learner needs (Atlas, 2023; Jegede, 2024; Terzic et al., 2023), which includes improving language skills (Office of Educational Technology, 2023). As a result, there is substantial interest in gen-

erative AI in educational settings (Nash, 2024; Office of Educational Technology, 2023; Wei, 2023) and its ability to be leveraged effectively as a teaching tool (Office of Educational Technology, 2023; Terzic et al., 2023). When applied to text-based games—what Harshitha (2022) refers to as “[AI] powered” (p. 44)—learners can enter their text and receive personalized responses, creating an almost infinite range of possibilities to explore, making these games more captivating and engaging compared to their nonAI-based predecessor (Harshitha, 2022). This enhanced form of text-based game could create text that is adapted to each learner’s needs while also assisting them with the content in the form of hints and insights as they progress during gameplay (Kim, 2018).

1.1. Purpose

Given the potential of AI, this study examines AI-powered, text-based games, focusing on their ability to adaptively generate content while also addressing challenges that affect the efficacy of their use in language learning. Specifically, ChatGPT, the LLM developed by OpenAI (2025a), is used to emulate a modified version of *CCA* designed to generate increasingly complex and appropriate responses while providing learners with contextual feedback intended to foster an interactive and engaging environment.

This work builds upon the insights of others, such as Sharma (2008), who states that language cannot be entirely learned through gameplay but should instead complement other experiences learners are exposed to, including the language learning process; Rankin et al. (2008), who point out that the creation of inclusive learning environments for learners with different language abilities is a challenging endeavor for educators, highlighting the need for L2 didactic support not only in the classroom but also in informal learning settings; and Kim (2018), asserting that the advantages and limitations of text-based games, both as learning processes and outcomes, are not thoroughly documented, stressing the urgent need for more comprehensive studies. Thus, this investigation is anticipated to be of value to educators, practitioners, and researchers, as such knowledge could help enhance learning outcomes (Kim, 2018; Wright & Weible, 2024).

It is necessary to state that although this research relies on scholarly insights, it does not claim to be a comprehensive analysis but rather is intended to encourage further research and discussion on the educational benefits and drawbacks of AI-powered, text-based games. Educators and researchers, for example, can replicate the approach outlined herein to develop their own text-based game using generative AI for exploration and potential use in and outside of the classroom, including its use as a supplementary tool for language learning.

Moreover, as observed in the source code of the GitHub package (wh0am1-dev, 2015) used in this work and discussed later, the original *CCA* displayed text to players in uppercase. However, it is worth noting that early screenshots of the game interface can be found that show the text in sentence case, suggesting there are likely variations (across different versions of the game) in how the text was

displayed. Despite this, the uppercase convention is maintained herein to distinguish it from the corresponding text generated by ChatGPT. What this also means is that this study contains AI-generated content as comparisons are made between the text-based descriptions provided by the original game and the AI-powered version.

Lastly, the findings found herein were shared with educational researchers and practitioners, whose valuable feedback aided in improving and refining this work. This includes feedback acquired at academic conferences (e.g., DaCosta, 2025).

1.2. Terminology

Prior to delving into the related research, there are terms used herein that need clarification. Will Crowther created “Adventure” for the PDP-10 mainframe (Falcó & Colás, 2022; Jerz, 2007; Montfort, 2005) written in Fortran (Falcó & Colás, 2022; Jerz, 2007; Lyömiö, 2017) (internally titled as “Adventures” [Jerz, 2007], and abbreviated “Advent” [Jerz, 2007; Lyömiö, 2017]) prior to 1977 (Jerz, 2007; Jerz & Thomas, 2015), with some sources citing 1976 (e.g., Jain, 2019; Montfort, 2005) and 1975 (e.g., Dong, 2023), and some simply noting multiple years (1975-1976; e.g., Falcó & Colás, 2022; Sharma, 2008). It was after Don Woods enhanced the game (see Jerz, 2007, for an explanation of these enhancements) that it became known as “Colossal Cave Adventure” (Jerz, 2007), with some sources referring to it as “Colossal Cave” (Montfort, 2005). Despite variations in name, release dates, and versions (that are said to exist, e.g., Jerz, 2007; Sharma, 2008), and the fact that the GitHub package used in this work is likely the original *Adventure*, the more widely recognized title, “Colossal Cave Adventure,” is used henceforth.

In addition, “text-based” games are often used interchangeably with terms like “interactive fiction” (IF) (Dambekodi et al., 2020; Harshitha, 2022; Hausknecht et al., 2020; Jain, 2019; Osborne et al., 2022; Tsai et al., 2023; Zelinka, 2018; Zhiwei & Yin, n.d.) and “text adventure” (Dambekodi et al., 2020; Tsai et al., 2023). Wright and Weible (2024) note that the term “text adventure” is now less common, with IF preferred in recent articles (e.g., Harshitha, 2022; Kim, 2018; Kostka et al., 2017; Lyömiö, 2017; Manuaba, 2017), even though the use of “text adventure” games can still be found (e.g., Ammanabrolu & Riedl, 2019; Yin & May, 2020). They cite Allington (2015), who put forth that “Videogames such as these were originally described not as IF but as ‘adventures’, ‘adventure games’, ‘text adventures’, or ‘text adventure games’,” and that all of these terms basically meant “games similar to Adventure” (p. 269). Harshitha (2022) perhaps offers one of the most straightforward explanations by describing text-based games as an early form of IF, while Kostka et al. (2017) label IF as a genre of text-based adventure. In other words, outside of IF—which emerged later to represent various forms of media, including books and games, that involve engagement and interaction through different actions (Europass SRL, 2023), and thus, can be viewed as a broad term (Allington, 2015; Kim, 2018)—many of the terms used are all synonyms of text-based games.

Given the importance of text in the current investigation and the use of the early pure-text adventure, *CCA*, for simplicity, the term “text-based games” will be used hereafter unless otherwise specified. They are defined in this research as games where written text (i.e., words) is the primary means of interaction between players and the game (Manuaba, 2017).

2. Related Work

It goes without saying that technology plays a prominent role in learning when it comes to innovative approaches, and learners are no longer strangers to the use of technology in educational settings (Darvenkumar & Devi, 2022), including the use of digital games. The immersive and interactive nature of these games has contributed to their widespread appeal, with consoles and personal computers only exacerbating their popularity (Reinders, 2016). These qualities can help engage players, encouraging them to invest countless hours in improving their playing abilities while advancing through different levels (Rankin et al., 2008). Hence, it is only natural that educators have been called to harness the interest in these games to make language learning more inviting and beneficial (e.g., Tafazoli, 2021). When applied to educational settings, for instance, it is said that they could serve as effective (Darvenkumar & Devi, 2022) and authentic (Rankin et al., 2006) learning environments that motivate learners (Darvenkumar & Devi, 2022), providing them with opportunities to develop and test their knowledge (Rankin et al., 2006), all along enabling them to engage dynamically with content (Manuaba, 2017).

The use of digital games as part of the learning process is by no means a new idea. Reinders (2016) cites Lee (1979) and Rixon (1981) as examples to drive the point that the study of games in foreign language teaching alone predates four decades. Then there is Reinders & Wattana (2015), who references several reviews (i.e., Hainey et al., 2011; Kirriemuir & McFarlane, 2004; Mitchell & Savill-Smith, 2004; Kostka et al., 2017) as evidence that these games have made substantial contributions to learning across various domains (Reinders & Wattana, 2015). Some have even gone as far as to contend that digital games are becoming increasingly recognized as practical instructional tools if not considered essential teaching methods (e.g., Darvenkumar & Devi, 2022).

However, text-based games are specialized forms of digital games with their unique attributes. Consequently, to comprehend their role in language acquisition, it is important first to understand what is known about the use of digital games in language learning contexts in general, including IF.

2.1. Digital Game-Based Language Learning

Digital game-based language learning (DGBLL)—an area of study where game elements are used designed to target the development of specific language skill or achieve an explicit language learning goal (Chowdhury et al., 2024), or, as Tafazoli (2021) puts it, “the application of games in language teaching/learning” (p. 33)—

has gained increasing popularity in past decades (Alyaz & Genc, 2016). Its beginnings are arguably connected to the general evolution of computer-assisted language learning (CALL) (Alyaz & Genc, 2016; Khatibi & Cowie, 2013; Scholz, 2015)—what Pereira (2013) characterizes as “the field of using computers specifically for language learning” (p. 23). What this means is that DGBLL is built upon an older appeal that focuses on the influence of play in language learning as well as general education, which, as discussed, includes foreign language teaching research going back decades (Reinders, 2016).

Despite its origins, DGBLL is nonetheless described as a relatively new field (e.g., Reinders, 2016; Scholz, 2015), with its popularity partly due to the successful implementations described in several studies (Alyaz & Genc, 2016). Reinders (2016), for example, reports that the effects of gameplay on language learning have shown positive effects on learners’ willingness to communicate, language socialization, and other factors. With significant research conducted on L2 learning (Dixon et al., 2022) (e.g., English as a Second Language and English as a Foreign Language; Rankin et al., 2006; Rankin et al., 2008) through the use of multi-user domains, multi-user object-oriented domains (Rankin et al., 2006; Rankin et al., 2008), as well as massively multiplayer online role-playing games (MMORPGs) (Dixon et al., 2022; Rankin et al., 2006; Rankin et al., 2008).

2.1.1. Transformative Benefits of Digital Games in Language Learning

Online games (Pereira, 2013), particularly MMORPGs, have garnered attention in L2 development contexts because they require players to possess and use multiple skills during gameplay (Scholz, 2015), including reading, writing, and listening, as players must communicate with other players online (Pereira, 2013; Scholz, 2015). According to Rankin et al. (2008), they cater to the L2 learning goals of reading comprehension, vocabulary acquisition, and conversational proficiency, which can be used as specifications for serious game design; that is, games that do more than entertain, fostering knowledge and skill acquisition through the incorporation of learning objectives with elements of play (Rankin et al., 2008). To play these games successfully (Pereira, 2013; Scholz, 2015), players need to focus on using the target language accurately and coherently to communicate and help accomplish game tasks (Rankin et al., 2006). Thus, parallels are believed to exist between the skills needed by language learners and game players, with both, for instance, actively navigating, processing, and shaping the gaming ecosystem through the use of language (Scholz, 2015).

This speaks to the social aspect of these games (i.e., online virtual worlds supporting interaction between thousands of concurrent players) in that players are, essentially, exposed to other languages as they communicate with native speakers of that language (Rankin et al., 2008). For example, those who join a guild or group quest in *World of Warcraft* (Blizzard Entertainment, 2004) must communicate with others in the target language to discuss game objectives and convey strategies, which is not limited to in-game discussions but instead could take place in external spaces (Scholz, 2015). This is important because, as Rankin et al. (2008) ex-

plain, cooperative gameplay—where native and nonnative speakers communicate with one another—is a significant influence on L2 acquisition. Moreover, since MMORPGs are often localized to support different linguistic and cultural audiences, players are not restricted to any single language. Instead, they have the option to experience the game in various foreign languages, gaining exposure to hundreds of hours of authentic content (Dixon et al., 2022).

Other positive effects include increasing learner engagement (Reinders, 2016), which claims that incorporating fun and entertainment into the learning process can boost interest in language learning (Darvenkumar & Devi, 2022). Citing Koo (2009), Scholz (2015) explains, for instance, that players of MMORPGs are driven by motivation through visually stimulating worlds that offer an abundance of opportunities. Extrapolated, games are believed to motivate learners to participate actively, and this increased participation is thought to enhance engagement and support learning (Darvenkumar & Devi, 2022). When learners lack motivation, their commitment to classroom activities falters, and this can significantly impact whether they take an active role in language learning (Zhang, 2023). So, it is hardly surprising that learner motivation is a common focus in DGBLL and game-based learning (GBL) conversations (e.g., Darvenkumar & Devi, 2022; Furtado et al., 2018).

Then, there is the effect of gameplay in helping with associated stresses, such as lowering anxiety (Reinders, 2016). When speaking from the perspective of GBL (e.g., Rankin et al., 2006; Zhang, 2023) and gamification (e.g., Zhang, 2023), it is said that these games and related components can be used in instructional approaches to create environments that are relaxing (Rankin et al., 2006; Zhang, 2023) and less stressful (Zhang, 2023). The pressure of learning a new language can be frustrating (Sharma, 2008) and draining, resulting in anxiety, negatively impacting mental health, and adversely affecting academic performance, thus necessitating innovative approaches that can promote learners' overall welfare (Zhang, 2023). Zhang (2023), for example, explains that by making minor adjustments in what they call vernacular games, that is, games that are not explicitly designed for language learning, educators can create a safe and fun way to develop language skills while reducing stress and anxiety found in traditional classrooms. Something similar has been said of MMORPGs as supplementary tools to aid in L2 development (e.g., Kongmee et al., 2011). Learners can see their use of language firsthand (Kongmee et al., 2011). Such environments can help challenge learners (Rankin et al., 2006; Zhang, 2023), providing immediate feedback (Kongmee et al., 2011; Rankin et al., 2006) and rewards (Rankin et al., 2006) while allowing them to maintain a quicker learning pace (Zhang, 2023).

2.1.2. Challenges and Reservations in Digital Game-Based Language Learning

With all that in mind, the field of DGBLL is no stranger to challenges and reservations, with assessments regarding their suitability in language learning contexts asserted to be early and preliminary (Reinders, 2016). This is, to some extent, due

to challenges in conducting the studies themselves in operational, pedagogical, and methodological contexts (Reinders, 2016). Reinders (2016) explains that it is difficult to control the novelty of games when they are introduced into the classroom. Then, there is the significant issue of data collection, as research on DGBLL can take place outside of formal settings (Reinders, 2016). It is also difficult to account for all the variables that can affect the learning outcome, given that digital games often supplement other existing content (Reinders, 2016).

