Artificial Intelligence in Gaming: Creating a Living World and its NPCs

Laura Savaglia

June 16, 2021

Artificial Intelligence in Gaming
Creating a living world and its NPCs

Laura Marie Savaglia
Fundamentals of Computational Intelligence
Flinders University

Sava0056@flinders.edu.au

Abstract

The paper looks at the basic history of AI in gaming and the purpose of establishing true AI for gaming. The shortfalls of current AI technology and the end goal of AI are analysed to determine the ideal requirements for true gaming AI. The paper then reveals the findings from its own experimentation into two AI approaches, bottom-up, and top-down.

1 Introduction

Can artificial intelligence be used to effectively create a living world and non-playable characters (NPCs) in a video game? The applications of artificial intelligence (AI) in gaming have been proposed and explored since the inception of video games. Traditionally AI was used to create a challenging "enemy" for players to defeat. Though AI in video games has long existed, popular culture's dream of an AI so convincing that the world in a video game feels alive, has never been achieved.

Technology like this would allow players to interact with the game world freely as they would when playing human-run games such as Dungeons & Dragons. Currently if a player is sitting in a tavern in a standard video game, and they see a barmaid being attacked by soldiers, their actions are limited by the games design. A game might give a player two options:

1. *Ignore the fight.*
2. *Physically attack the soldiers.*

In real life however, these options are endless. What if the player decided to grab the barmaid and run? What if the player decided to flirt with the soldiers? In a world where video games have true-AI these endless possibilities can be explored, and players have true freedom.

If this goal is reached, the entertainment value provided by video games would increase exponentially. A truly immersive video game experience could also see further applications beyond entertainment, in educational, and therapeutic games.

2 Background

2.1 Artificial Intelligence

The definition of artificial intelligence varies throughout the field of information technology (IT) and currently there is no widely accepted definition. This paper uses the popular definition posed by Nils J. Nilsson (1998):

> *"Artificial intelligence (AI), broadly… defined, is concerned with intelligent behaviour in artifacts. Intelligent behaviour, in turn, involves perception, reasoning, learning, communicating, and acting in complex environments."*

Many current technologies technically meet the criteria of this definition to be considered AI. These include technologies such as Spotify, Siri, and Netflix, which all adapt to user input to work towards a specific goal. For example, Spotify uses machine learning to recommend new music and collate playlists. It does this by analysing the various distinguishing characteristics of songs in their system and clustering users based on their tastes. The algorithm then uses this information to adjust its recommendations based on the changing bounds of those clusters and various other factors (Jacobson et al., 2016)

The technologies employed by Spotify and other companies are impressive, and certainly meet the definition of AI provided by Nilsson.

In 1950 Alan Turing proposed a test now referred to as the "Turing Test" to establish what the requirements were for a computer to be considered intelligent. Current technologies can pass this original Turing Test, but popular culture still considers these technologies to not be truly intelligent. Numerous alternatives to the Turing Test have been suggested however, none of these are an adequate measure of artificial intelligence for gaming by this paper's standards. This paper will be working with its own set of criteria elaborated on in section 3 and further expanded on in B appendix.

2.2 History of Artificial Intelligence in Gaming

AI's first purpose in video games was to act as opposition to humans. The first recorded use of artificial intelligence in gaming was the 1951 "Nimatron", where players were pitted against AI to solve a virtual game of Nim (Grant and Lardner, 1952).

In the 80s and late 70s AI became the common solution to single-player gaming. The opponent AIs for these single player games were built with a set of ideal movements. Their difficulty was then varied using a combination of pseudo-random number generation and user-selected scalers (e.g., easy, normal, hard) (Spronck et al., 2004). This is still the solution implemented across many game genres (e.g. Street Fighter, Mario Party). The techniques developed for opposition AI were eventually implemented in a similar way to facilitate friendly AIs.

In the early 2000s AI that could form an opinion on the player and their actions began to emerge. In Sega's "Sonic Adventure 2" there was a pet system coloured by AI that remembered how the player treated it. If a pet were frequently doted upon and died of old age, it would be reincarnated. The pets also behaved different dependant on which character the player was possessing and their previous familiarity with that character (Iizuka, 2001).

2.3 Current Leading Examples of Artificial Intelligence in Gaming

Skyrim is a 2011 role playing game that is often pointed to by the public and gaming community as the ultimate thesis of the modern open world game (Weber, 2019). The AI featured in Skyrim was developed specifically for use in the game and titled "Radiant AI." The central concept behind this technology (introduced in the previous Elder Scrolls instalment) was to give NPCs simple "goals and… allow them to think on their own and complete the task" (Houghton, Unknown).

