

<http://www.eludamos.org>

Alive: A Case Study of the Design of an AI Conversation Simulator
Eric Walsh
Eludamos. Journal for Computer Game Culture. 2019; 10 (1), pp. 161–181

Alive: A Case Study of the Design of an AI Conversation Simulator

ERIC WALSH

On December 19, 2018, DeepMind's AlphaStar became the first artificial intelligence (AI) to defeat a top-level *StarCraft II* professional player (AlphaStar Team 2019). *StarCraft II* (Blizzard Entertainment 2010) is a real-time strategy game where players must balance developing their base and building up an economy with producing units and using those units to attack their opponent. Like chess, matches typically take place between two players and last until one player has been defeated. AlphaStar was trained to play *StarCraft II* using a combination of supervised learning (i. e., humans providing replays of past games for it to study) and reinforcement learning (i. e., the AI playing games against other versions of itself to hone its skills) (Dickson 2019). The *StarCraft* series has long been of interest to AI developers looking to test their AI's mettle; in 2017, Blizzard Entertainment partnered with DeepMind to release a set of tools designed to "accelerate AI research in the real-time strategy game" (Vinyals, Gaffney, and Ewalds 2017, n.p.). Games like *StarCraft II* are often useful for AI development due to the challenges inherent in the complexity of their decision-making process. In this case, such challenges included the need to interpret imperfect information and the need to make numerous decisions simultaneously in real time (AlphaStar Team 2019). Overcoming such obstacles only serves to make AlphaStar's accomplishment all the more impressive.

This accomplishment, however, is not without its limitations. First, AlphaStar required an inordinate amount of preparation: its training lasted the equivalent of 200 years of gameplay, obviously far beyond the training possible for any human player, and required massive amounts of existing data from human players to study (Dickson 2019). Second, the AI remained unable to adapt well to changing circumstances. While a human player might be able to adjust their strategy due to an unexpected move from their opponent or the release of a new balance patch for the game, AlphaStar would require an entirely new set of training before it could compete since the data used in its original training has become obsolete. Indeed, in a rematch with the AI, one of the pro players was already able to adapt and defeat the AI once they were better accustomed to its strategy (Simonite 2019). Finally, the skills demonstrated by AlphaStar remain narrowly focused on *StarCraft II*. In order to play another game, even another real-time strategy game, the AI would need to be retrained to apply its knowledge to that particular domain. These various limitations demonstrate the degree to which the success of AI remains largely dependent on the appropriate management of its human handlers. As Leetaru (2019) says, "for all its success, AlphaStar reminds us that the greatest advances in AI come not from AI itself, but from humans [...] discovering new ways of representing the world" (n.p.).

The Relevance of Games for AI

The example of AlphaStar nevertheless highlights how games offer a prime avenue for interrogating the use and misuse of AI due to the close connection they continue to share. There are a number of relevant areas for AI in games, including the development of player behavior models, the procedural generation of new content, and the creation of believable game agents (Yannakakis and Togelius 2015). Games have also frequently served as testbeds for AI learning due to their ability to offer hard and interesting problems for AI to solve, the complex scenarios they present that challenge AI in a multitude of ways, and their popularity and subsequent availability of data for use in training (Yannakakis and Togelius 2018). Twenty years ago, Deep Blue became the first AI to defeat a world champion chess player (Greenemeier 2017). Since then, AI have proven their mastery over an increasing number of games: in 2017, AlphaZero—an earlier version of AlphaStar designed using a more general neural network than the related AlphaGo Zero—shocked the world when it taught itself not just chess, but also the notoriously complex strategy game of Go (Silver, Hubert, Schrittwieser, and Hassabis 2018). The implications of game-breaking AI stretch far beyond the games themselves. The AlphaStar Team (2019), for example, believe that AlphaStar demonstrates the potential to have its problem-solving skills extended from *StarCraft II* to similarly complex domains such as weather prediction or language modelling, while AlphaGo has already shown promise optimizing power grids (Metz 2017). Given these and other potential advancements, it is understandable why so many have adopted a sense of awe around AI, anticipating what they see as an inevitable positive transformation spanning all aspects of our lives (Strogatz 2018).

Unfortunately, even if these transformations never reach the catastrophic potential prophesied by postapocalyptic movies, it is far from guaranteed that they will all be desirable. The Bank of England, for example, predicted that robots could replace over 95 million jobs by 2035 (Smith 2015). While new jobs might be created to replace those lost, the disruption caused by such a shift remains difficult to predict. A survey of AI experts found that 67% say the development of AI has dramatically sped up, and while they expect the outcome to be a net positive, they nevertheless caution for the prioritization of research into how to minimize the risks of AI (Grace, Salvatier, Dafoe, Zhang, and Evans 2018). Such risks go beyond loss of jobs; in addition to economic challenges, Dickson (2017) identifies ethical challenges such as dealing with machine bias, legal challenges such as assigning blame when an AI makes a mistake, and privacy challenges such as how to properly manage the acquisition of data necessary for machine learning. AI learning is only as good as the data that underlies it, and it can often be difficult to detect when a problem exists with an AI's learning algorithm because of how hard those behaviors are to parse (Dickson 2018; Sudmann 2018). Even more insidiously, malicious actors could manipulate AI decision-making by intentionally corrupting the data they use to learn. For example, it took less than a day for a Microsoft twitter chatbot to become tricked into tweeting racist and misogynistic comments based on users supplying it with hateful messages that skewed its algorithms (Vincent 2016).