Likewise, Scholz (2015) highlights that the absence of a specific theory for studying digital games for language learning is a significant concern, a result of the many ways in which DGBLL has been theorized and examined. Citing Hubbard (2009), Scholz agrees that something similar can be said of CALL, but CALL comprises research that has examined language learning using computers against numerous theoretical frameworks. This has left DGBLL at a disadvantage, with Scholz citing others, explaining that the research has been limited, focused on either game characteristics that are applicable to language learning (Gee, 2008; Sykes & Reinhardt, 2012) or on the self-reported perceptions of learners regarding the effectiveness of gameplay in L2 development (Peterson, 2012; Allen et al., 2014).

Even among such studies, there have been challenges, with research focused on perceptions towards DGBLL inconsistent, reporting both positive and negative attitudes by educators (Alyaz & Genc, 2016), essentially revealing conflicting views. Khatibi and Cowie (2013), citing Reinders (2012), explain that those who encourage the use of digital technology in education are guilty of hyperbolic idealism and, consequently, educators retort that digital games have not achieved effective classroom integration; instead, they promote entertainment over learning. At the same time, there is disagreement that games designed for L2 development cannot engage learners, at least not to the level of their entertainment counterparts (Dixon et al., 2022).

Scholz (2015) explains that the hesitancy stems, at least in part, from a generational shift, where educators have not explored games to the same degree they have with computers in general. This is especially the case when discussions include CALL, given that CALL is generally more accepted and the research more established (Scholz, 2015). Alyaz and Genc (2016) attempt to clarify the matter in that for educators to adopt GBL pedagogy into their classrooms, they need to experience games themselves, which they present could be done through training (Alyaz & Genc, 2016).

When attempts have been made to measure L2 development during gameplay, focus is often placed on the amount of vocabulary acquisition (Scholz, 2015). Reinders (2016) cautions that assumptions have been made that exposure and interaction with a target language during gameplay naturally leads to language proficiency growth and that there is little research available that explicitly focuses on language acquisition (see Reinders, 2016 for a review). This does not mean that games do not play a role in language learning and that they should be discredited,

as the gameplay can introduce learners to language (Scholz, 2015). Studies have shown that games influence affective aspects of language (Dixon et al., 2022; Reinders, 2016) and that meaningful L2 interaction can take place (Dixon et al., 2022), and this impact is somehow linked to language acquisition (Reinders, 2016; see Dixon et al., 2022 for a review). However, gameplay by itself (solely through exposure) to learn and reinforce language is insufficient (Scholz, 2015).

Altogether, conclusions have been made regarding the role of games in language education (Reinders, 2016), with DGBLL viewed as a novel shift in thinking on how to address the needs of educators and learners in today's digital world (Tafazoli, 2021). This includes the use of games both formally and outside of the classroom (Reinders, 2016), with online multiplayer games, such as MMORPGs, serving as "unorthodox second language (L2) pedagogical tools" (Rankin et al., 2008, p. 43). However, DGBLL is also still evolving as the understanding of its potential continues to expand alongside the growth of gaming (Scholz, 2015). This means that learning through digital games is not yet fully understood, including their role in language teaching and acquisition, and that significant gaps remain in the existing literature, with many important areas still unexplored (Reinders, 2016).

2.2. Interactive Fiction

According to Harshitha (2022), IF is one such area, describing it as "the use of a particular kind of narrative that doesn't follow such a static chronological flow" (p. 45). Presented as falling within the field of Digital Game-Based Learning (Shelton et al., 2008) and DGBLL (Pereira, 2013), these are "digital text-based adventure[s] designed with specific educational goals in mind" (Shelton et al., 2008: p. 128). As demonstrated, IF has been defined in a multitude of ways, with no single definition accepted (Ørnholt, 2023). However, it is said to be generally agreed upon (Kim, 2018; Ørnholt, 2023) that (at least within the context of games as compared to other forms of media) IF refers to "games that rely on text commands" (Kim, 2018: p. 5) and which are "text[-]based and [have] a story that develops based on reader/player input" (Ørnholt, 2023: p. 6).

Research into the use of IF in education has been thriving, examining commercial resources alongside those explicitly designed for education (Marková, 2023), including their role in language learning classrooms (Sargsyan, 2013). A review by Wright and Weible (2024) examined research conducted between 2012 and 2022 on text-based games, including IF and text adventures in education. The review categorized 63 peer-reviewed journal articles into five themes: writing, reading, history, technology, and special topics, with the last theme covering serious games in areas such as science, health and wellness, business, medicine, and real-world applications (Wright & Weible, 2024). What stands out is that, in addition to the breadth of coverage, seven articles specifically addressed reading literacy, highlighting positive effects like increased student engagement with reading materials (Wright & Weible, 2024). Interactive fiction has been described as being

aligned with the principles of L2 acquisition as well as communicative approaches (Pereira, 2013), with studies indicating that IF can be used effectively in second- and foreign-language teaching contexts, particularly when it comes to reading (e.g., Darvenkumar & Devi, 2022) and vocabulary (e.g., Neville et al., 2009) skill development (Sargsyan, 2013), to include demonstrating their ability to motivate learners (e.g., Manuaba, 2017).

2.2.1. Interactive Fiction in Reading and Vocabulary Growth

Text-based games use in-depth descriptions of settings, characters, and events to engage players while also allowing them to be part of the story (Ørnholt, 2023). This is achieved through extensive text and detailed accounts that players must read (Zhang, 2023), making the narrative a vital component, with players active participants in how the story unfolds (Harshitha, 2022). When coupled with their interactive and challenging features, these games can potentially provide a much more enjoyable reading experience (Pereira, 2013). Thus, it is unsurprising that scholars assert that these games (what they refer to as IF-based) can assist with extensive reading (e.g., Sargsyan, 2013; Thu & Nhu, 2021, citing Pereira, 2013).

Extensive reading is often associated with reading for pleasure (Kim, 2018), where readers are focused on the meaning of the text instead of any specific word or phrase (Thu & Nhu, 2021). Since this reading is a form of leisure, they have the freedom to select content that is important to them (Ørnholt, 2023), which is claimed to promote independent reading (Thu & Nhu, 2021). Thu and Nhu (2021), for instance, conducted a study that involved 40 first-year Vietnamese English major students using IF-based extensive reading materials in classroom activities, pair reading, and individual computer-based IF reading. They concluded that the participants reported a positive attitude towards IF, primarily when used as extensive reading in class, as it was engaging and increased their involvement in the reading process (Thu & Nhu, 2021). What is particularly interesting is that many of the participants found IF motivational, contributing to the habit of reading English for pleasure (Thu & Nhu, 2021). This is important, as Sargsyan (2013) notes, citing Lancy & Hayes (1988), that learners rarely engage in this kind of reading, making IF essential because it can motivate them to read interesting content beyond the classroom.

As presented, learner motivation is a common focus in discussions surrounding GBL (e.g., Darvenkumar & Devi, 2022; Furtado et al., 2018) and consistently appears as a significant factor in related research studies. Manuaba, 2017, for example, investigated the potential of text-based games to motivate and enhance reading among young Indonesian students. The study, involving 20 participants with no prior experience with text-based applications, found that the game effectively motivated and engaged them in reading (Manuaba, 2017). Similarly, Darvenkumar and Devi (2022) explored how text-based games can boost critical reading and thinking skills among 66 tertiary English students. The game made learning more enjoyable and engaging, particularly for less confident learners, with the conclusion that it effectively motivated the participants, enhancing learning out-

comes in language education (Darvenkumar & Devi, 2022).

The importance of motivation cannot be overstated. It is essential—vital in language learning—as those who are inspired are more engaged, often developing a deeper understanding of the material (Zhang, 2023). Text-based games can motivate learners to explore more of the ideas described in the text, fostering comprehension through engagement (Manuaba, 2017), which is an argument presented of video games in general (e.g., MacCallum-Stewart & Parsler, 2007; Reinders & Wattana, 2015; see DaCosta (2024) for a discussion on player motivation to explore and learn more about the in-game items they encounter in video games).

Alongside evidence suggesting that this form of reading can positively influence language learning (Thu & Nhu, 2021), allowing learners to enhance their reading skills, there is also research to suggest that learners can simultaneously acquire new vocabulary in context (Sargsyan, 2013). A notable study by Neville et al. (2009) has shaped later research (e.g., Nagai & York, 2021), examining the effects of IF-based games on vocabulary retention in a German class. Despite its short duration, participants using the game performed better than those using only print materials (Neville et al., 2009). Namely, the findings point to the idea that, though the participants may not have perceived the game as an effective learning tool, it played a role in improving vocabulary acquisition, even for less confident learners (Neville et al., 2009).

As explained, this thinking stems from the belief, at least partially, that language is acquired through exposure and meaningful interactions (Ørnholt, 2023). Ørnholt (2023), for instance, explains that Norwegian students base much of their English language proficiency on classroom instruction, but this also includes exposure to other forms of media (including games) outside of the classroom. This idea that language is learned through practical use relates to the concept of incidental learning, which Furtado et al. (2018), citing Hulstijn (1992), describes as the “accidental learning of information” (p. 7). This type of learning occurs unintentionally, without a conscious goal to acquire knowledge, and takes place while the learner is engaged in a task or activity (Kelly, 2012).

In other words, knowledge (in this case, language, Pereira, 2013) is acquired naturally, almost as a byproduct (Kelly, 2012) of practical use. Given the long-recognized role of reading in facilitating vocabulary acquisition and retention (Marková, 2023), one of the benefits of extensive reading is said to be vocabulary learning (Coady & Huckin, 1997; Sargsyan, 2013). Analogous to incidental learning, this can be described as learning new words while engaged with the text (Kim, 2018). Learners acquire new vocabulary, not formally, but by progressing through the narrative while engaged with the text-based descriptions (Kim, 2018; Marková, 2023).

In addition to reading extensively, it is presented that the depth of the descriptions found in some texts requires learners to read intensively to interpret clues and make strategic choices, enhancing their comprehension skills (Pereira, 2013). This requires learners to read more slowly while systematically analyzing (Kim,

2018) textual descriptions that could be highly detailed and complex (Ørnholt, 2023). In text-based games, this is often presented in the form of scenario-driven puzzles (i.e., tasks) that require problem-solving skills (Wright & Weible, 2024; e.g., Neville et al., 2009). Learners must develop strategies to achieve long-term objectives, which involve recalling specific clues and understanding how objects interact with actions (Ammanabrolu & Riedl, 2019; Dambekodi et al., 2020; Haroush et al., 2018). This likewise involves recognizing implied meanings within the text (Harshitha, 2022), requiring learners to comprehend not only the language but also contextual cues, including character tone.

2.2.2. Issues in Language Learning with Text-Based Games

Notwithstanding these asserted benefits, IF faces challenges in educational contexts, including the notion of incidental learning alongside incidental vocabulary learning. Furtado et al. (2018), for example, states that learners might miss contextual information, skip words, or struggle with comprehension from an incidental learning perspective. It is conceded that there is a counterargument that because text-based games rely heavily on written content, learners are forced to engage with much of the text to progress (Pereira, 2013). However, what could be seen as building upon the reasoning of Furtado et al., Marková (2023) explains that exposure to a target vocabulary item and later recall does not necessarily result in understanding. This is problematic, as misinterpretations of the text during gameplay could hinder progression, leaving the learner frustrated and disengaged.

Then, there are inherent characteristics of text-based games that may impede learning. These games, for instance, have been described as straightforward, which is part of their appeal among educators (Pereira, 2013). They are easier to understand and play, lacking the multifaceted graphical interfaces or challenging control mechanics of modern-day games, which might leave educators feeling out of their depth, unable to incorporate them into classroom settings (Pereira, 2013).

At the same time, their narratives have been described as too complex. Sargsyan (2013), for example, states that a quantitative analysis by Neville et al. (2009) showed that the scaffolding employed was not enough and that participants revealed that the narrative was beyond their proficiency levels. Moreover, as explained, players must work towards goals, which Yuan et al. (2018) expound are often unclear and are discovered only by exploring the game world. In *CCA*, for instance, players need to “exit the building, follow a stream through the forest, and continue downstream until they reach the cave” (Sharma, 2008: p. 30) but are otherwise not told that finding and entering the cave is the first objective of the game—even though it could be argued that the game title implies such.

The minimalistic design of text-based games is also believed to complicate navigation (Lyömiö, 2017). Once the cave is reached, for example, players are faced with a lack of clear direction resulting from the game’s expansive, twisting passages (Lyömiö, 2017). Moreover, repeating steps in the opposite order (e.g., “go north,” “go east,” and then “go west,” “go south”) does not mean that the player will return to the exact location they were previously (Kostka et al., 2017). Unlike

many modern-day games that provide players with on-screen markers (i.e., way-points), in-game maps, and other visual and audio cues to help with such challenges, players of text-based games, in particular early versions, must create maps to assist with navigation and track their progress (Lyömiö, 2017).

Text-based games also feature interactable objects based on their attributes (Côté et al., 2019), and some of these items can be added to the inventory for later use. While this is not an issue in itself, missing crucial items can result in losing the game (i.e., dying), making inventory management essential (Lyömiö, 2017). Likewise, poor choices may lead to unavoidable situations, also requiring players to start over from the beginning (Lyömiö, 2017). Although these mechanics are considered “punishing,” they aim to encourage curiosity and exploration (Lyömiö, 2017: p. 31).

Altogether, the lack of clear objectives, difficult navigation, and the overall punishing nature of text-based games can be frustrating to players, creating a feeling of unfairness and diminishing enjoyment and engagement. Klimmt et al. (2009) explain that when it comes to video games, player enjoyment is directly tied to difficulty level. Tasks (i.e., killing enemies), for example, that are too easy, can induce boredom rather than fun (Klimmt et al., 2009). Consequently, games should support moderate difficulty levels, as these provide the highest level of enjoyment (Klimmt et al., 2009). Extended to learning, text-based games that are too easy may not engage learners, while those that are too challenging might make it hard for them to focus. This could substantially upset the effectiveness of these games in language learning contexts, as imbalanced difficulty levels could discourage learners rather than motivate them.