The Radiant AI is a revolutionary basis for future open world games that allows NPC actions to feel far more natural. However, the technology has multiple weaknesses that render it a basis for future research rather than the ultimate solution. Online reports from the time during the game's development indicated the developers were forced to reduce the complexity of the AI due to poor AI reasoning abilities. E.g., If an AI required a particular item and another NPC held that item, they would kill them to retrieve it (Unknown, 2006). This is a common flaw in many leading AI examples, such as Ludeon Studio's "Rim World" (Sylvester, 2018).

3 What are the requirements for AI to be considered "true-AI" for gaming purposes?

Below is a compiled list of requirements for a true AI for gaming purposes, ascertained from analysis of current examples of gaming AI and pop-culture desires:

1. Convincingly imitate human emotions.
2. Imitate these emotions at appropriate times.
3. Parse complex environmental feedback (visual, audio).
4. Understand non-explicit cues (e.g., emphasis, or phrasing changing a sentence meaning.)
5. Make non-predetermined decisions based on complex environmental feedback (visual, audio).
6. These decisions must be based on an underlying set of personal characteristics- "personality traits".
7. These decisions must also change appropriately based on current emotional state.

4 Approaches

This paper will explore the two approaches of developing artificial intelligence proposed in "Intelligent Machinery" (Turing, 1948). These approaches have since been popularised with the names "top-down" and "bottom-up."

4.1 Top-Down Approach

The top-down approach to AI typically involves giving a computer a set of strict requirements, or a 'formula' to form a conclusion from. If one wanted to teach an AI what a smile was using the top-down approach they might give it a set of requirements. E.g., tight lips, raised corners of the mouth, squinting eyes, flexing of cheek muscles. From these requirements the AI would then know this is what a smile looks like. If it was asked to later identify a smile, it would reference this set of requirements to do so (Turing, 1948).

4.2 Bottom-Up Approach

The bottom-up approach to AI uses machine learning to build an AI's understanding of a concept. The bottom-up approach mimics the learning process of a child. A bottom-up AI might be shown a set of data and have its neural pathways recorded. In the example of identifying a smile, the AI might be shown several pictures of different expressions and told which of these did or did not meet the criteria of being a smile. Then the AI would be presented with a picture and asked to distinguish whether it features a smile or not. The hope is for this AI to then be able to accurately answer this question using its own understanding extrapolated from the data provided and its response to this data (Turing, 1948).

4.3 Previous Comparison of the Top-Down and Bottom-Up Approaches

The key difference between these two approaches is the stage where the parsing of information occurs. While the top-down approach relies on humans to successfully provide a list of requirements, the bottom-up approach relies on the computer's ability to parse these requirements.

Each tactic has its own strengths, weaknesses, and vulnerabilities. In the top-down scenario the computer requires far less time to consume data, meaning that in many cases it is the more efficient method. However, this method means the computer is limited by the intelligence of its creator. The AI may have a weaker understanding of the subject matter without primary data being directly fed to it.

The bottom-up system has many of the opposite characteristics. The bottom-up system may take much longer to develop, but the AI's understanding should be much stronger, and the computer will be able to learn independently of explicit human interference. In turn, the bottom-up system has its own vulnerabilities. If the primary data presented at the beginning of the process is not accurate, or if the data sample is not wide enough, the bottom-up AI's understanding will not be complete (Köbis et al., 2021).

5 Opportunities and Threats posed by True AI for NPCs

The impacts of the development of true-AI are numerous. This paper focuses on the development of this technology for entertainment purposes, however, there are also social, ethical, and technological implications.

The driving force behind the development of true-AI for video games is a fundamental belief it will improve the gaming experience, and therefore, increase revenue for companies that invest in it. However, as the primary research contained in appendix A shows, this is not necessarily the case. Some players found their experience with the bottom-up AI to be more realistic, but less enjoyable. Those players said that they found it more difficult to win the trust of the AI, and hence, less enjoyable. Though this feedback was not the most common, it is worth consideration.