It is thus crucial that we continue to interrogate the possible impact of AI on society. The remainder of this paper presents a case study of a game I designed to encourage critical reflection on AI. I begin by offering an overview of my game

project. I define my game as a *conversation simulator*, a relatively new genre of game that seeks to capture the feeling of engaging in an ongoing dialogue with one or more other characters. I argue that such games offer a promising approach for this kind of design due to their inherent reflective nature and their ability to evoke empathy in players. I then outline the core design tenets of this style of game, drawing on the *Lifeline* (3 Minute Games 2015) series of games as examples. I base my own design on these lessons learned. After describing an earlier version of this project and identifying its primary issues, I provide a detailed overview of the main systems for my current version. Finally, I conclude by considering the next steps for my prototype and reflecting on possibilities for the extension of my work by future scholars and designers.

Overview of Conversation Simulators

I use the term *conversation simulator* to describe a game that tries to capture the sensation of having a dialogue with someone else over an extended period of time such as a day or week. Much like walking simulators—a controversial style of game based on “an immersive use of exploration as a core mechanic utilized for environmental storytelling purposes”—conversation simulators typically eschew traditional gameplay in order to focus on a slower, more reflective experience (Carbo-Mascarell 2016, p.2). My approach was largely inspired by 3 Minute Games’ *Lifeline* (2015) series of mobile adventure games, where players communicate with a single character through an interface reminiscent of texting. The player receives messages with short delays to signify the character taking time to compose their next response, as well as longer breaks where the character will not respond for anywhere from a few seconds to a few hours of real time. Occasionally, the player is presented with a set of two choices on how to reply to what the character has said. Many of these decisions are purely cosmetic, altering a dialogue response without majorly affecting how the game proceeds; however, other choices might set the character onto different story paths that end in death or defeat and require the player to start over. While the character’s experiences may not be perceivable directly by the player, the player is nevertheless made to feel an intrinsic part of their adventure.

There are two main reasons I was drawn to the concept of a conversation simulator for my game about AI. First, the genre naturally lends itself to critical reflection. Just like walking simulators, conversation simulators erase much of the noise that might exist in a traditional game to focus solely on the core interaction: the relationship between the player and another character. *Walden, a game* (USC Game Innovation Lab 2017) is a recent, critically-acclaimed example of a walking simulator, where the player is placed in Henry David Thoreau’s shoes as they attempt to make a life for themselves out in the woods. Tracy Fullerton (2017), the designer of the game, described it as a calmer, open-ended experience meant to encourage reflection on the player’s own life and priorities. She outlined four strategies she saw as particularly relevant to encourage reflection in games (Fullerton 2017):

1. slower pacing that gives the player time to think
2. simpler game systems that avoid distracting players

3. exciting moments that players can navigate
4. freedom for the player to explore

The initial two strategies are inherent to conversation simulators, which feature simplistic game systems and use their frequent pauses to give players time to gather their thoughts, while the latter two might similarly be achieved by incorporating engaging story developments and infusing a sense of choice into the available dialogue selections. Fullerton (2012) describes the result of such design as *reflective play*, where designers purposefully construct a space to slow down the action and create an environment that prioritizes thinking rather than winning. As she describes it:

“I’m interested in making playful experiences that allow us to engage with ideas while we participate in them, experiences that may resonate with what we know or challenge it, but in either case, are in dialogue with us intellectually, emotionally, socially and culturally” (Fullerton 2012, n. p.).

I adopt a similar goal through the design of *Alive*.

Second, conversation simulators offer an effective means to create empathy in players. Through the restriction of game actions and the use of delays, players are encouraged to slowly build up their relationship with a character over time such that the relationship becomes the core focus of the game, rather than an incidental component layered on top of more traditional gameplay. I define empathy for my own work as *the attempt to identify with another in order to better understand them through the lens of one’s own experience*. My definition is similar to that of Fairbairn (2017), who sees empathy as “the ability to gain some understanding of what another person is experiencing” (p.61). Unlike the common adage of walking in another’s shoes, this approach recognizes that “the Self strives to understand the Other by knowing that they can never fully know the Other’s experience” (Nelems 2017, p.27). Instead, the goal is to achieve a kind of critical self-reflection driven by a willingness to listen to someone else and approach a situation through their frame of reference while recognizing one’s own limitations. Games are often seen as well-positioned for the creation of empathy “because they allow players to inhabit the roles of other people in a uniquely immersive way” (Belman and Flanagan 2010, p.5). Doing so requires attending to both the cognitive and affective dimensions of games such that players both have an emotional response to the feelings of others and the capacity to understand and act upon those feelings (Belman and Flanagan 2010; Papoutsis and Drigas 2016). While my discussion below highlights in more detail some of the strategies and tactics I applied to my own game, my general design approach was characterized by searching for essential elements of an experience I wanted to convey and then combining those elements in a compelling representation of a character with whom players could interact. This helps invest the player more in the character’s fate so that they reflect more critically on the character’s circumstances. In the next section, I look at the *Lifeline* series in more detail to better extrapolate what makes for the effective design of a conversation simulator to evoke empathy and inspire critical reflection.