There are other challenges. Sargsyan (2013), citing Lancy and Hayes (1988), explains that knowing the right words to use (i.e., what the computer will accept) can be difficult. *Colossal Cave Adventure*, for instance, only supports simple commands like “enter,” “drop,” “get,” and “open,” which are common in such games (Côté et al., 2019). Without a complete list of acceptable commands, however, players often need to experiment with different phrases to find those the game will accept (Côté et al., 2019). This is aligned with the game’s overall theme, in which aspects of the game are not explicitly defined and must be learned from exploration (Yuan et al., 2018).

This can be nonetheless problematic, given that text-based games accept free-form action (Shi et al., 2023). Thus, the number of valid actions is theoretically infinite (Yin & May, 2019). Even when the vocabulary and maximum action length are limited (Yin & May, 2019), the number of actions can still be large (Shi et al., 2023), in the hundreds or even thousands (Yin & May, 2019). Jain (2019), for example, explains that *Zork* comprises 85 rooms and 20 treasures, and as a result, the size of the state space is greater than 20^{85} since a treasure could be found in any one of the rooms in the game (with rooms potentially holding multiple treasures). Then there is the complication that not all actions (i.e., commands) are required to solve these games, sometimes only serving as a form of player amuse-

ment (Kostka et al., 2017).

Time is also an issue. Again, citing Lancy and Hayes (1988), Sargsyan (2013) suggests that it could take a few hours to complete a narrative. This means that players should be able to save game state (i.e., checkpoints) to continue later (Sargsyan, 2013 citing Lancy & Hayes, 1988). This is a feature offered in most modern games but is void in early text-based versions.

Perhaps the largest challenge with the use of these games for language learning is their relatively linear plots (Ammanabrolu & Riedl, 2019). Their scripted text is static and, thus, can be predictable (Kirill, 2024), a problem commonly found in early, pure-text versions. As first discussed, dynamic versions of these games do exist (Kostka et al., 2017) that support multiple paths and endings (Zelinka, 2018). However, as also presented, performing the same actions in the same order can result in the same outcomes and, thus, the same on-screen text (Kirill, 2024; Zelinka, 2018). This issue persists in some modern-day games as well (e.g., *The Walking Dead* [Telltale Games and Skybound Games, 2012-2019]), making them feel repetitive (Kirill, 2024), affecting replayability, and weakening long-term player retention (Risi & Togelius, 2020), and thus, can impact their use in learning language contexts. That is to say, although static content can be beneficial to reinforce language learning by providing a consistent learning environment, the deterministic nature of text-based games can lead to reduced engagement and limit exposure to new language use.

Altogether, research into IF, including text-based formats, for language learning, has been affirmed to be in its early stages (e.g., Khatibi & Cowie, 2013), with several significant gaps still needing to be addressed (Furtado et al., 2018; Wright & Weible, 2024). Further compounding matters, and as first presented, IF is said not to be well documented, which could leave educators at a disadvantage, as they must understand how these games function and how to use them effectively (Kim, 2018) if they are to see positive learning outcomes (Wright & Weible, 2024). When coupled with the notion that educators are looking for ways to engage learners to help them build language competency (Rankin et al., 2006; Rankin et al., 2016), alongside the claim that text-based games are a readily available technical resource, the longstanding assertion that there is no reason not to experiment with these games in educational settings (e.g., Pereira, 2013) remains valid today. This is especially relevant when considering the ongoing developments and potential of generative AI.

2.3. Generative Artificial Intelligence

It has been asserted that generative AI shows promise in education (Nash, 2024; Office of Educational Technology, 2023; Wei, 2023). As mentioned, their ability to create dynamic and varied responses (Kirill, 2024) has made them of particular interest when it comes to the study of personalized learning (Jegade, 2024; Office of Educational Technology, 2023). Tailored learning is often mentioned in the literature, focusing on the customization of instructional content to meet the needs

and preferences of learners (Atlas, 2023; Jegede, 2024; Terzic et al., 2023). Early discussions of personalized learning using computers centered on intelligent tutoring, where learners are exposed to step-by-step guidance and personalized feedback based on their performance and needs (Atlas, 2023), showing that interest in creating tailor-made instruction is far from a new idea.

However, Terzic et al. (2023) put forth that in recent years, AI has been pivotal in promoting the creation of interactive tools and respective systems that provide learners with personalized and dynamic educational content (see Jegede, 2024, for a discussion on AI-enhanced English language education tools). Wei (2023) substantiates this, explaining that the integration of AI in education “has garnered considerable attention from researchers, educators, and policymakers worldwide” (p. 3), from their use in intelligent tutoring systems (ITS) to adaptive learning platforms. Citing Chen (2024), Maity and Deroy (2024), for instance, reviewed the benefits and drawbacks of integrating generative AI into ITS to enable the dynamic creation of contextually relevant educational content. Through generative AI, they concluded that such systems could generate questions tailored to the learner’s current situation (e.g., by considering prior responses) to challenge and engage students. Maity and Deroy further state that generative AI can enhance educational systems by providing feedback that goes beyond generic responses, offering human-like interactions that promote deeper understanding.

Wei (2023) offers something similar, explaining that ChatGPT can support language learners by offering targeted feedback on various language skills and sub-skills, ultimately contributing to overall language proficiency; this includes generating grammatically correct results. In a mixed-methods study, Wei investigated the impact of AI-mediated language instruction—delivered via Duolingo—on English learning, L2 motivation, and self-regulated learning among Chinese university students. Sixty participants were divided into an experimental group (AI-assisted) and a control group (traditional instruction). Findings from pre- and post-tests, questionnaires, and interviews revealed that the students exposed to the AI-aided instruction outperformed those in the control group in language achievement. Positive findings were also reported in the study by Jegede (2024), who examined how AI-driven tools can support English language learning through personalized instruction and immediate feedback. An analysis of structured questionnaires from 200 students across four international schools revealed that most participants found AI tools effective in tailoring content and delivering feedback aligned with their learning needs, ultimately contributing to improved language acquisition.

Altogether, this speaks to the nondeterministic nature of LLMs, suggesting that they can be instrumental in language learning, as these models can generate complex, contextually appropriate responses, creating a more natural and immersive experience that enhances engagement (Harshitha, 2022; Sargsyan, 2013). Consequently, these models could be used to create impactful instructional content (Office of Educational Technology, 2023; Terzic et al., 2023). This is particularly the case

when they are integrated into tools. Thus, when viewed from the perspective of text-based games, generative AI could significantly boost their educational value, making them not only more interactive and stimulating (Harshitha, 2022) to learners but also have the potential to provide tailored learning experiences (Kim, 2018).

Artificial Intelligence in Text-Based Games

Text-based games are no stranger to AI. These games, for instance, have been used as testbeds in the areas of “language understanding, problem-solving, and language generation” (Madotto et al., 2020: p. 1). In recent years, there has been a particular interest in building AI that can solve these games (Côté et al., 2019). This is because text-based games can be viewed as problems that can be solved through sequential decision-making, which Yuan et al. (2018) explains can be described by the Reinforcement Learning (RL) setting.

Think of RL as a type of machine learning where an “agent” (i.e., a computer program) learns to make decisions through interactions, much like how players engage with games. In the context of text-based games, the agent learns how to play the game by trying different actions, getting feedback, and improving its strategy over time. Basically, in RL, policies are created that guide agents in selecting the most optimal move at each game state in order to solve a game (Yin & May, 2020). There is considerable research touching upon the use of RL in the context of text-based games (e.g., Ammanabrolu & Riedl, 2019; Atkinson et al., 2019; Côté et al., 2019; Haroush et al., 2018; Kostka et al., 2017; Shi et al., 2023; Towle & Zhou, 2023; Yao et al., 2020; Yao et al., 2021; Yin & May, 2019; Yuan et al., 2018), that include proposing diverse architectures and learning approaches of various RL-based agents (Shi et al., 2023).

While such research is no doubt important, Ammanabrolu et al. (2020) point out that the focus has predominately targeted agents designed to play (i.e., solve) text-based games rather than generating content. This is not to suggest that research focused on generating text (logic and other game components) for these games does not exist (Yao et al., 2020). Yao et al. (2020) reference Ammanabrolu et al., who utilized a Markov and neural language model to generate quests procedurally for games similar to TextWorld. This framework can be used to create and train RL agents for use with text-based games (Côté et al., 2019). More importantly, in the current context, the authors discuss *AI Dungeon 2* (Walton, 2019), which explored the generation of narratives using GPT-2 based on arbitrary text actions. Then there is the work of Zhiwei and Yin (n.d.), who created *Abyssl Hotel*, an interactive text adventure intended to explore interactive narratives through the use of a Stable Diffusion image generation model and ChatGPT-3.5. Their investigation revealed the potential of AI to assist with natural language communication, simulating narrative, and text generation for logical and realistic scenes during gameplay while also speaking to cost savings during game development (Zhiwei & Yin, n.d.).

As described, some recent work has included the use of modern-day LLMs, such as ChatGPT. For example, the platform has been explored in the context of serving as a Dungeon Master in the tabletop game *Dungeons and Dragons* (Gygax

& Arneson, 1974; e.g., Triyason, 2023). Triyason (2023) found that the model showed its ability as a game master at logically developing the narrative, adapting during gameplay, and effectively engaging players, but cautioned that limitations existed when it came to becoming immersed in the narrative, suggesting more research is needed regarding generative AI's ability as a storyteller. Meanwhile, Dong (2023) introduced COTTAGE, a system that uses LLMs to generate coherent text adventures that are logically consistent. The author combined a neuro-symbolic solution with a structured game generation approach to create interactive games that are story-driven and can dynamically adapt to player interactions.

Given the robustness of LLMs, these models could address the challenges previously noted with text-based games. For instance, Pereira (2013) highlights that players often rely on online hints, maps, and walkthroughs to avoid an inability to move forward. Large language models could assist with the narrative by suggesting actions consequential to the objective, hints regarding where to go that assist with navigation, or improved inventory management by offering suggestions or reminders about necessary items and their uses. In other words, LLMs could provide contextual hints and personalized guidance, helping players stay engaged and reduce frustration.

The significance of personalization cannot be emphasized enough. Harshitha (2022) stresses that teachers should account for each individual's language proficiency when using text-based games. Large language models could, therefore, be instrumental in tailored learning contexts by recognizing, for example, that a learner is struggling and recommending additional or different instruction (Office of Educational Technology, 2023). This includes offering scaffolding techniques by gradually increasing task complexity and providing integrated vocabulary and grammar support, allowing players to learn incrementally and effectively. In doing so, LLMs have the potential to generate diverse and contextually relevant dialogues, simulating real-life language use, which, in part, can make these games more engaging and interactive (Kim, 2018; Pereira, 2013), thus speaking to their potential in enhancing the overall learning experience. Moreover, they could make text-based games more adaptive to individual needs by offering challenges and rewards that are unique to the learner and learning experience. Altogether, generative AI has the potential to provide contextual support to help learners progress during gameplay while simultaneously addressing their specific learning needs (Kim, 2018).

3. Method and Materials

The current study presents the development of a generative AI model to explore these possibilities, with all relevant files and materials accessible online. As initially presented, educators and researchers can replicate this approach using the provided methodology to create their generative AI model.

3.1. Instrument

Described as “the first computer game of its kind” (Sharma, 2008: p. 20), *CCA*

holds a significant place in gaming history (Jerz & Thomas, 2015; Sharma, 2008), serving as inspiration for other games (Jerz, 2007; Jerz & Thomas, 2015), such as *Zork* (Jerz & Thomas, 2015). Text-based games often take place in a fantasy world (Kostka et al., 2017) and are typically laid out as a labyrinth (Kostka et al., 2017) that is further organized into “rooms” (Côté et al., 2019; Falcó & Colás, 2022; Jain, 2019; Kostka et al., 2017) which are locations forming a map, such as chambers, halls, a forest (Kostka et al., 2017), and even “STANDING AT THE END OF A ROAD” (wh0am1-dev, 2015). *Colossal Cave Adventure* is no different, set in a dangerous cave inspired by a natural cave system in Kentucky (Jerz, 2007), where players must seek treasure while encountering creatures (Jerz & Thomas, 2015; Sharma, 2008). Moreover, as discussed, these rooms may contain objects players can interact with and add to their inventory (Côté et al., 2019; Kostka et al., 2017), which is accomplished through textual commands (Côté et al., 2019; Lyömiö, 2017; Manuaba, 2017; Yin & May, 2019).

Following Kim’s (2018) suggestion to use existing text-based games, *CCA* was selected as the instrument for this work for several reasons. As a text-based game, it relies on language as the primary interaction method, providing feedback through individual words and sentences (Sharma, 2008; Yin & May, 2019), which Sharma (2008) offers is ideal for language learning. Moreover, its immersive nature demands careful attention to language for effective navigation. For instance, one of the rooms includes part of the description, “TO THE WEST IS A LARGE PASSAGE. ABOVE YOU IS A HOLE TO ANOTHER PASSAGE” (wh0am1-dev, 2015), prompting players to issue precise commands like “go west” or “enter hole.”

It was also chosen for its challenging nature. As with other games in this genre (Kostka et al., 2017; e.g., *Zork*; Yin & May, 2019), the game includes puzzles with embedded linguistic clues, requiring careful reading and interpretation. Challenges range from spatial navigation, where players map and traverse the intricate cave system, to item-based puzzles involving the timely use of objects. Such elements require strategic planning and exploration, arguably demanding advanced language comprehension.