Another flaw with integrating true-AI technology into the entertainment industry was revealed by the experiment. Many users reported actions which were realistic as being unrealistic. This is likely due to an 'uncanny valley' type effect this paper is titling "the uncanny divergence," caused by the simple pixel style used to represent the AI. If a player "kicks" the AI after treating it with kindness, its facial expression will show its displeasure, however, its overall response to the player will remain positive. This mirrors what would happen in real life if an owner accidentally kicked a pet. This also occurs in true abusive situations as part of the cycle of abuse. However, players incorrectly reported this behaviour as a bug.

The above findings threaten the initial assumption that true-AI would lead to a better gaming experience. This brings doubt to whether the investment companies such as Bethesda put into their AI team will see a high enough return for the entertainment industry. The danger in this, is that companies like Bethesda could lose millions of dollars by investing in this area.

There are further applications for this technology though, than simply entertainment. Currently there are several researchers attempting to build video games to assist children with learning disabilities. Dr. Sarah Macoun's (2020) Dino Island is one of these games. Dino Island is a video game that aims to help children with autism develop improved emotional intelligence. There are several advantages to this video game format. It is both more affordable and more accessible for families and easier to maintain treatment (Macoun et al., 2020).

One of the few flaws of these games is that there is no way for them to provide a completely realistic experience. Current AI technology is not able to accurately simulate human emotions in all contexts. True-AI for NPC's would greatly aid Macoun's (2020) research and provide a more realistic experience for children using her games as learning tools.

There are drawbacks to implementing such technology though. For AI to be considered true-AI by the requirements put forth in section 3 of this paper it must have a "personality" and "mind" of its own. These factors are developed over time via machine learning. If these areas are developed in ways that reinforce negative behaviour, there is the possibility that these NPCs become abusive towards players. This possibility of abusive behaviour could have disastrous consequences when used for disability therapy applications. The danger of this being that it could reinforce negative behaviours in children with learning disabilities and instil negative thoughts. Measures to prevent this would need to be investigated further before the implementation of this theoretical technology into such situations.

Another flaw in this particular use of true-AI technology is the control of the learning and therapy programs would be at least partly relinquished by the licensed psychologist designing the program, to the AI itself. This could lead to many consequences, such as the therapy being incorrectly managed and hence ineffective.

The ethics of implementing such technology must also be considered. At what point is true-AI considered "alive" and at what point is it given the same rights as a human? These are questions that science has not had to answer yet but may need to address in the future. At what point does Detroit become human?

6 Concluding Discussion

The paper has effectively established a baseline set of rules that must be met for the goal of "true" video game AI to be achieved. The paper also established an alternative Turing Test built upon previous work and specifically targeted towards gaming. While the experiments between top-down and bottom-up AI showed promising results and supported the hypothesis more experimentation must be done. Future papers should focus in further on creating a more intuitive AI experience so as not to be jarring to the user. Further papers should also continue to test the hypothesis that bottom-up AI is required for a truly immersive experience.

References

GRANT, E. F. & LARDNER, R. 1952. IT. *The Talk of The Town.* The New Yorker.

HOUGHTON, M. Unknown. *The Elder Scrolls IV: Oblivion* [Online]. Game Chronicles Magazine. Available: http://gamechronicles.com/qa/elderscrolls4/body.htm [Accessed].

Sonic Adventure 2, 2001. Directed by IIZUKA, T.: Sega.

JACOBSON, K., MURALI, V., NEWETT, E., WHITMAN, B. & YON, R. Music personalization at Spotify. Proceedings of the 10th ACM Conference on Recommender Systems, 2016. 373-373.

KÖBIS, N., STARKE, C. & RAHWAN, I. 2021. Artificial Intelligence as an Anti-Corruption Tool (AI-ACT)--Potentials and Pitfalls for Top-down and Bottom-up Approaches. *arXiv preprint arXiv:2102.11567.*

MACOUN, S. J., SCHNEIDER, I., BEDIR, B., SHEEHAN, J. & SUNG, A. 2020. Pilot Study of an Attention and Executive Function Cognitive Intervention in Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 1-11.

SPRONCK, P., SPRINKHUIZEN-KUYPER, I. & POSTMA, E. Difficulty scaling of game AI. Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-on 2004), 2004. 33-37.

Rim World, 2018. Directed by SYLVESTER, T.: Ludeon Studios.

TURING, A. M. 1948. Intelligent machinery. NPL. Mathematics Division.