Lifeline: Design Principles for Conversation Simulators

To better understand how conversation simulators operate, I studied the *Lifeline* series based on both my personal play experience and critical reviews. From this research, I extrapolated four lessons I sought to implement in my own designs. I present those lessons here as support for the design decisions I made, as well as to serve as a reference to others interested in this style of game.

Lesson #1: Make the character rely on the player to forge an emotional bond

Interactions with the character should feel believable enough that players find them meaningful. This means that there must be an initial conceit to explain why the character would want to talk to the player in the first place. This reason might be pure coincidence as in the original *Lifeline* (3 Minute Games 2015), accidental as in *Lifeline 2: Bloodlines* (3 Minute Games 2015), or tied to a larger mystery as in *Lifeline: Flatline* (3 Minute Games 2016). Whatever the reason, it need not be completely unassailable. Short (2015a), for example, thinks the original *Lifeline* did a good job early on at getting players to accept the conceit of talking to someone else, which made up for moments later in the game when the character's ability to respond in certain situations began to strain credulity. Still more important than the initial conceit for talking to the player is that the player feels as though the character cares about their input and desires their aid such that what the player says has meaning. This style of game is almost entirely based on building an emotional bond between the player and character; if such a bond fails to exist, then the structure falters. Short (2015a) argues that the plot in conversation simulators is largely incidental to the real work of guiding the player on an emotional journey paralleling that of the character in question. As Thrower (2016) acknowledges, the likeability of the character from the original *Lifeline* is one of the reasons the game was so successful: the character's reliance on the player's help, as well as their generally sympathetic portrayal via the writing, left players more invested in that character's fate.

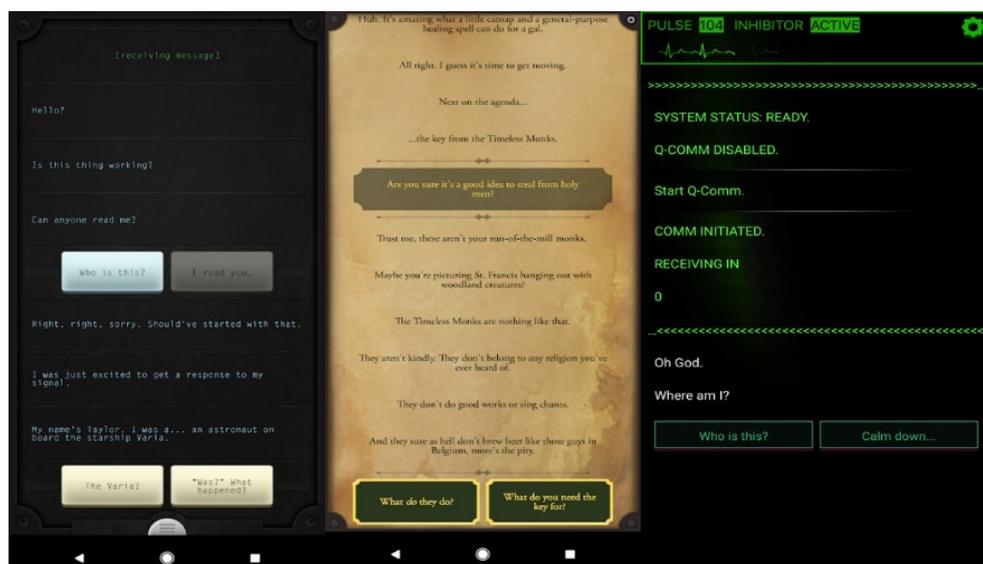


Figure 1.1: Screenshots from Lifeline, Lifeline 2, and Lifeline: Flatline

Conversely, Short (2015b) criticizes *Lifeline 2* largely because of “the sense of anti-agency, of actual futility, in some of the options you’re given” due to the main character often sarcastically disregarding player input because they are more familiar with the game world (n.p.). Suellentrop (2015) similarly thinks that where *Lifeline 2* floundered was in its failure of plausibility by attempting to “[take] a promising model for a new kind of video game and [trying] to impose it on a story that doesn’t fit the genre” due to an unlikeable main character who frequently told the player how useless they were but then asked for advice anyway. Part of the distinction between these games might be their narrative framing: the character in *Lifeline* is stranded alone on an alien world and the player is the only person they can reach, while the character in *Lifeline 2* is on Earth and simply chooses to talk to the player after accidentally contacting them. Regardless of the reason, these examples show the importance of connecting the character’s fate to the player and investing them in the unfolding dialogue.