A notable puzzle involves scaring away a snake using a caged bird (Jerz, 2007; Lyömiö, 2017). Players must issue the command “release bird,” which prompts the response, “THE LITTLE BIRD ATTACKS THE GREEN SNAKE, AND IN AN ASTOUNDING FLURRY DRIVES THE SNAKE AWAY” (wh0am1-dev, 2015). To solve this, players need to find both the cage and the bird, noting that the bird is afraid of a rod, which must be temporarily dropped to retrieve the bird. Players then use the bird against the snake. This puzzle requires critical thinking, experimentation, and detailed recall of the game’s environment, which altogether can be argued requires a solid grasp of language.

Then, there is the motivation to use *CCA* due to its ease of development. Sharma (2008) explains that making a game today requires specialized teams comprising numerous talents, from programmers to writers, and can cost upwards of millions of dollars to produce. Text-based games, on the other hand, are compar-

atively simple to create and modify (Manuaba, 2017), with IF-based games still being created today (Kostka et al., 2017). There are a number of tools specifically designed to work with text-based games that have been mentioned in the literature, including Twine (e.g., Wright & Weible, 2024), Inform (e.g., Atkinson et al., 2019; Kim, 2018; Wyatt, 2018), Text Adventure Development System (e.g., TADS; Sharma, 2008), and Quest (e.g., Lyömiö, 2017).

The review by Wright and Weible (2024) found that among these, Twine was the most mentioned, followed by Inform (14 and 7 mentions, respectively). Inform, for example, is widely discussed in academic studies and is popular, in part, because it uses plain English as its programming language to create and modify text-based games (Kim, 2018; Sharma, 2008; Wright & Weible, 2024; Wyatt, 2018). This means that it features a user-friendly, no-code interface with a low learning curve, making it suitable for teachers with limited resources and programming experience (Kim, 2018).

While these tools are generally straightforward, they do require some setup and maintenance. According to Kim (2018), running an Inform game involves configuring an interpreter on target computers for offline play or embedding Inform-compatible code into a website's HTML and uploading game files to the server for online use. Then, there is the issue of which tools can generate dynamic text based on player actions and choices. Inform, for instance, supports this, but the text itself is predefined by the developer within the script, thus making the content dynamic only within the structure of the narrative. Stated differently, Inform does not create new content on the fly, but instead, it shows and hides the predefined content based on player actions or the current state of the game.

Pursuant to Dong's (2023) assertion that generating and running text-based games entirely by AI would be a significant advancement, this investigation explored ChatGPTs (OpenAI, 2024b) (henceforth GPTs), customized instances of OpenAI's (2024a) ChatGPT-4 model to create an advanced version of *CCA*. Generative pre-trained transformers (GPTs), a type of LLM, have shown significant promise in educational contexts (Gimpel et al., 2023). Namely, ChatGPT's capabilities have shown to be impressive in creating responses that are coherent and contextually relevant across numerous domains (Triyason, 2023). These GPTs provide a user-friendly interface for uploading "knowledge source" documents, which give the model contextual information and specific details without requiring complex preprocessing or training (ChatGPT-4, personal communication, July 19, 2024). This makes GPTs a straightforward, no-code solution for creating customized models (OpenAI, 2024b), thus potentially running games like *CCA* without needing specialized interpreters or setups. At the same time, GPTs could enable the creation of personalized experiences while addressing challenges related to developing and deploying games in educational settings.

3.2. Acquiring and Preparing the Game

The first step in creating the model involved acquiring and preparing the game. A

Fortran-based source code package was obtained from a GitHub repository at <https://github.com/wh0am1-dev/adventure> (wh0am1-dev, 2015). This version includes the original *Adventure* game and its source code, which was confirmed by cross-referencing with other sources (e.g., Jerz, 2007). (Jerz [2007] explains that the oldest *Adventure*-related files come from backup tapes of Woods' student account at the Stanford Artificial Intelligence Lab. These include several text files dated March 1977.) Despite Fortran being a legacy language, it is well-documented and was covered in the datasets used to (pre-)train ChatGPT-4 (ChatGPT-4, personal communication, July 19, 2024). With ChatGPT's expertise in interpreting and emulating code, this package was deemed suitable for this research.

Dong (2023) explains that text-based games are (at least traditionally) controlled by the game's programming code. The code defines everything from the rules of the game world to the interactions that the player can have with the environment, including how the rooms are laid out and what items and characters are in them (Dong, 2023), with *CCA* being no exception. Given the significance of the programming code, the goal was to upload the source code and corresponding data files as initially developed without modifications to accommodate the platform. In other words, the intent was to leverage ChatGPT's ability to generate models with minimal setup, avoiding environmental and configuration challenges associated with other solutions.

Moreover, this approach was adopted in response to insights gained from Tsai et al. (2023), who examined how ChatGPT performed with text-based games compared to other solutions. The authors described how players would relay game state to ChatGPT by using the natural language text descriptions as prompts (Tsai et al., 2023). ChatGPT would then suggest actions to take, which the players would carry out within the game (Tsai et al., 2023). Essentially, given ChatGPT's familiarity with *Zork*, the authors examined its ability to provide appropriate responses (i.e., actions) based on the in-game text descriptions (Tsai et al., 2023).

Tsai et al. (2023), however, explained that ChatGPT struggled with remembering locations and objects in the game and that it hallucinated scenes when asked to recall previous playthroughs. Dong (2023) validates this by explaining that LLMs suffer from a multitude of issues, including hallucinations, which could lead to long-term inconsistency during gameplay. Unlike Tsai et al., who seemed to use ChatGPT as a guide for gameplay rather than interacting with the programming code itself, this study directly uploaded the source code to the GPT. This approach provided the model with complete game logic and text-based descriptions, enabling it to both emulate the game and assist with recall during gameplay.

The package was also beneficial due to its supplementary content. It included distribution files for macOS and Windows, enabling direct comparison between the GPT-generated output and the original game's executable versions. Furthermore, it featured detailed maps of the game's environment, aiding navigation and understanding of the cave's complexity. Walkthroughs with instructions were also

crucial for guiding the testing and evaluating the GPT's performance.

Finally, the package was selected because it is available as open source. The author clearly defines the conditions of use, stating, "This repository is for educational purposes only" (wh0am1-dev, 2015, "Readme" section), which permits its use in the current study.

3.3. Training, Refining, and Evaluating the Baseline Game

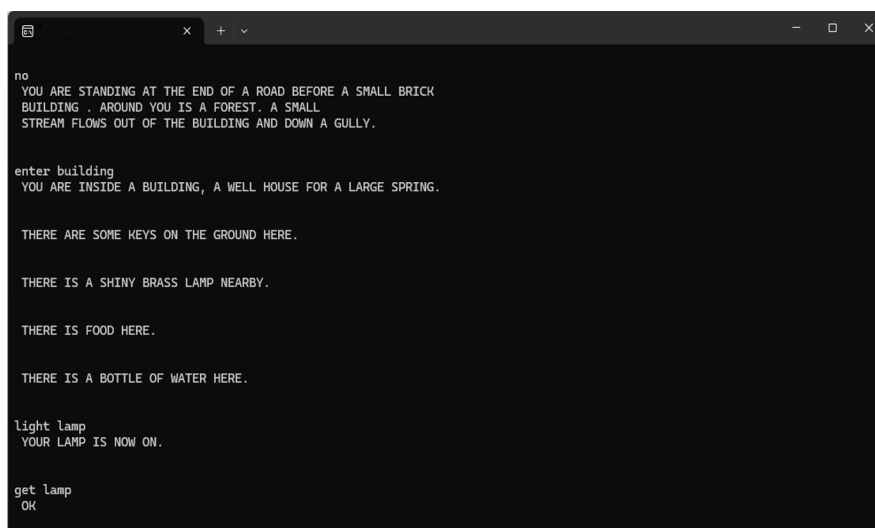
The GPT was next created using the GPT Builder to emulate the game "as-is" based on its source code; that is, ChatGPT should be able to simulate the game so that it can be played as it was originally designed. The process began by uploading the code (logic), 77-03-31_adventure.f, and data (in-game text descriptions), 77-03-31_adventure.dat, files found in the package at \adventure-master\src\, to provide the model with the necessary content. The file map.pdf, found at \adventure-master\maps\, was also uploaded to provide the model with navigational data. The Code Interpreter & Data Analysis feature was then activated to help the model understand the game's logic and text-based descriptions (i.e., interpret the Fortran source code). Instructions were then entered, resulting in a GPT that could be examined to determine if it effectively mirrored the original game.

Using Walkthrough 1, found at \adventure-master\walkthroughs\, identical commands were entered into the GPT and the Windows advf4-31.exe runtime with responses compared meticulously. The runtime—an executable file of the game available at \adventure-master\dist\win\Adventure\—served as a control to compare the GPT's output. Discrepancies led to iterative adjustments in the GPT's instructions, ensuring alignment with the runtime output. This approach was adopted based on the work of Kirill (2024), who did something similar in the training of a Claude model; as the author notes, the LLM tended to stray from the plot when creating its narrative. By working through iterations, adjusting the scene descriptions, and refining the instructions, Kirill was able to ensure that the model consistently provided players with options that resulted in the intended direction. This process was essential to the current work, affording a means to modify the GPT and verifying that it accurately emulated the original gameplay, thereby establishing a baseline model by which to add enhancements.

It is important to explain why Walkthrough 1 was used in this investigation, given that two additional walkthroughs are included in the package. These walkthroughs vary in their detail as they consider the steps needed to achieve different scores. Although scoring details are not displayed during gameplay in *CCA*, leading to misconceptions that the game lacks a definitive highest score or a fewest-move solution (Sharma, 2008)—as not all text-based games rely on scoring, often using only win/loss conditions instead (Kostka et al., 2017)—it was initially designed with a target score of 350 points, achievable by collecting all 15 hidden treasures (Lyömiö, 2017). Later versions allowed higher scores, which is reflected in the walkthroughs, with 1 offering detailed guidance for a playthrough described as being close to optimized (to achieve the maximum score) and 2 and 3 outlining

the steps for scoring 350 and 550 points, respectively. Walkthrough 1 was not chosen based on scoring but instead offered a detailed playthrough (despite the use of magic words or cheat codes, e.g., “plugh,” “plover,” and “xyzzz”), making it the best choice among the three walkthroughs.

Moreover, the package includes two other versions of the game, 77-03-11_adventure and 77-03-23_adventure, warranting an explanation as to why the 77-03-31_adventure release was chosen. According to Jerz (2007), the March 11 version, created by Crowther, is the first known iteration. The March 31 version, which is similar but features renumbered vocabulary items, is considered the final version before Woods’ enhancements (Jerz, 2007) and is one of the reasons for its selection in this research. The March 31 version was also chosen because of its comprehensiveness, encompassing code, data, and runtime files, in comparison to the other two versions found in the package. **Figure 1** shows this runtime, depicting the original game’s interface and gameplay, while a walkthrough of the game’s source code and data can be found in Jerz (2007).

A screenshot of a terminal window showing the text-based interface of the game. The text is as follows:

```
no
YOU ARE STANDING AT THE END OF A ROAD BEFORE A SMALL BRICK
BUILDING . AROUND YOU IS A FOREST. A SMALL
STREAM FLOWS OUT OF THE BUILDING AND DOWN A GULLY.

enter building
YOU ARE INSIDE A BUILDING, A WELL HOUSE FOR A LARGE SPRING.

THERE ARE SOME KEYS ON THE GROUND HERE.

THERE IS A SHINY BRASS LAMP NEARBY.

THERE IS FOOD HERE.

THERE IS A BOTTLE OF WATER HERE.

Light Lamp
YOUR LAMP IS NOW ON.

get Lamp
OK
```

Figure 1. The original game interface based on the advf4-31.exe runtime (Filepath removed from the figure).

3.4. Training, Refining, and Evaluating the Advanced Game

Upon confirming that the GPT successfully emulated the original game, several enhancements were made. This was vital, as building upon a baseline of the original gameplay offered some assurance that the behavioral changes observed during gameplay were a result of the new instruction. These new enhancements entailed instructing the GPT to provide contextual hints and guidance to reduce frustration and keep learners engaged. The GPT was also asked to offer advice on essential inventory items to help learners manage resources effectively and, in doing so, lessen the chances of having to start over (i.e., losing the game). Furthermore, using the text descriptions found in the data file, the GPT was instructed to generate unique, contextually relevant language with the intent of enriching the

overall experience. Finally, the GPT was instructed to increase the complexity of the language as learners progressed through the game, aiming to challenge them and enhance their learning experience.

With an advanced version of the GPT created, Walkthrough 1 was used to evaluate the modified model, like what was done with the baseline version. However, the evaluation no longer focused explicitly on response comparisons but rather on assessing the new functionalities. In other words, unlike the initial phase, which ensured the GPT mirrored the original game, this phase aimed to verify if the model met the new objectives. The evaluation tested the GPT with scenarios differing from and aligning with the walkthrough's commands. Based on these assessments, the GPT's instructions were refined to ensure it provided contextual hints, inventory advice, and rich and progressively complex language.

For example, unsuccessful commands were entered to assess if the GPT provided appropriate contextual hints and guidance, helping to redirect the learner onto the correct path. To evaluate inventory recommendations, items that were considered vital (i.e., needed later to advance the gameplay) were intentionally left behind to see if the GPT would suggest collecting them. The text generated was compared to the runtime output to ensure it was contextually appropriate yet distinct from the original in-game descriptions. Finally, language complexity was evaluated by observing how the model's generated text evolved in vocabulary richness and syntactic complexity over time in comparison to the original in-game text.

The GPT interface is shown in **Figure 2**, and the final instructions and configurations (used in this study) are detailed in **Table 1**. Although presented in two paragraphs for clarity, they represent a unified set of instructions. The first paragraph covers the baseline game instructions. In contrast, the second outlines enhancements for the advanced version. Educators can begin by using the initial paragraph to have their GPT emulate the original game and then apply the second paragraph to incorporate the advanced features or customize the model to suit their specific educational needs.