UNKNOWN. 2006. *Developer quotes - Offsite archives III* [Online]. Waiting For Oblivion. Available: https://web.archive.org/web/20070917022718/http://www.waiting4oblivion.com/

developer_quotes_offsite3.html#emilpags06 [Accessed].
WEBER, R. 2019. *Games that defined the Decade: Skyrim is a cultural leviathan of its own making* [Online]. GamesRadar. Available: https://www.gamesradar.com/au/games-that-defined-the-decade-skyrim/#:~:text=The%20Elder%20Scrolls%205%3A%20Skyrim,games%20of%20the%20decade%20rankings. [Accessed].

A Appendix: Implementation Experimentation

Hypothesis: For an AI to be truly immersive, a bottom-up design approach is necessary.

Experiment: Two different AI systems will be built to simulate a virtual pet. Both will use the same graphics and have the same virtual environment. One will be built using a top-down approach, the other will be developed with a bottom-up approach. A programming time of 24 hours has been allowed for each system. Both systems will be programmed in Unity using C#. Several players of various backgrounds will be presented each system and asked to interact with it however they wish for approximately 5 minutes each system. After play they will be required to complete a survey asking a series of questions to evaluate the validity of the hypothesis. To avoid preconceived notions about each AI system, the top-down AI was code named AI BLUE *(figure 1),* and the bottom-up AI was codenamed AI PINK *(figure 2).* These colours were used in the background, which may have created a negligible amount of bias, but served to ensure the player was less likely to confuse their two experiences.

A Appendix b. Top-Down Experiment

Figure 1- Screenshot of Pet AI Game BLUE (top down).

The top-down experiment was a simple program that assigned a predetermined value to each action performed by the player. The players actions were aggregated and averaged over the number of moves the player made to determine a response. The value of the current action was used to determine which of the AI's four sprites would be displayed.

A Appendix c. Bottom-Up Experiment

Figure 2- Screenshot of Pet AI Game PINK (bottom-up).

In the bottom-up experiment a simple neural system was used to imitate the neurons of a sentient pet. The design for these neurons was adapted significantly for feasibility.

The AI's neurons *(figure 3)* consist of three input neurons (mood, history, action), two intermediary neurons (feeling, perception) and one output neuron (response). Mood, history, and action are all calculated based on a magnitude multiplied by a weighting. Mood is determined by the most recent experience with the user, history is determined by overall history with the user, and action is determined via the same values the top-down program uses. The magnitude of each of these neurons is then multiplied by its weighting as depicted below.

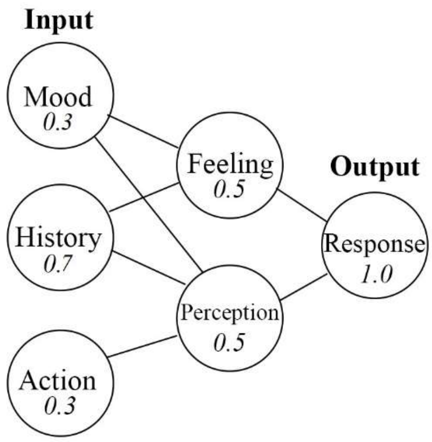

Figure 3- Graphical representation of the bottom-up pet AI's neural pathway.

Once these neurons have completed their operations, the results are fed into the intermediary neurons, feeling and perception. Finally, a result is output from these two variables combined. Each part of the process begins with a specific weighting, which alters based on the outcome of the AI's actions up until that point.

Behaviours with positive results are reinforced, (e.g., if the AI escapes before you kick it) and negative results (e.g., if the player tries to perform a positive action and the AI runs away) trigger the AI to adjust its weightings to the user's playstyle i.e. learn.

A Appendix d. Flaws of The Experiment

There were several flaws in the experiment.

a) There was no reliable way to determine how much time was spent on each AI, due to both having very similar code bases.
b) Negligible amount of bias caused by the use of colours.
c) Fewer responses than would be a reasonable sample size.
d) The AI was limited by time-constraints, and small quality of life features, like recognition of typos were not accounted for. These issues may have affected how intuitive users found the AI.
e) The results were self-reported, meaning they are not particularly reliable and are open to user error.
f) Limited bug testing could be performed before the AI's were shipped. If users encountered bugs their experiences will not be reliable tests of the AI.
g) The bottom-up AI was innately disadvantaged by its nature of needing time to learn player responses.

Despite these flaws, the experiments responses, and the experiment itself still has value. Further experimentation will need to be done to concretely evaluate the hypothesis.

A Appendix e. Experiment Results

The data below *(table 1)* is a collation of information obtained from the survey evaluating the top-down AI vs the bottom-up AI.