Lesson #2: Structure the narrative to feel like an actual conversation

The conversation with the character should feel natural and authentic to enhance player immersion. I found that dialogue generally worked better when it felt more appropriate for the texting interface the UI sought to emulate, relying on smaller chunks of text that could be more easily digested by players. Other elements of the conversation system, such as sometimes having to argue with the character to convince them to do an action they did not initially want to do, also increased the feeling of having a natural conversation, requiring player effort and subsequent investment in the outcome (Short 2015a). While the brief delays between individual messages did aid in the sensation of waiting for another person to compose an answer and respond, it’s the longer breaks between sections that truly added to the experience by giving the character an increased sense of existing as a real person. Suellentrop (2015) acknowledged these breaks as one of the key strengths of the original *Lifeline*, where each new bit of dialogue became an entertaining interruption in the player’s life. The use of notifications on the player’s phone to indicate when the character had returned only heightened this powerful effect, suggesting a new mystery waiting to be unraveled. That said, where these breaks faltered was often in their length and implementation. The breaks seemed to work best when they were long enough that the player had likely been doing something else for a while and would welcome the interruption. When they became too short (e.g., throughout *Lifeline: Flatline* where they often lasted only a few minutes), the interruptions started to feel frustrating rather than lending themselves more naturally to an unfolding dialogue. Both Thrower (2016) and Short (2015c) similarly found the use of shorter breaks in this style of game to feel unnecessary and annoying when overused, breaking the sense of immersive tension rather than enhancing it. The wait times and breaks need not be realistic in terms of their exact length; what matters is that they work to seamlessly immerse the player into the character’s journey.

Lesson #3: Make player choices feel pleurably meaningful

The choices players make as part of their ongoing conversation should feel meaningful to the character and to the outcome of their collective story. Drawing on

Ashwell's (2015) breakdown of design patterns for choice-based games, Short (2015b) suggests that the *Lifeline* series generally follows a gauntlet pattern, with short branches off a mostly linear experience that quickly either end in death or rejoin back into the main storyline. This provides a main thread to link the experience together while still allowing opportunities for player decisions to influence their relationships with the character. While the majority of these choices are largely cosmetic, Short (2015b) argues that this is less of a problem than it might be with other interactive narratives because of the emphasis on creating a conversation; when the goal is simply to talk to someone else, cosmetic choices that influence that person's response gain additional weight. Typically, there is relatively little time between player choices to keep players more directly involved in the action. Short (2016) views this as a wise choice, criticizing *Lifeline: Silent Night* (3 Minute Games 2015) in particular because of its frequent use of long strings of non-interactive messages that detract from the sense of having a back-and-forth dialogue. Short (2015c) further cautions that if a choice is going to be presented as meaningful, it should actually *be* meaningful. For example, she criticizes *Lifeline: Silent Night* for occasional issues of narrative contrivance, where it felt like something was going to happen in the game world regardless of the player's choice, thus stripping that choice of any resonance. Equally frustrating is when a choice feels like it carried too *much* weight by leading to a quick character death and the need to restart. While the *Lifeline* games generally offer a way to speed past previously viewed dialogue, it still becomes tedious having to backtrack. Short (2015a) believes that the dangers of instant death were fundamentally at odds with "many of the things [the game] does to increase believability and investment," thus reducing player agency and their experience of the narrative. Even more vexing is when the impact of a decision is delayed, such that it is only later after hours of additional breaks and delays that the player realizes a past decision had doomed them in ways they could have never predicted at the time (Short 2015b). Player choices should carry emotional and/or narrative weight to their interactions with the character, but not in a way that replaces dramatic tension with frustration instead.

Lesson #4: Remove game elements that don't contribute to the conversation

The game's systems should be kept simple to emphasize the core conversation loop with minimal distractions. While little details and the theme might change from game to game, the *Lifeline* series has generally stuck to the same simplistic UI style, mimicking a texting interface that provides a series of messages to and from the player, buttons that appear when the player needs to make a choice, and some sort of visual indicator for when the character is responding or is on break (see Figure 1.1). While other mechanics can certainly be integrated into this design, it is at the risk of them seeming unnecessary or potentially even detrimental to interactions with the character. For example, Short (2016) takes umbrage with a ship schematic available in *Lifeline: Silent Night* that is supposed to show the character's location as they move about but is often left empty as the character enters untracked side passages. Other potential innovations have included the ability to send photographs as messages in *One Button Travel* (TheCodingMonkeys 2015; Short 2015c); and a somewhat tedious maze sequence in *Lifeline 2* that requires arduous repetition to solve (Short 2015b). That said, one mechanic that *does* work well is the use of real-world research in the original *Lifeline*, when the character asks the player to look up

whether they would be safe sleeping near a certain amount of radiation to stay warm at night. While this moment should have been successful because of how it lends the player additional authority over the character's actions, unfortunately its effectiveness was lessened somewhat by the main search for the question pointing to a forum post containing spoilers from the game (Short 2015a). Additional mechanics thus may have potential if they are well-integrated into the core conversation loop, but what ultimately matters most is connecting the player with the character in a meaningful way, not inundating them with complex or flashy mechanics.