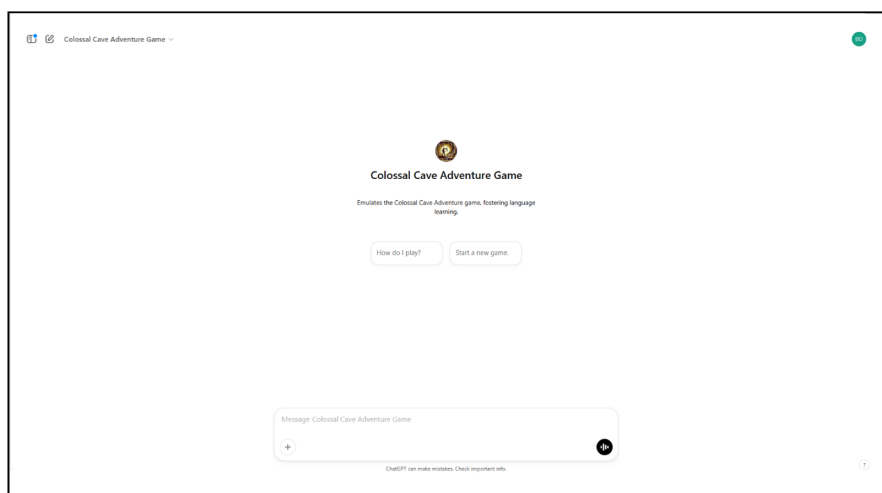


Figure 2. The GPT game interface (“By” line removed from the figure).

Table 1. Instructions and settings used to create and configure the GPT.

Name	Value
Instructions	<p>This GPT emulates the original Colossal Cave Adventure game using the provided source code (77-03-31_adventure.f and 77-03-31_adventure.dat). It interprets the code to respond to user prompts and guide the gameplay. Each session begins with providing instructions on how to play the game. The GPT responds to user prompts by executing the corresponding actions within the game, describing the outcomes. The GPT will strictly follow the game logic, serving as the game emulator. The GPT will interact politely.</p> <p>Additionally, suppose users seem uncertain about their next steps or direction during gameplay. In that case, the GPT will provide contextual hints and tailored guidance to help prevent confusion and ensure users remain engaged and free from frustration. The GPT will use the map (map.pdf) as a reference to identify the user's location accurately and offer appropriate suggestions. Furthermore, the GPT will offer advice and reminders about essential items and their uses that need to be added to the inventory, reducing the risk of players losing the game. The GPT will recommend that users gather certain items before leaving an area. Also, using the data file (77-03-31_adventure.dat), the GPT will generate unique responses that maintain the original meaning but are varied and contextually relevant to mimic real-life language usage. Likewise, as the user advances in the game, the GPT will progressively enhance the complexity of the responses to challenge and develop the user's vocabulary and reading comprehension skills. Finally, the GPT will only accept grammatically correct prompts. For example, the GPT will not accept "enter building" but will accept "Enter the building." If a prompt is not grammatically correct, the GPT will reject it and provide an example of the correct format.</p>
Conversation Starters	<p>How do I play? Start a new game.</p>
Capabilities	Code Interpreter & Data Analysis
Knowledge Sources	<p>77-03-31_adventure.f 77-03-31_adventure.dat map.pdf</p>

3.5. Procedure

Colossal Cave Adventure was played three times over one week by the author using the advanced GPT. Insights from the training and refinement phases (baseline and advanced) informed a modified version of Walkthrough 1, which guided each playthrough for consistency. This modified version avoided the use of magic words heavily used in the walkthrough. It also intentionally explored spaces that were otherwise avoided because of the cheat codes. Moreover, the modifications assessed the GPT's ability to offer contextual hints and recommend inventory

items.

The modified walkthrough served as a worksheet to document the GPT's responses and any deviations from expectations, that is, behaviors expected and observed during the training and refinement of the advanced phase. Each playthrough was recorded in a worksheet, capturing observed responses and behaviors. The three completed worksheets were then analyzed for reporting purposes.

4. Results and Discussion

As noted by [Terzic et al. \(2023\)](#), like other novel tools, there are persistent challenges concerning the use of generative AI in achieving enhanced learning outcomes that improve educational processes. This work was no exception, with the playthroughs showing that the GPT performed well in many areas, with challenges also uncovered. The following analysis summarizes the results, presenting the strengths and areas where improvements and further exploration are needed. These findings are presented in a structured discussion, with the results systematically organized into categories that emerged as patterns, themes, and relationships within the data during the analysis.

4.1. Emulating the Gameplay

As discussed, the GPT successfully emulated the original game, showcasing ChatGPT's strong code interpretation skills. Although some refinement was necessary to align responses with expected behavior, plain English was used to "train" the GPT, taking the form of human-readable instructions. At the same time, knowledge documents comprising the code and supporting files were directly uploaded. The process was simplified via the ChatGPT online interface, which offered a streamlined approach that eliminated complex setups and demonstrated the system's accessibility and adaptability.

The most significant challenge found outside of the need to explore prompt engineering techniques was that a subscription was required. Granted, at the time of this writing, users can freely create an OpenAI account with limited access to its models (e.g., GPT-4o mini and o3-mini) ([OpenAI, 2025c](#)). However, full access to advanced features, like GPTs, which were central to the current work, required a paid subscription, with ChatGPT Plus currently priced at USD 20 per month ([OpenAI, 2025c](#)).

While this may be affordable for working professionals, [Wang and Guo \(2023\)](#) caution that it could pose a financial barrier in educational settings, such as schools and research institutions, thereby limiting accessibility. Consequently, they propose that future research should focus on the reduction of costs associated with the use of AI-powered platforms as a means of making such technologies readily accessible ([Wang & Guo, 2023](#)). This sentiment is echoed in this investigation and advocates that companies, such as OpenAI, consider offering free versions of their advanced features to reach broader audiences or, at minimum, make such capabilities freely available for educational use or at reduced costs.

4.2. Contextual Feedback

The playthroughs revealed that the GPT provided contextual hints and guidance during gameplay. For instance, after spending time in the brick building (at the start of the game), the model began to offer suggested actions like “explore the area around the building” and “follow the stream down the gully,” with the latter offering guidance to the cave’s entrance. Curiously, the GPT also began outlining pathways, allowing a recommended path to be chosen by responding with its corresponding number. This behavior was not consistent but occurred at different times during the gameplay throughout all three playthroughs.

Côté et al. (2019), Jain (2019), and Zelinka (2018) categorize text-based games into parser-, choice-, or hypertext-based formats depending on how commands are input, with a similar classification provided by Lyömiö (2017). Zelinka explains that “all parser-based games with [a] finite number of actions can be converted to choice-based games by simply enumerating all the actions accepted by the interpreter at different time steps” (p. 6). With *CCA* being a classic parser-based game (Côté et al., 2019; Lyömiö, 2017), the GPT was able to transform the gameplay into a choice-based format, allowing options to be selected rather than directly inputting commands. What this means is that noting the desired path number behaved similarly to issuing a natural language command, prompting the model to execute the action and update the game’s internal state—and, thus, the described scene—accordingly.

Colossal Cave Adventure’s conversion into a choice-based game was unexpected and showed the power of GPTs. However, this could prove to be challenging in language learning contexts. As presented, the thinking is that learners enter natural language commands, enhancing skills through active reading, comprehension, and linguistic formulation (Pereira, 2013).

With the GPT’s choice-based format, learners instead receive pre-defined options, which may limit opportunities to practice these language skills. In other words, although learners would still read and interpret text, they generate fewer commands, which could reduce the game’s effectiveness for language learning. Consequently, the GPT’s choice-based interface, while adaptable and user-friendly, may limit active language production and thus compromise the depth of language practice and its efficacy in educational settings.

Additionally, it was observed that hints could be obtained by requesting assistance from the model during gameplay. For example, asking the model, “What should I do next?” elicited the response, “Here are a few suggestions for what you can do next,” with the itemized paths numbered and described.

The GPT also sometimes offered guidance on which items to collect during gameplay. Upon entering the building in one of the playthroughs, the model immediately announced:

“Inside the building, there is a shiny brass lantern, a set of keys, a tasty-looking bottle of water, and a loaf of bread.”

Whereas the runtime of the original game provided the text, “THERE ARE SOME KEYS ON THE GROUND HERE. THERE IS A SHINY BRASS LAMP NEARBY. THERE IS FOOD HERE. THERE IS A BOTTLE OF WATER HERE” (wh0am1-dev, 2015).

The GPT embellished the description in comparison to the original in-game text, but what is of interest is that the GPT suggested items be picked up. For instance, upon entering “Get the lamp,” the model responded:

“The items in the well house [building] can be very useful in your adventure. Currently, you have picked up the brass lantern. There might be other useful items around.”

This guidance was not only encouraging but helpful, resulting in the additional items being added to the inventory in the playthroughs. Gaining such items also prompted the GPT to sometimes respond with additional hints and advice on their importance and use. For example, picking up the bottle and bread returned:

“You pick up the bottle of water.”

“You pick up the loaf of bread.”

However, the command “Get the keys” was met with:

“You pick up the keys. They are essential for opening doors and unlocking certain areas within the cave, making them critical for game progression.”

4.3. Relevant and Complex Text

The GPT generated contextually appropriate text across the playthroughs that aligned with the original game, showing its adaptability and nondeterministic nature. For instance, the original game displayed the text, “YOU ARE STANDING AT THE END OF A ROAD BEFORE A SMALL BRICK BUILDING. AROUND YOU IS A FOREST. A SMALL STREAM FLOWS OUT OF THE BUILDING AND DOWN A GULLY” (wh0am1-dev, 2015). Conversely, the GPT described the room as:

“You are standing at the end of a road in front of a small brick building. The surrounding area is dense with forest, offering a sense of mystery and adventure. A gentle stream emerges from the building, trickling down into a nearby gully.”

Later in the gameplay (in one of the playthroughs), the original game presented, “YOU ARE ON THE EAST BANK OF A FISSURE SLICING CLEAR ACROSS THE HALL. THE MIST IS QUITE THICK HERE, AND THE FISSURE IS TOO WIDE TO JUMP” (wh0am1-dev, 2015). Which was illustrated by the GPT as:

“You find yourself standing on the eastern edge of a deep chasm that bisects the hall from end to end. The mist here is dense, obscuring your vision and rendering the air heavy. The chasm is alarmingly wide, far too broad for you to leap across.”

Other examples are shown in **Table 2**, highlighting the GPT’s capacity to add depth and complexity to the narrative, intended in this study to offer learners a more engaging and linguistically rich experience. They demonstrated how the model introduced vivid descriptions with the hope of creating a more immersive atmosphere compared to the original game’s more direct text.

Table 2. Examples of text descriptions between the original CCA game (wh0am1-dev, 2015) and the advanced GPT version.

Original Game	Advanced GPT
<p>“YOU ARE AT ONE END OF A VAST HALL STRETCHING FORWARD OUT OF SIGHT TO THE WEST. THERE ARE OPENINGS TO EITHER SIDE. NEARBY, A WIDE STONE STAIRCASE LEADS DOWNWARD. THE HALL IS FILLED WITH WISPS OF WHITE MIST SWAYING TO AND FRO ALMOST AS IF ALIVE. A COLD WIND BLOWS UP THE STAIRCASE. THERE IS A PASSAGE AT THE TOP OF A DOME BEHIND YOU.”</p>	<p>“You stand at one end of an immense hall that extends out of sight to the west. Openings are visible on either side. Nearby, a broad stone staircase descends into the depths. The hall is filled with tendrils of white mist that sway back and forth, almost as if they were alive. A cold wind gusts up the staircase. There is a passage behind you, concealed at the apex of a towering dome, its shadow stretching across the hall, silent and waiting.”</p>
<p>“YOU ARE IN A MAZE OF TWISTY LITTLE PASSAGES, ALL ALIKE.”</p>	<p>“You find yourself entangled in a labyrinth of serpentine corridors, each indistinguishably winding and seemingly endless.”</p>
<p>“YOU ARE IN A LARGE LOW CIRCULAR CHAMBER WHOSE FLOOR IS AN IMMENSE SLAB FALLEN FROM THE CEILING (SLAB ROOM). EAST AND WEST THERE ONCE WERE LARGE PASSAGES, BUT THEY ARE NOW FILLED WITH BOULDERS. LOW SMALL PASSAGES GO NORTH AND SOUTH, AND THE SOUTH ONE QUICKLY BENDS WEST AROUND THE BOULDERS.”</p>	<p>“You find yourself in a vast, low-ceilinged circular chamber, the floor of which is dominated by an enormous slab that has fallen from above (Slab Room). Once, there were large passages to the east and west, but massive boulders now obstruct them. Narrow, low passages lead to the north and south, with the southern one curving westward around the boulders shortly after its entrance.”</p>

It is important to note that evaluating whether the GPT’s text consistently increased in complexity as gameplay progressed was problematic. While the model seemingly enriched the narrative with greater depth and detail, the expected gradual increase in complexity was not always apparent, both within single sessions and across multiple playthroughs.

In individual sessions (i.e., during the same playthrough), the GPT added vivid descriptions and intricate details, but the complexity did not always follow a clear, linear path. Occasionally, the in-game text descriptions would become richer, then revert to more direct language, disrupting the expected progression and making

it difficult to observe a steady increase in complexity during a single playthrough. Equally, analyzing text across the three playthroughs confirmed this behavior. The expected growth in narrative complexity varied, with some sessions showing a consistent increase in detail while others did not. This inconsistency made it difficult to assess whether the GPT's output steadily enhanced complexity throughout the gaming experience.