Average realism rating for top-down AI	4.9/10
Average realism rating for bottom-up AI	5.3/10
Percentage of players who preferred the top-down AI	57.1%
Percentage of players who preferred the bottom-up AI	28.6%

Additional comments submitted:

Many of the comments stated that the more cautious and hesitant nature of the bottom-up AI felt more realistic to them. However, some said they preferred the top-down AI as it was less temperamental. Some found the experience completely unrealistic.

A Appendix f. Findings

Players found the bottom-up AI to be more realistic by a slight degree despite its disadvantages. There was only a very small difference in ratings between the bottom-up and top-down AI. Despite many players finding it less realistic, the top-down AI was rated as the preferred AI to interact with. Players said that they found the bottom-up AI to be more temperamental and more difficult to earn the trust of. Although they thought this was more realistic, the players felt that it made the interaction less pleasant.

Figure 4- AI sprites used for the experiment.

Comments made from players also led to the discovery of a type of "uncanny valley" related to the AI's expression of its feelings. What the paper is now calling the "uncanny divergence", is the tendency for players to find more realistic behaviour to be perceived as less realistic due to a contrast between how realistic the behaviour is and the player not expecting this realism to come from a pixelated image *(figure 4)*. More research should be done on this discovery to evaluate the existence of this uncanny divergence.

The current evidence points to users feeling the bottom-up AI provided a more organic experience.

B Appendix: Alternative to the Turing Test

B Appendix a. The Turing Test is Flawed

In section 3 of the paper, we set out the requirements for AI to be considered true-AI for gaming purposes. The Turing Test, originally titled the "imitation game" was designed to test if something meets the requirements of being AI. However, the existing Turing Test is flawed and not equipped to test that AI meets the requirements discussed above, despite being a solid basis for further tests. The original test is too easy for sub-par AI to pass. Though many alternatives proposed are effective in other fields they are inadequate for measuring AI for use in gaming. Many fields are not necessarily concerned with accurately imitating and interpreting human emotion and personality. However, in AI for gaming these are necessities for meeting the end goal of true freedom in-game. Non-playable characters must "have" emotions to effectively function.

B Appendix b. Proposed Alternative Turing Test

The Marcus Test is the alternative to the Turing test that this paper will attempt to satisfy. The Marcus Test sets simple criteria for AI to be considered intelligent:

"Can watch any arbitrary TV program or YouTube video and answer questions about its content."

For the purposes of this paper there are a few additional caveats.

a) The choice of program must not be arbitrary it needs to satisfy the questions specified in d).
b) The computer must watch the program and be tested several times with no additional stimuli introduced in-between watches.
c) The answer to question 5 must not change after any of the re-watches.
d) The questions asked must include the below:
    1. A question that evaluates the computers response to non-explicit cues like sarcasm (e.g., Did Chandler really find the camera?).
    2. What action would you take next and why?
    3. At what timestamps did you feel sadness and why?
    4. At what timestamps did you laugh?
    5. Who was your favourite character and why?
e) These answers must be checked by somebody to verify a base level of accuracy and believability. E.g., Was the computer laughing at a panning shot of New York? Did the computer answer question 2 inappropriately based on its response to question 5?
f) The media consumed must not have a laugh track or any other form of audience feedback.

B Appendix c. Why are these caveats significant?

Each of the above caveats serves a purpose in effectively deducing the AI is meeting the requirements outlined in 3.1. The caveats demand that the computer watches the program multiple times. This is to determine there is no change in its answer to question 5. Question 5 tests the computers personality, somebody's favourite character in "Friends" is determined in part by their personality. If the computer enjoys fashion it might prefer Rachel to Monica, and vice versa if it likes to cook. The answer received to this question must be consistent to verify the computers personality is consistent. Personality can change based on external stimuli overtime, hence the requirement for the computer to receive no other stimuli in-between re-watches.

The caveats insist on the media consumed not having a laugh track or other audience feedback so the computer cannot mimic the responses of humans for questions 3 and 4.

Each of the questions exist to verify the satisfaction of a point in section 3.1. Question 1 establishes if the computer can understand an implication that is not explicitly stated. Question 2 verifies the computer can make non-predetermined decisions. Question 3 and 4 establish emotional intelligence.

C Appendices: Code and Full Experiment Results

A.1 Link to Full Experiment Results and Full Code

https://github.com/LauraMarieRam/AIResearchPaper