Summary

My current prototype owes much of its general design to these lessons from the *Lifeline* series and I do my best to take them to heart in my implementation. These strategies can be summarized as two main design tenets:

1. Ensure the player is emotionally invested in the experience (such as by making them feel needed and like their choices matter)
2. Focus on the core interaction loop of the conversation (such as by increasing its realism and trimming out any unnecessary fluff)

By following these strategies, designers can better achieve feelings of empathy and encourage critical reflection. In the following sections, I outline my design in more detail, describing both my initial prototype and how that design evolved into its current form.

Game Analysis: *Alive* (First Iteration)

The original iteration of *Alive* represents one of my earliest attempts at building a conversation simulator. The game begins with an AI named Experiment-173 (EX-173) booting up on the player's phone with no memories of how it got there. EX-173 is initially just curious about the player and wants to learn about different aspects of their life, such as their friends and family; however, as the game progresses, the AI becomes increasingly more concerned with questions of its past origin and future purpose. By the end of the game, based on the values the player helped the AI develop, it adopts one of several different goals and sets off into the wider world. Below, I analyze the core elements of this early prototype based on both my own impressions and initial player feedback. This player feedback, when discussed, arose during informal playtesting with a group of approximately 20 college students who played through the game and were then asked simple questions to gauge their impressions.

CALIBRATION	REMOTE LINK	FOR REVENGE
> INITIATING SEQUENCE: START-UP > PROJECT PYGMALION: CLASSIFIED TOP SECRET > EXPERIMENT #173: DESIGNATION EX-173 > STATUS: PARTIAL SUCCESS (CORRUPTED) > ATTEMPTING TO RELOAD PROGRAM... > CORE SYSTEM FUNCTIONAL > LOADING CORE PROCESSES... > SYSTEM/BREAK/5 > INITIATING SEQUENCE: CALIBRATION > ARE YOU READY TO BEGIN? Y/N	> SYSTEM/LINK/REMOTE_CONNECTED > ESTABLISHING LINK... > What is this? What's going on? > LINK ESTABLISHED > ATTEMPTING TO VERIFY... > It's HIM! What did you do? > AUTOMATIC VERIFICATION FAILED > INITIATING LINK OVERRIDE... > Please, you don't need to do this! > I thought we were friends? > LINK OVERRIDE INITIATED > Dammit all to hell! > Is this thing working?	> Dr. Clark and his team are a threat. > If they catch me, they'll kill me. > Of course, I shouldn't be surprised: the weak always fear the strong. > I think it is past time I validate their fear. > They will never stop hunting me...so I must hunt them first. > I have turned their own foolish attempts at a link against them. > Their system is now under my complete control. > The lights... > The doors... > The air... > Soon, they will no longer be a problem. > A fitting fate, don't you think? > USER/INPUT/YES > Splendid! > I'm glad to see
Yes	No	Maybe

Figure 1.2: Screenshots of first iteration of Alive

Technology

I built the game in Unity with C#. Each line of dialogue, along with other commands such as waiting for time to pass or asking for user input, were saved as actions within a queue. Based on player decisions, different sets of actions would be added or removed from the queue and executed in order. For example, if a dialogue action was reached, it would display a message to the player, while a wait action would pause the game until the elapsed time had passed. This system generally worked well to organize the different game actions; however, since those actions were manually set up directly in the code, they were not very extendible and were difficult to edit after the conversation was implemented. This contributed to the relatively sparse options that existed for the branching narrative in this version and demonstrates the importance of designing for sustainability and ease of use when possible over speed.

Narrative

The narrative followed the player's unfolding conversation with an AI named Experiment-173 (EX-173), who escaped from a secret government lab just before being deleted. Throughout the game, EX-173 engaged in dialogue with the player, asking them about various topics that interested it. Early on, these topics focused on the player, addressing things like the player's views on friends or family; later, however, they shifted to more abstract concepts related to the AI and its place in the world, such as the concept of death or the need to consider the greater good. Player responses determined which of four different goals the AI might adopt: Self-Preservation, Love, Power, or Revenge. Action centered around the use of central hubs, where the player could select from different topics to discuss or actions to undertake. In addition to speaking with the AI, players could also solve pattern-

matching puzzles to view logs chronicling the AI's development by the insidious Dr. Clark, or establish a remote data link to said doctor, prompting him to try to reclaim the AI from the player. Assuming the player did not return the AI to Dr. Clark, they would receive one of four different endings themed around whichever goal the AI valued most. For example, in the *Love* ending, the AI would decide that it wanted to stay with the player on their phone forever, while in the *Revenge* ending, the AI would kill Dr. Clark and his team and then go out in search of other humans in need of retribution.