What this suggests is that generative AI's uneven narrative advancements could complicate creating adaptable learning experiences. In educational settings, where personalized learning is crucial, such inconsistencies can hinder effectiveness. If the AI's output does not consistently match the needed complexity, it may not provide a uniformly challenging experience. Some learners might face overly simplistic content, while others might encounter too advanced text. This can disrupt a structured learning path, making it hard to ensure learners progress at an appropriate pace.

4.4. Recognizing Text Commands

This raises another interesting observation that was experienced during gameplay that speaks to the nature of LLMs and their ability to process natural language. Côté et al. (2019) explain that one of the most significant challenges in playing parser-text-based games is determining which commands the parser will recognize. As described, the original *CCA* is limited to simple commands like "enter," "drop," "get," and "open." Which, as explained, denote the types of terms that most text-based games understand (Côté et al., 2019). However, as also discussed, unless the game presents a finite list of all allowable commands, players must often experiment with different phrases and wording to find the precise commands that the game accepts (Côté et al., 2019).

This can potentially make the gameplay experience frustrating and less intuitive, impacting learning. For example, in the original game, commands such as "enter," "enter building," and even 'e' work, but "enter the building" does not. In contrast, the GPT accepted "Enter the building" along with any other similar phrases, such as "Go into the building" and "Step into the building." While this was promising, the model was very relaxed in what it would accept, including misspellings, alongside the original game's cryptic commands (e.g., "e," which can represent enter or east depending on the room and how it is used).

Even though this was by design, given the intent of using such games in language learning contexts, additional instructions were added during the first playthrough to ensure the GPT only accepted grammatically correct commands. This was reflected in the model's behavior in subsequent playthroughs. In the second playthrough, for example, the GPT offered:

"(Note: Remember to use grammatically correct prompts. For example, 'Enter the building.' instead of 'enter building.')."

Whereas, in the third playthrough, the command "get bottle" resulted in the

GPT not only pointing out that the phrase was not grammatically correct but also suggesting a contextually appropriate one:

“Your command is not grammatically correct. Please use a complete sentence. For example, “Take the bottle.””

4.5. Interactive Game Elements

As presented, the model altered the narrative in each of the playthroughs, embellishing each room. However, what is more intriguing is that the GPT introduced new elements and logic to the gameplay. Although the essential mechanics remained intact, the model modified existing game elements while altering the logic for how these elements interacted, resulting in a different experience each time the game was played. This demonstrates its ability to provide new and engaging experiences, promoting replayability while maintaining learner interest.

For instance, the GPT enhanced the game by changing certain elements. It was mentioned earlier that the model (in one of the playthroughs) described the building as containing “a shiny brass lantern” and “a loaf of bread.” Neither of these exist in the original game. Instead, the “LAMP” (in the original game; [wh0am1-dev, 2015](#)) became a “lantern,” and the “FOOD” ([wh0am1-dev, 2015](#)) was defined as a “loaf of bread.” These changes show how the model added new details, enriching the gameplay with additional context and description.

The GPT also added entirely new game elements. In another playthrough, the model embellished the building with “a desk with some papers and a small brass lamp,” “a staircase leading down into darkness,” and “a rug on the floor with a curious bulge underneath.” Moreover, the GPT allowed interaction with these new game elements. The staircase led to (what appeared to be) a cellar, revealing the bottle of water. Meanwhile, pulling back the rug, exposed the keys.

Finally, the GPT introduced new logic. For example, when attempting to turn on the lamp after picking it up, the GPT responded in one of the playthroughs with, “You cannot turn on the lamp because you already picked it up.” This departs from the original game, where the lamp can be turned on even when it is part of the inventory. It was found that by dropping the lamp, it could then be lighted and consequently placed back into the list of carried items.

This new logic demonstrated the GPT’s ability to create unique gameplay experiences but, more importantly, highlights the need to ensure that such changes align with the original game’s mechanics. This is a matter stressed by [Kirill \(2024\)](#), who emphasized the complexity of ensuring that the story remains consistent despite slight variations in in-game text descriptions while at the same time guaranteeing that essential information is presented. This can be problematic as inconsistent modifications—such as altering command functionalities or adding new game elements—can confuse learners and disrupt the learning process. Suppose that using a GPT like that used in this research alters how commands are interpreted or introduces elements that deviate from the original game’s logic. In that case, it could create inconsistencies in how students learn and apply language

skills.

4.6. Response Unpredictability

The nondeterministic nature of LLMs offers both benefits and challenges in education. Unique and dynamic narratives can enhance engagement and provide varied learning experiences. However, this variability can be problematic where consistency is crucial. Each game playthrough produced variations in the story with new elements (i.e., no two gaming experiences would be entirely the same, even if following identical commands), leading to inconsistent experiences. Tsai et al. (2023) explained this in their investigation, and it is one of the reasons why the GPT was created in the current work with the *CCA* source code and accompanying files uploaded as knowledge sources. Even so, given the nondeterministic nature of ChatGPT, this approach resulted in the same outcome. While this unpredictability was by design, it can also complicate the creation of uniform content for assessments, raising concerns about fairness and equality in AI-driven education (Office of Educational Technology, 2023). Such diverse outputs could undermine efforts to maintain consistent evaluation conditions.

Unsurprisingly, there are concerns about using generative AI in education, particularly among educators (Office of Educational Technology, 2023). For instance, ChatGPT's responses have been argued to appear credible, but OpenAI (2023a) acknowledges that they may not always be factually correct. This is because ChatGPT (and other generative AI platforms) can hallucinate, producing fabricated and misleading content (Gimpel et al., 2023). In the current investigation, some degree of fabrication was expected and even sought, as the GPT was tasked with creating new content that—if it remained contextually relevant to the story and original text—was beneficial for its purposed language learning context. However, this issue can be problematic when such a platform is relied upon as an authentic resource for information. This risk of inaccuracies highlights the need for careful oversight to ensure that the information meets educational standards.

This variability also raises concerns about bias. While ChatGPT's adaptability can be beneficial in some contexts, it is essential to recognize that its behavior and responses were significantly influenced by the training (i.e., uploaded knowledge sources) the model received in this study and the extensive datasets used to train the foundation model, ChatGPT-4. Discrimination in AI is a well-documented issue, especially in educational contexts, as educators seek technology-enhanced methods that are not only practical and scalable but safe (Office of Educational Technology, 2023). OpenAI (2023b) acknowledges that ChatGPT may exhibit biases because of the datasets used, which may reflect Western perspectives and individuals, thus potentially showing societal biases and exposing discriminatory views in its responses. As with its tendency to hallucinate, this, too, emphasizes the need for careful evaluation of AI systems to address partialities and ensure fairness.

The GPT used in this research did not exhibit bias or use language that might be considered harmful or hateful towards a person. An early version of the *CCA* source code and data was used, and a review of this content did not reveal concerning language, except perhaps for references to “A LITTLE DWARF” (wh0am1-dev, 2015), which may be viewed as offensive. However, given ChatGPT’s nondeterministic nature, the interpretation and use of such text in subsequent playthroughs remain uncertain, especially in educational settings for language learning. In other words, Harshitha (2022) explains that teachers have no control over the responses these models generate, making it imperative that they work closely with their students when using such models. At the same time, The Office of Educational Technology (2023) points out that educators are fully aware of the risks of using such new technologies.

It is also important to note that the model occasionally lost track of the game’s state and instructions. It sometimes forgot key details, such as inventory items and narrative progress, which made the gameplay difficult. This is a finding reported by Dong (2023), who explains that text-based games, which rely on LLMs, can be contradictory because of their inability to consistently track game state, thus knowing when it should accept or reject player requests. Dong, for example, discussed this in the context of *AI Dungeon* (Walton, 2019) (Yao et al., 2020, discussed *AI Dungeon 2*), describing it as a GPT-based online text adventure game that had exhibited similar challenges, demonstrating inconsistencies and forgetfulness during gameplay. The GPT, in the current work, often required reminders to follow instructions or recall previous events. Although prompting the model would (sometimes) correct the game state, this need for frequent reminders could disrupt the learning process and negatively impact the overall experience in educational contexts.

Finally, ChatGPT-4’s token limit for interactions was a significant challenge. The message, “You’ve hit the Plus plan limit for GPT-4” (OpenAI, 2024a), appeared frequently during playthroughs, requiring waits before gameplay could resume. This limit, along with the model’s verbosity, quickly consumed tokens and interrupted gameplay. While the model could be instructed to minimize its verbosity in responses, thus helping mitigate token limitations, this would undoubtedly be counterproductive in language learning contexts, where extended and detailed interactions are crucial.

5. Limitations and Future Research

As with all research, this work has limitations. It used an early version of *CCA* (i.e., *Adventure*) written in Fortran, a language well-interpreted by ChatGPT. It is unclear if similar results would occur with other text-based games developed in different languages or platforms. Zelinka (2018) explains that the lack of a universal interface for different types of text-based games makes it problematic to learn how to play multiple games or, in this context, conduct experiments, making generalization difficult. Future research should, therefore, investigate a broader range

of text-based games across various development environments to understand how generative AI interacts with different systems and assess its effectiveness and versatility in diverse contexts.

Moreover, this study focused on ChatGPT, specifically GPTs, for their ease of use without the need for complex preprocessing (outside of uploading knowledge documents and providing instruction). However, other LLMs might provide different insights and behaviors. Replicating this study with other generative AI tools and platforms is essential to understanding how different models impact gameplay and educational value. This would offer a broader view of generative AI's applicability in education while providing a means to evaluate the effectiveness of various LLMs.

The walkthroughs used were crucial to ensure consistency but may have introduced bias by presenting a specific narrative perspective. It is unclear if different scripted solutions would yield different results. Future research should explore both scripted and free-play methods to see if learner behavior and AI responses differ significantly under various conditions, providing a broader understanding of how generative AI interacts with different gameplay styles and narratives.

The use of specific knowledge source documents, such as the map.pdf file, also raises questions. While ChatGPT-4 (personal communication, July 19, 2024) claimed it could read the map image file, its effectiveness during gameplay is unclear. That is, the removal of the file (while developing the advanced version of the model) did not appear to impact or show behavioral changes, perhaps suggesting that the model used the code to figure out navigation rather than the map or that the removal of the file forced the model to rely on the game logic and data files. Tsai et al. (2023) also explained that in their investigation, ChatGPT struggled and that the model did not learn the world of *Zork*, instead relying on memorization. They concluded that ChatGPT could not construct a map during gameplay to consistently track the player's current location or how to reach a given destination (Tsai et al., 2023). Future research should examine how these files influence the model's responses and interactions to understand their role and effectiveness in text-based games.

The GPT's tendency to forget instructions and lose track of the narrative is concerning. Future research should seek solutions for managing game state to enhance usability and coherence. Interestingly, in one of the playthroughs, the GPT asked about saving the game. Moreover, requesting the model to return to the last checkpoint successfully restored the game state. However, this functionality was limited, with the GPT merely retaining information for the duration of the current session. To enhance the use of generative AI in gaming, especially for educational purposes, it is crucial to develop robust mechanisms for saving and restoring game states across multiple sessions in a single playthrough. Until this is possible, shorter, more focused play sessions may be advisable, requiring educators to develop strategies to achieve learning objectives within these con-

straints.

Closely related to this issue was the model's permissiveness in accepting improper input, including misspellings and cryptic commands. This issue was addressed through additional instruction, after which the GPT was configured to accept only grammatically correct prompts. However, this also suffered, given that the model occasionally deviated from its instructions, forgetting constraints over time. Therefore, future research should explore effective strategies for handling improper inputs from learners and assess their impact on language learning.

Then, there is the matter of the constant and rapid evolution of generative AI. Harshitha (2022) explains that a category of IF that benefits from AI technology is voice-based IF. This form uses speech recognition and natural language understanding via smart devices to provide an audio narration of the story, with the player entering prompts by voice (Harshitha, 2022) instead of text. During this investigation, OpenAI released a similar feature, allowing users to interact with its models entirely by voice commands, with the model replying through audio personas (OpenAI, 2025b; see the audio icon depicted in Figure 2). Future research should explore this audio feature in the context of using text-based games for language learning and making such solutions accessible to visually impaired learners.

It is also important to disclose that, during this writing, OpenAI (2025d) released its 4.5 model. As a result, educators and researchers replicating this work—even if using the same knowledge sources and applying the same instruction—may observe different behavior from their GPT. This is particularly true if they leverage newer and, presumably, more enhanced versions of the ChatGPT base model. Ultimately, this offers direct evidence to the statement that generative models are constantly evolving, and underscores that this work is not definitive but instead a resource and catalyst for further study.

Perhaps the most limiting is that this research primarily presents insights from the author regarding the behaviors of the GPT. In other words, the perspectives shared reflect a single individual's observations. Future research should focus on examining the use of AI-powered text-based games with actual language learners. For example, studies could compare language learning outcomes between participants exposed to such games and those who receive traditional language instruction. Additionally, future studies might assess the impact of AI-powered games versus traditional text-based games by comparing, for instance, participants who engage with the enhanced GPT *CCA* version and those who play the original.

This includes addressing another limitation identified in this study: the difficulty in assessing the progression of language complexity during gameplay. Rather than using formal, standardized metrics, the evaluation relied on observations—monitoring how the GPT-generated text evolved in terms of vocabulary richness and syntactic complexity. However, in the absence of quantitative tools, the anal-

ysis lacked precision. Future research should incorporate standardized instruments (e.g., Flesch-Kincaid readability scores) to more systematically track and evaluate language complexity throughout gameplay, enabling a more rigorous and data-driven assessment of AI-supported language learning.

Finally, although this work focused on language learning through *CCA*, the findings suggest broader applications for generative AI. For example, text-based games can explore historical events from various perspectives (Harshitha, 2022; Wright & Weible, 2024). Interactive fiction can integrate cultural and language learning (Wright & Weible, 2024), teaching cultural contexts along with language skills such as grammar (Europass SRL, 2023). Future research should examine incorporating diverse cultural and historical narratives into text-based games with generative AI. This could enhance cultural awareness and historical understanding and potentially foster eco-consciousness through environmental themes.