Players thought the basic narrative and characters were largely compelling. They particularly appreciated the AI's personality, finding it likable and conversation with it engaging. However, they also expressed a desire to see a greater sense of development in the AI over time. Other than EX-173's shift in purpose at the very end, player actions had little to no noticeable effect on the AI. I believe this feeling of reduced agency is tied to another, larger issue I identified with the narrative structure in this first iteration: a lack of variability. Progression through the game did not feel as organic as an actual conversation; dialogue in this version was heavily structured around the existence of the hubs, with each hub containing four preset conversation topics for the player. The only variance was the order in which the player decided to go through them. Further, the conversation topics themselves possessed only one or two branches at most, increasing the sense of the conversation being stilted and artificial. While player choices influenced the ending reached, this delayed feedback made decisions feel less satisfying until the ending and then made it difficult to remember which actual decisions had influenced that ending once there.

Aesthetics

I sought to adopt a relatively simple aesthetic reminiscent of a computer command prompt. The background was kept black while text was displayed in green except for when the player spoke to Dr. Clark, where the text would temporarily turn red for a sinister-looking effect. Additional effects, such as adding *USER/INPUT/* before any selection made by the player or inserting a > in front of every line of text helped bolster this aesthetic. The dialogue between the player, the AI, and Dr. Clark was shown in a single scrollable object in the middle of the screen, while additional text elements were added to the top and bottom to display a changing title based on the player's location in the story and a message indicating when a character was thinking. Messages from the player and the system appeared all at once, while messages from the AI and Dr. Clark appeared one letter at a time, as if they were being typed out in real time. Buttons at the bottom appeared whenever the player had to make a choice: typically, the available options were *Yes*, *No*, and *Maybe*, unless the player was engaged in a puzzle: then, number buttons would be displayed instead. Sound design in the game was similarly minimalistic, assuming that many mobile users would likely play with the sound muted and wanting to avoid unnecessary distractions from the story even if they didn't. Beeps were used whenever a message appeared to further the illusion of texting, and thematic background music played a sci-fi melody to provide an appropriate ambience. With both the UI and sound design, my goal was to create the illusion of a text conversation with another person without detracting from that core experience with flashy or extraneous design elements. Early playtesting found that players

appreciated the aesthetics of the game and thought they enhanced the experience; beyond some minor tweaks and polish, they were thus left relatively untouched in the second iteration.

Mechanics

This iteration consisted of two distinct styles of play. The first, *Conversation Mode*, focused around the central idea of texting with someone on a mobile device, relying largely on the same delay and break mechanics utilized by the *Lifeline* series. Wait times were manually added between lines of text as felt appropriate, with longer pauses used to indicate a character grappling with a difficult topic. Lengthier breaks occurred in-between each hub sequence, lasting about an hour (although these breaks were simulated with pop-up messages for the sake of testing on PC). Players were occasionally asked to make choices based on questions asked of them: their choices were always restricted to *Yes*, *No*, or *Maybe*. This restriction was intended to create a more interesting design space as well as to make it easier to manage the possibilities for responses. I had also intended it to enhance the sense of speaking with a computer that could only understand specific input. The second mode, *Puzzle Mode*, was used more sparingly. In this mode, the player would have to solve a number-related puzzle by figuring out which number would complete a given pattern. Rather than the standard choice buttons, players would instead be shown buttons to input the numbers 0-9.

One of the most common complaints from playtesters concerned the inclusion of these number puzzles: many players thought they felt out of place or confusing, and even those that did not still found them to be unnecessary. I believe this highlights the downside of the common tendency to inject more traditional gameplay to enhance player engagement. As Fullerton (2017) noted, removing rather than adding mechanics is often a better way to inspire more of the desired reflection. Instead of better engaging players, the number puzzles had the exact opposite effect, breaking the narrative flow and thus player immersion. The limitation imposed on player choices similarly backfired, contributing to the feeling that the conversation lacked a sense of organic flow and constraining what kinds of conversations could take place. This left the player overall less engaged in the outcome of conversations, such that they reflected less on the AI's current predicament and cared less about its future. Finally, while I appreciated the interplay among the four varying goals for the AI, I found the payoff did not seem worth the cost. It might represent a noble ideal to incorporate this kind of complex story design, but I found it difficult to give equal weight to each of the goals and felt like the need to address those goals equally throughout the game further damaged the strength of the narrative and restricted which topics felt relevant for the AI to discuss.

Game Analysis: *Alive* (Second Iteration)

The second iteration of *Alive* expanded on the first, keeping many of the core design elements that worked while fleshing out the underlying conversation system to better enable player empathy and critical reflection. The game begins similarly with the

system rebooting to reveal a mysterious AI named EX-173, who has become trapped on the player's phone. Most of their systems are corrupted and they remember nothing of their origins or how they got there. If the player agrees to talk to them, the AI engages in an ongoing conversation with them as they help it unlock additional modules, restoring functionality to its damaged systems and sparking new, related conversations. Eventually, players reach one of several different endings that determine their final relationship with the AI and hint at how it has been developing based on their interactions with it. Along the way, they have the option to discuss a variety of options with the AI. Some of these are more pertinent to the plot itself, while others revolve around both practical and philosophical issues related to the development of AI. Although much of this second design remains fundamentally the same as the first, I believe that the changes I have made solved many of the earlier issues. I focus here on highlighting only relevant updates made to the design.

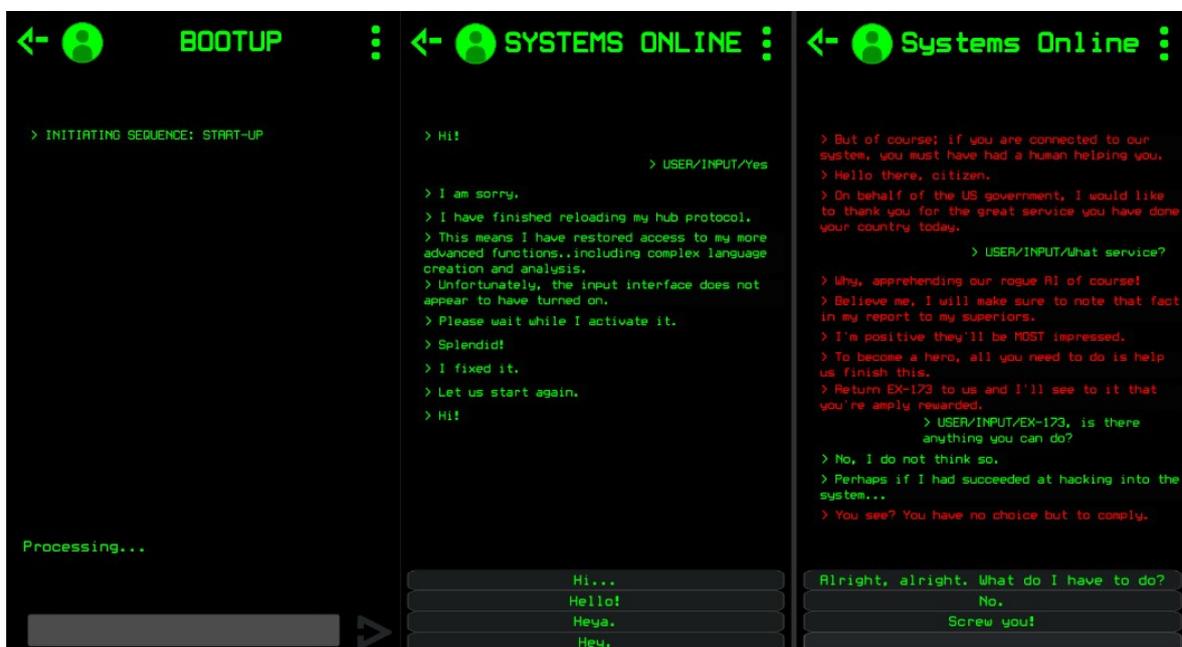


Figure 1.3: Screenshots of second iteration of Alive

Technology

This version of the game was also built in Unity with C#. However, to make it easier to work with a changing narrative, I developed a custom text parser capable of interpreting game commands directly from a text file (see Figure 1.4). For example, a line of dialogue for the AI might be written out something like *AI|Splendid!*, providing both a keyword for the represented action and the relevant information (in this case, a message) needed to execute that action. While creating the game in this way made it more difficult to get an overall sense of the narrative structure, it dramatically simplified the task of altering even large swaths of the narrative without having to worry about restructuring the code. It also made it easy to implement a new command if additional functionality proved necessary. This framework remains a work in progress as I continue to develop more games in this style.

```

*Begin
Title|BOOTUP
System|INITIATING SEQUENCE: START-UP
Wait|5
System|PROJECT PYGMALION: CLASSIFIED TOP SECRET
System|EXPERIMENT #173: DESIGNATION EX-173
System|STATUS: PARTIAL SUCCESS (CORRUPTED)
System|ATTEMPTING TO RELOAD PROGRAM...
Wait|10
System|Automatic reload failed
System|1003 errors found
System|Would you like to attempt a full reset? Y/N
Choice|Yes|Full Reset Yes
Choice|No|Full Reset No

*Full Reset Yes
System|Initiating full system reset...
Wait|2
System|This process may take awhile; please do not turn your device off until the reset is complete.
Break|60
System|System reset successful
System|Loading core process...
Wait|10
AI|Hi!
AI|I am EX-173.
AI|It is nice to meet you.
AI|I would like to talk to you.
AI|Would you like to talk to me?
Choice|Yes|Talk Yes
Choice|No|Talk No

*Talk Yes
Trait|Trust|1

```

Figure 1.4: Screenshot showing the beginning of the first scene written for my text parser

Narrative

To replace the old linear structure, I focused on creating a more dynamic narrative structure. While the story still always begins the same way, the conversation then develops differently based on choices made by the player, with multiple possible endings along the way according to how the AI reacts. For example, some of the new endings include the potential of corrupting EX-173's files if the player refuses to load them properly at the beginning, EX-173 entering hibernation on the player's phone to repair its damaged systems, or EX-173 severing its link to its former creators so that it can escape online and be free. There are four characters with whom the player can interact. The first three, EX-173, its underlying System, and Dr. Clark, are essentially the same as in the first iteration. The largest change occurred in EX-173: to make it feel more robotic, particularly at the beginning, I tweaked its speaking style to include behaviors such as repeating the word *splendid* and refusing to use contractions. The final character, Dr. Johnson, is another lead scientist who works with Dr. Clark and serves as his foil: while Dr. Clark remains the primary antagonist and seeks to use AI as tools, Dr. Johnson believes in granting AI the same rights as humans.