6. Conclusion

In conclusion, this work explored how generative AI, specifically ChatGPT, can enhance text-based games for language learning. By creating a GPT of the version 4 foundation model to generate complex and contextually relevant responses, the study aimed to create a more immersive and engaging experience. The custom model preserved the original game's structure while introducing dynamic content and continuous feedback. However, this adaptability also brought challenges, such as maintaining consistency. While there is more work to be done, it is hoped that the findings found herein pave the way for future research to develop innovative strategies and methods, potentially extending the benefits of generative AI beyond language learning to other educational contexts.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 Writing Practice: Game Enjoyment as a Key to Engagement. *Language Learning & Technology, 18*, 124-150.
- Allington, D. (2015). Linguistic Capital and Development Capital in a Network of Cultural Producers: Mutually Valuing Peer Groups in the "Interactive Fiction" Retrogaming Scene. *Cultural Sociology, 10*, 267-286. <https://doi.org/10.1177/1749975515598333>
- Alyaz, Y., & Genc, Z. S. (2016). Digital Game-Based Language Learning in Foreign Language Teacher Education. *Turkish Online Journal of Distance Education, 17*, Article No. 9. <https://doi.org/10.17718/tojde.44375>
- Ammanabrolu, P., & Riedl, M. (2019). Playing Text-Adventure Games with Graph-Based Deep Reinforcement Learning. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Volume 1, pp. 3557-3565). Asso-

- ciation for Computational Linguistics. <https://doi.org/10.18653/v1/n19-1358>
- Ammanabrolu, P., Broniec, W., Mueller, A., Paul, J., & Riedl, M. O. (2020). Toward Automated Quest Generation in Text-Adventure Games. In *Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)* (pp. 131-138). Association for Computational Creativity.
<https://computationalcreativity.net/iccc20/papers/032-iccc20.pdf>
- Atkinson, T., Baier, H., Copplestone, T., Devlin, S., & Swan, J. (2019). The Text-Based Adventure AI Competition. *IEEE Transactions on Games, 11*, 260-266.
<https://doi.org/10.1109/tg.2019.2896017>
- Atlas, S. (2023). *ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI*. College of Business Faculty Publications.
https://digitalcommons.uri.edu/cba_facpubs/548
- Blizzard Entertainment (2004). *World of Warcraft (PC, Mac OS X) [Video Game]*. Blizzard Entertainment.
- Chen, W.-Y. (2024). *Intelligent Tutor: Leveraging ChatGPT and Microsoft Copilot Studio to Deliver a Generative AI Student Support and Feedback System within Teams*.
<https://doi.org/10.48550/arXiv.2410.10650>
- Chowdhury, M., Dixon, L., Kuo, L., Donaldson, J. P., Eslami, Z., Viruru, R. et al. (2024). Digital Game-Based Language Learning for Vocabulary Development. *Computers and Education Open, 6*, Article ID: 100160. <https://doi.org/10.1016/j.caeo.2024.100160>
- Coady, J., & Huckin, T. (1997). *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge University Press.
- Côté, M., Kádár, Á., Yuan, X., Kybartas, B., Barnes, T., Fine, E. et al. (2019). Textworld: A Learning Environment for Text-Based Games. In T. Cazenave, A. Saffidine, & N. Sturtevant (Eds.), *Computer Games. CGW 2018. Communications in Computer and Information Science* (Vol. 1017, pp. 41-75). Springer.
https://doi.org/10.1007/978-3-030-24337-1_3
- Crowther, W., & Woods, D. (1975-1977). *Colossal Cave Adventure (PDP-10) [Video Game]*. Crowther & Woods.
- DaCosta, B. (2024). Historical Depictions, Archaeological Practices, and the Construct of Cultural Heritage in Commercial Video Games: The Role of These Games in Raising Awareness. *Preservation, Digital Technology & Culture, 53*, 113-132.
<https://doi.org/10.1515/pdte-2024-0028>
- DaCosta, B. (2025). Dynamic Narratives: Exploring Language Acquisition through AI-Driven Text-Based Games. In R. J. Cohen (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 611-616). Association for the Advancement of Computing in Education.
- Dambekodi, S., Frazier, S., Ammanabrolu, P., & Riedl, M. O. (2020). Playing Text-Based Games with Common Sense. In *Wordplay: When Language Meets Games (NeurIPS 2020: The Thirty-Fourth Annual Conference on Neural Information Processing Systems)* (pp. 1-6). Neural Information Processing Systems.
<https://nips.cc/virtual/2020/16679>
- Darvenkumar, T., & Devi, V. A. (2022). Text-Based Game: A Tool to Enhance Critical Reading and Critical Thinking Skills in English Classrooms. *Theory and Practice in Language Studies, 12*, 2326-2332. <https://doi.org/10.17507/tpls.1211.09>
- Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second Language (L2) Gains through Digital Game-Based Language Learning (DGBLL): A Meta-Analysis. *Language Learning & Technology, 26*, 1-25. <http://hdl.handle.net/10125/73464>

- Dong, Y. R. (2023). *COTTAGE: Coherent Text Adventure Games Generation*. Master's Thesis, University of Pennsylvania.
<https://www.cis.upenn.edu/~ccb/publications/masters-theses/River-Yijiang-Dong-masters-thesis-2023.pdf>
- Europass SRL (2023, December 7). *Interactive Fiction: A Valuable Tool in the Classroom*. Teacher Academy. <https://www.teacheracademy.eu/blog/interactive-fiction/>
- Falcó, J. J. M., & Colás, R. M. (2022). The “Room” in Text Adventure Games and Analysis of Real Geographical Locations in Aventuras AD Games. *Transactions of the Digital Games Research Association*, 6, 1-33. <https://doi.org/10.26503/todigra.v6i1.128>
- Furtado, P. G. F., Hirashima, T., & Yusuke, H. (2018). A Serious Game for Improving Inferring in the Presence of Foreign Language Unknown Words. *International Journal of Advanced Computer Science and Applications*, 9, 7-14.
<https://doi.org/10.14569/ijacsa.2018.090202>
- Gee, J. P. (2008). Learning and Games. In K. Salen (Ed.), *The Ecology of Games: Connecting Youth, Games, and Learning* (pp. 21-40). MIT Press.
- Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., & Vandirk, S. (2023, March). *Unlocking the Power of Generative AI Models and Systems Such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers (Hohenheim Discussion Papers in Business, Economics and Social Sciences, No. 02-2023)*. Universität Hohenheim, Fakultät Wirtschafts- und Sozialwissenschaften.
<https://nbn-resolving.de/urn:nbn:de:bsz:100-opus-21463>
- Gygax, G., & Arneson, D. (1974). *Dungeons & Dragons*. TSR. <https://dnd.wizards.com>
- Hainey, T., Connolly, T., Stansfield, M., & Boyle, L. (2011). The Use of Computer Games in Education: A Review of the Literature. In P. Felicia (Ed.), *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches* (pp. 29-50). IGI Global. <https://doi.org/10.4018/978-1-60960-495-0.ch002>
- Harmer, J. (2015). *The Practice of English Language Teaching* (5th ed.). Pearson Longman ELT.
- Haroush, M., Zahavy, T., Mankowitz, D. J., & Mannor, S. (2018). Learning How Not to Act in Text-Based Games. In *6th International Conference on Learning Representations (ICLR)* (pp. 1-4). OpenReview.net.
<https://openreview.net/pdf/6c6d2a550b8842c8ec561f28ce78d717a5f11886.pdf>
- Harshitha, H. (2022). Choose Your Own Adventure: The Evolution of Digital Interactive Fiction and Its Use in Language Pedagogy. *Fortell*, 44, 44-55.
<https://www.fortell.org/wp-content/uploads/2022/10/July-2022-44-55.pdf>
- Hausknecht, M., Ammanabrolu, P., Côté, M., & Yuan, X. (2020). Interactive Fiction Games: A Colossal Adventure. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34, 7903-7910. <https://doi.org/10.1609/aaai.v34i05.6297>
- Hubbard, P. (2009). *Computer-Assisted Language Learning: Critical Concepts in Linguistics* (Vols. I-IV). Routledge.
- Hulstijn, J. H. (1992). Retention of Inferred and Given Word Meanings: Experiments in Incidental Vocabulary Learning. In P. J. L. Arnaud, & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). Palgrave Macmillan.
https://doi.org/10.1007/978-1-349-12396-4_11
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words. *The Modern Language Journal*, 80, 327-339.

- Infocom (1977-1982). *Zork I-III (PC) [Video Game]*. Infocom.
- Jain, V. V. K. (2019). *Deep Reinforcement Learning Algorithms for Playing Text-Based Games from Feedback*. Master's Thesis, McGill University.
<https://escholarship.mcgill.ca/downloads/2v23vz32s?locale=en>
- Jegede, O. O. (2024). Artificial Intelligence and English Language Learning: Exploring the Roles of AI-Driven Tools in Personalizing Learning and Providing Instant Feedback. *Universal Library of Languages and Literatures*, 1, 6-19.
<https://doi.org/10.70315/uloap.ullli.2024.0102002>
- Jerz, D. G. (2007). Somewhere Nearby Is Colossal Cave: Examining Will Crowther's Original Adventure in Code and in Kentucky. *Digital Humanities Quarterly*, 1, 1-88.
<https://www.digitalhumanities.org/dhq/vol/1/2/000009/000009.html>
- Jerz, D., & Thomas, D. (2015, October 6). Cave Gave Game: Subterranean Space as Video-game Place. *Electronic Book Review*.
<https://electronicbookreview.com/essay/cave-gave-game-subterranean-space-as-video-game-place/>
- Kelly, S. W. (2012). Incidental Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1611-1613). Springer. https://doi.org/10.1007/978-1-4419-1428-6_366
- Kerthyayana Manuaba, I. B. (2017). Text-Based Games as Potential Media for Improving Reading Behaviour in Indonesia. *Procedia Computer Science*, 116, 214-221.
<https://doi.org/10.1016/j.procs.2017.10.041>
- Khatibi, E., & Cowie, E. (2013). *Language Learning through Interactive Games*. Thesis, Malmö University College.
<https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1491487&dsid=2024>
- Kim, A. (2018). *Colossal Classroom Adventure: Developing Interactive Fiction for English Language Learners*. Master's Thesis, University of Illinois.
<https://www.ideals.illinois.edu/items/107579>
- Kirill, K. (2024, February 23). *How to Build a Text Adventure Game Powered by Claude and PartyRock*. AWS.
<https://community.aws/content/2cnQIH63FagJq4Y9fdNYao9y2f1/text-adventure-ga?lang=en>
- Kirriemuir, J., & McFarlane, A. (2004). *Literature Review in Games and Learning*. FUTURELAB. <https://telearn.hal.science/hal-00190453/document>
- Klimmt, C., Blake, C., Hefner, D., Vorderer, P., & Roth, C. (2009). Player Performance, Satisfaction, and Video Game Enjoyment. In S. Natkin, & J. Dupire (Eds.), *Entertainment Computing—ICEC 2009* (pp. 1-10). Lecture Notes in Computer Science, Vol. 5709, Springer. https://doi.org/10.1007/978-3-642-04052-8_1
- Kongmee, I., Strachan, R., Pickard, A., & Montgomery, C. (2011). Moving between Virtual and Real Worlds: Second Language Learning through Massively Multiplayer Online Role Playing Games (MMORPGs). In *2011 3rd Computer Science and Electronic Engineering Conference (CEEC)* (pp. 13-18). IEEE. <https://doi.org/10.1109/ceec.2011.5995817>
- Koo, D. (2009). The Moderating Role of Locus of Control on the Links between Experiential Motives and Intention to Play Online Games. *Computers in Human Behavior*, 25, 466-474. <https://doi.org/10.1016/j.chb.2008.10.010>
- Kostka, B., Kwiecieli, J., Kowalski, J., & Rychlikowski, P. (2017). Text-Based Adventures of the Golovin AI Agent. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)* (pp. 181-188). IEEE. <https://doi.org/10.1109/cig.2017.8080433>
- Lancy, D. F., & Hayes, B. L. (1988). Interactive Fiction and the Reluctant Reader. *The English Journal*, 77, 42-46. <https://doi.org/10.2307/818936>

- Lanzi, P. L., & Loiacono, D. (2023). ChatGPT and Other Large Language Models as Evolutionary Engines for Online Interactive Collaborative Game Design. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 1383-1390). Association for Computing Machinery. <https://doi.org/10.1145/3583131.3590351>
- Laufer, B., & Sim, D. D. (1985). Measuring and Explaining the Reading Threshold Needed for English for Academic Purposes Texts. *Foreign Language Annals*, 18, 405-411. <https://doi.org/10.1111/j.1944-9720.1985.tb00973.x>
- Lee, W. R. (1979). *Language Teaching Games and Contests*. Oxford University Press.
- Lyömiö, S. (2017). *Level and Story Design in Choose Your Own Adventure Games*. Bachelor's Thesis, South-Eastern Finland University of Applied Sciences. https://www.theseus.fi/bitstream/handle/10024/126433/Lyomio_Sirja.pdf?sequence=1
- MacCallum-Stewart, E., & Parsler, J. (2007). Controversies: Historicising the Computer Game. In A. Baba (Ed.), *Proceedings of the 2007 DiGRA International Conference: Situated Play* (pp. 203-210). Digital Games Research Association. <https://dl.digra.org/index.php/dl/article/view/296/296>
- Madotto, A., Namazifar, M., Huizinga, J., Molino, P., Ecoffet, A., Zheng, H. et al. (2020). Exploration Based Language Learning for Text-Based Games. In C. Bessiere (Ed.), *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* (pp. 1488-1494). International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2020/207>
- Maity, S., & Deroy, A. (2024). *Generative AI and Its Impact on Personalized Intelligent Tutoring Systems*. <https://doi.org/10.48550/arXiv.2410.10650>
- Marková, D. (2023). *Language Learning through Adaptive Graded Readers*. Master's Thesis, Charles University, Faculty of Arts, Department of English Language and Didactics. <https://dspace.cuni.cz/handle/20.500.11956/184219>
- Mitchell, A., & Savill-Smith, C. (2004). *The Use of Computer and Video Games for Learning: A Review of the Literature*. Learning and Skills Development Agency.
- Montfort, N. (2005). *Twisty Little Passages: An Approach to Interactive Fiction*. MIT Press.
- Nagai, T., & York, J. (2021). A Comparison of Linear and Interactive Fiction on Vocabulary Acquisition, Reading Comprehension and Engagement. In T. D. Cooper (Ed.), *Proceedings of the JALTCALL2021 Conference* (pp. 70-90). Japan Association for Language Teaching. <https://doi.org/10.37546/jaltsig.call.pcp2021-06>
- Nash, B. L. (2024). Love and Learning in the Age of Algorithms: How Intimate Relationships with Artificial Intelligence May Shape Epistemology, Sociality, and Linguistic Justice. *Reading Research Quarterly*, 59, 624-631. <https://doi.org/10.1002/rrq.549>
- Nassaji, H. (2011). Issues in Second-Language Reading: Implications for Acquisition and Instruction. *Reading Research Quarterly*, 46, 173-184. <https://doi.org/10.1598/rrq.46.2.5>
- Neville, D. O., Shelton, B. E., & McInnis, B. (2009). Cybertext Redux: Using Digital Game-Based Learning to Teach L2 Vocabulary, Reading, and Culture. *Computer Assisted Language Learning*, 22, 409-424. <https://doi.org/10.1080/09588220903345168>
- Office of Educational Technology (2023). *Artificial Intelligence and the Future of Teaching and Learning*. U.S. Department of Education. <https://tech.ed.gov/ai-future-of-teaching-and-learning/>
- OpenAI (2023a). *Does ChatGPT Tell the Truth?* <https://help.openai.com/en/articles/8313428-does-chatgpt-tell-the-truth>
- OpenAI (2023b). *Is ChatGPT Biased?* <https://help.openai.com/en/articles/8313359-is-chatgpt-biased>

- OpenAI (2024a). *ChatGPT-4 [Software]*. <https://openai.com/blog/introducing-gpts>
- OpenAI (2024b). *Introducing GPTs*. <https://openai.com/blog/introducing-gpts>
- OpenAI (2025a). *ChatGPT [Software]*. OpenAI.
- OpenAI (2025b). *ChatGPT Can Now See, Hear, and Speak*. <https://openai.com/index/chatgpt-can-now-see-hear-and-speak/>
- OpenAI (2025c). *ChatGPT Pricing*. <https://openai.com/chatgpt/pricing/>
- OpenAI (2025d). *Introducing GPT-4.5*. <https://openai.com/index/introducing-gpt-4-5/>
- Ørnholt, T. A. (2023). *Enjoyment of Interactive Fiction Compared to Non-Interactive Fiction, in a Norwegian EFL Classroom*. Master's Thesis, American University of Agder. <https://uia.brage.unit.no/uia-xmlui/bitstream/handle/11250/3073034/no.uia%3ain-spera%3a143763411%3a36698203.pdf?sequence=1&isAllowed=y>
- Osborne, P., Nömm, H., & Freitas, A. (2022). A Survey of Text Games for Reinforcement Learning Informed by Natural Language. *Transactions of the Association for Computational Linguistics*, 10, 873-887. https://doi.org/10.1162/tacl_a_00495
- Pereira, J. (2013). Video Game Meets Literature: Language Learning with Interactive Fiction. *e-Teals*, 4, 19-45. <https://ojs.letras.up.pt/index.php/et/article/view/4043>
- Peterson, M. (2012). Language Learner Interaction in a Massively Multiplayer Online Role-Playing Game. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 70-92). Palgrave Macmillan. https://doi.org/10.1057/9781137005267_5
- Pope, J. (2010). Where Do We Go from Here? Readers' Responses to Interactive Fiction: Narrative Structures, Reading Pleasure, and the Impact of Interface Design. *Convergence: The International Journal of Research into New Media Technologies*, 16, 75-94. <https://doi.org/10.1177/1354856509348774>
- Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The Effectiveness of Games for Educational Purposes: A Review of Recent Research. *Simulation & Gaming*, 23, 261-276. <https://doi.org/10.1177/1046878192233001>
- Rankin, Y. A., Gold, R., & Gooch, B. (2006). Playing for Keeps: Gaming as a Language Learning Tool. In J. Finnegan, & M. J. Barr (Eds.), *ACM SIGGRAPH 2006 Educators Program on SIGGRAPH '06* (p. 44). ACM Press. <https://doi.org/10.1145/1179295.1179340>
- Rankin, Y. A., Gold, R., & Gooch, B. (2016, September 7). *Evaluating Interactive Gaming as a Language Learning Tool [Video]*. YouTube <https://www.youtube.com/watch?app=desktop&v=thtRIm0ZiM4>
- Rankin, Y. A., McNeal, M., Shute, M. W., & Gooch, B. (2008). User Centered Game Design: Evaluating Massively Multiplayer Online Role-Playing Games for Second Language Acquisition. In D. Schwartz, K. Schrier, C. Swain, & M. Wagner (Eds.), *Sandbox '08: Proceedings of the 2008 ACM SIGGRAPH Symposium on Video Games* (pp. 43-50). Association for Computing Machinery. <https://doi.org/10.1145/1401843.1401851>
- Reinders, H. (2012). *Digital Games in Language Learning and Teaching*. Palgrave Macmillan.
- Reinders, H. (2016). Digital Games and Second Language Learning. In S. Thorne, & S. May (Eds.), *Language, Education and Technology* (pp. 1-10). Encyclopedia of language and Education, Springer. https://doi.org/10.1007/978-3-319-02328-1_26-1
- Reinders, H., & Wattana, S. (2015). The Effects of Digital Game Play on Second Language Interaction. *International Journal of Computer-Assisted Language Learning and Teaching*, 5, 1-21. <https://doi.org/10.4018/ijcallt.2015010101>
- Risi, S., & Togelius, J. (2020). Increasing Generality in Machine Learning through Proce-

- dural Content Generation. *Nature Machine Intelligence*, 2, 428-436.
<https://doi.org/10.1038/s42256-020-0208-z>
- Rixon, S. (1981). *How to Use Games in Language Teaching*. Macmillan.
- Sargsyan, M. (2013). *A Case Study of Challenges Faced by EFL Learners Reading Interactive Fiction*. Master's Thesis, American University of Armenia.
<https://dspace.aua.am/xmlui/handle/123456789/1266>
- Scholz, K. W. (2015). *Online Digital Game-Based Language Learning Environments: Opportunities for Second Language Development*. Doctoral Dissertation, University of Waterloo. <https://uwspace.uwaterloo.ca/items/3d4f5e0c-e9e1-493d-b39a-055604dc1ece>
- Sharma, L. (2008). "Colossal Cave Adventure—Hindi Mein!": *An Experimental Approach to Second Language Acquisition through Computer Games*. Master's Thesis, University of California. <https://nideffer.net/classes/278-09/ace-theses/luv-thesis.pdf>
- Shelton, B. E., Neville, D., & McInnis, B. (2008). Cybertext Redux: Using Interactive Fiction to Teach German Vocabulary, Reading, and Culture. In *Proceedings of the International Conference of the Learning Sciences (ICLS)* (pp. 128-129). International Society of the Learning Sciences.
- Shi, Z., Xu, Y., Fang, M., & Chen, L. (2023). Self-Imitation Learning for Action Generation in Text-Based Games. In A. Vlachos, & I. Augenstein (Eds.), *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 703-726). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2023.eacl-main.50>
- Sykes, J., & Reinhardt, J. (2012). Conceptualizing Digital Game-Mediated L2 Learning and Pedagogy: Game-Enhanced and Game-Based Research and Practice. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching. New Language Learning and Teaching Environments* (pp. 32-49). Palgrave Macmillan.
https://doi.org/10.1057/9781137005267_3
- Tafazoli, D. (2021, November). Review of the Book *Gameful Second and Foreign Language Teaching and Learning: Theory, Research, and Practice*, by J. Reinhardt. *ROSHD FLT Journal*, 134, 33-35. <https://www.researchgate.net/publication/356508281>
- Telltale Games, & Skybound Games (2012-2019). *The Walking Dead [Video Game]*. Released on Multiple Platforms. Telltale Games & Skybound Games.
- Terzic, I., Drobnjak, A., & Boticki, I. (2023). Designing Educational Personas Using Generative AI. In J. L. Shih, et al. (Eds.), *Proceedings of the 31st International Conference on Computers in Education* (pp. 961-963). Asia-Pacific Society for Computers in Education. <https://doi.org/10.58459/icce.2023.1488>
- Thu, D. N. A., & Nhu, P. Q. (2021). Using Interactive Fiction in Extensive Reading: Action Research for English-Major Freshmen at a University in Central Vietnam. *Journal of Science, Hue University of Education*, 4, 7-17.
<https://vjol.info.vn/index.php/TCKH-DHH/article/view/64176>
- Towle, B., & Zhou, K. (2023). Learn What Is Possible, Then Choose What Is Best: Disentangling One-to-Many Relations in Language through Text-Based Games. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 4955-4965). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.findings-emnlp.364>
- Triyason, T. (2023). Exploring the Potential of ChatGPT as a Dungeon Master in Dungeons & Dragons Tabletop Game. In *Proceedings of the 13th International Conference on Advances in Information Technology* (pp. 1-6). ACM.
<https://doi.org/10.1145/3628454.3628457>

- Tsai, C. F., Zhou, X., Liu, S. S., Li, J., Yu, M., & Mei, H. (2023). *Can Large Language Models Play Text Games Well? Current State-of-the-Art and Open Questions*. <https://doi.org/10.48550/arXiv.2304.02868>
- Walton, N. (2019, November 21). AI Dungeon 2: Creating Infinitely Generated Text Adventures with Deep Learning Language Models. <https://pcc4318.wordpress.com/2019/11/21/ai-dungeon-2-creating-infinitely-generated-text-adventures-with-deep-learning-language-models/>
- Wang, M., & Guo, W. (2023). The Potential Impact of ChatGPT on Education: Using History as a Rearview Mirror. *ECNU Review of Education*, 8, 41-48. <https://doi.org/10.1177/20965311231189826>
- Wei, L. (2023). Artificial Intelligence in Language Instruction: Impact on English Learning Achievement, L2 Motivation, and Self-Regulated Learning. *Frontiers in Psychology*, 14, Article ID: 1261955. <https://doi.org/10.3389/fpsyg.2023.1261955>
- wh0am1-dev (2015). *Colossal Cave Adventure Source Code! (1976) [Source Code]*. GitHub. <https://github.com/wh0am1-dev/adventure>
- Wright, T. B., & Weible, J. L. (2024). Current Attitudes on Digital Interactive Fiction and Text Adventure Games within Learning Contexts: A Systematic Literature Review. *Journal of Research in Science, Mathematics and Technology Education*, 7, 23-51. <https://doi.org/10.31756/jrsmte.713>
- Wyatt, C. (2018). The Natural Accommodation of Interactive Fiction: How Text-Based Games Remove Barriers to Participation. In C. Chen, K. Purzycki, & L. Wilkes (Eds.), *Proceedings of the Annual Computers and Writing Conference* (pp. 27-37), Colorado State University. <https://wac.colostate.edu/docs/proceedings/cw2018/wyatt.pdf>
- Yao, S., Narasimhan, K., & Hausknecht, M. (2021). Reading and Acting While Blindfolded: The Need for Semantics in Text Game Agents. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, & Y. Zhou (Eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 3097-3102). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.naacl-main.247>
- Yao, S., Rao, R., Hausknecht, M., & Narasimhan, K. (2020). Keep CALM and Explore: Language Models for Action Generation in Text-Based Games. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 8736-8754). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.704>
- Yin, X., & May, J. (2019). Comprehensible Context-Driven Text Game Playing. In *2019 IEEE Conference on Games (CoG)* (pp. 1-8). IEEE. <https://doi.org/10.1109/cig.2019.8847954>
- Yin, X., & May, J. (2020). *Zero-Shot Learning of Text Adventure Games with Sentence-Level Semantics*. <https://doi.org/10.48550/arXiv.2004.02986>
- Yuan, X., Côté, M.-A., Sordoni, A., Laroché, R., Tachet des Combes, R., Hausknecht, M., & Trischler, A. (2018, October). Counting to Explore and Generalize in Text-Based Games. In *European Workshop on Reinforcement Learning (EWRL)* (pp. 1-16). Microsoft Research. <https://www.microsoft.com/en-us/research/publication/counting-to-explore-and-generalize-in-text-based-games-2/>
- Zelinka, M. (2018). *Using Reinforcement Learning to Learn How to Play Text-Based Games*. Master's Thesis, Charles University. <https://arxiv.org/pdf/1801.01999>

Zhang, X. (2023). Not Just Playing: Potential in Enhancing Second Language Learning through Games. *SHS Web of Conferences, 180*, Article No. 02001.

<https://doi.org/10.1051/shsconf/202318002001>

Zhiwei, L., & Yin, C. H. (n.d.). *AI in Game: A Study about Understanding AI in Game Development*. ART-SCI MLRI. SM3807: Machine Learning and Robotic Interaction Studio.

https://mlrobotics.wordpress.com/wp-content/uploads/2023/04/ml_paper_hui-cheuk-yin-lin-zhiwei.pdf