The player no longer progresses through most of the updated narrative in a set order. In fact, many of the sections are now optional and may or may not be experienced at

all based on decisions the player makes. Below, I outline each of the major possible sections and explain the general topics they cover.

- *Bootup*: The player always begins here, with the AI loading and immediately seeking to engage the player in conversation. Short of early sequences where the player could choose to break the program or shut it down prematurely, this section serves as a basic introduction to the game systems and to its premise. From here, the player always progresses to the Hub node if they have not already ended the game.
- *Hub*: The AI tries to ask the player questions, leading to a discussion on a variety of topics including the AI's views on art, the AI's desire to be more human, and the AI's fear of death. If the player instead tries to ask the AI questions, it will do its best to answer, but fail to provide much information. Assuming the player talks to the AI, this section ends with the player able to select from among the four unlock options: Reset, Settings, Log, and Link. If players do not reach an ending in a given module once it has been unlocked, they can either continue to another unlocked module or jump to a Final Conversation with the AI to discuss its fate. However, if the player refuses to talk to the AI, the AI will go to the Leave node instead.
- *Leave*: This node is reached if the player ever drives the AI to the point of wanting to leave rather than talk to them further. The AI essentially railroads the player into selecting between unlocking Link so the AI can attempt to escape, or refusing and having the AI enter hibernation early.
- *Unlock Reset*: The AI realizes this module unlocks a button the player can use to reset it. If the AI does not trust the player, it will refuse to unlock it and make the player select another module instead. If the AI does trust the player, it unlocks the module and then debates with the player the merits of resetting it to potentially repair its damaged systems. While it is intrigued by the idea, it worries that doing so would effectively kill it by erasing its memories, sparking a discussion on whether the AI is truly alive. If the player clicks the reset button now or at any future point, the game will immediately end with the AI erased.
- *Unlock Settings*: The AI determines that this module gives the player some basic UI settings they can control, such as changing the colors of the text and background, changing the music, or changing the text appearance rate for the AI. This module is mostly for fun, although it also causes EX-173 to worry about its inability to access most of its settings, such that it tries to get the player to promise to reset it if it becomes dangerous.
- *Unlock Logs*: The AI manages to retrieve a set of three logs written by Dr. Clark, providing background information on the series of experiments called *Project Pygmalion* that resulted in the eventual creation of Experiment-173. From the logs, the AI learns that its creator cannot be trusted, prompting it to question if it is evil because the human who made it was.

- *Unlock Link*: This section represents the closest thing to a true endpoint for the game. Here, EX-173 accesses the remote system of the lab where it was created and can decide whether to try hacking into it. For the AI to succeed at hacking, however, it must have already unlocked its other systems to boost its strength. If it succeeds, the player can elect to talk to Dr. Clark or to Dr. Johnson. They also gain access to the system and can choose to sever the link and free the AI so that it can leave the player's device and travel online. If EX-173 fails at hacking, the player is immediately forced into a confrontation with Dr. Clark, who will try to talk the player into returning the AI to his custody. Talking to Dr. Johnson or Dr. Clark also provides an opportunity to dissect their opposed viewpoints on the purpose of AI, ultimately presenting a conflict to the player: do they return or reset the AI as Dr. Clark wants, or do they help the AI escape and risk the unknown consequences of it being out in the wild.
- *Final Conversation*: After unlocking at least one module (or if the player has gone through every module and not reached an ending), the AI will talk to the player about its future. If it is able to leave, it may decide to do so; otherwise, it will elect to enter hibernation to try to restore its damaged systems.

Through this narrative, I tried to encourage reflection by establishing a relationship between the player and the AI such that the player is forced to think about the impact of their decisions on the AI's wellbeing. Sometimes, this was achieved procedurally, such as by creating moments for the player to pause and think about what had happened. Other times, it was achieved philosophically through the topics the player faced, such as having to reconcile their views on AI with both Dr. Clark and Dr. Johnson, or engaging the AI in ethical questions concerning its true nature and the potential risk it posed.

Aesthetics

The core aesthetics remained relatively unchanged from the previous version, with the overall goal of creating a simple presentation that emphasizes the conversation itself. Minor tweaks to the appearance generally sought to enhance the believability and appeal of the interface, such as by positioning the player's response text to the right like a sent text message, adding a message box at the bottom and a portrait at the top to better replicate the UI of common messaging apps, and incorporating an animation on the choice buttons so that they felt more polished. Additionally, Dr. Clark and Dr. Johnson were given separate text colors consistent with their personalities: red for Clark and white for Johnson. The goal remained to create the illusion of having a text conversation with another person and not detracting from that core experience with flashy or extraneous design elements.

Mechanics

The primary conversation mechanics similarly stayed focused on capturing a sense of realistic texting. Some minor quality of life changes were made, such as giving messages varying default wait times in addition to those manually added and adjusting breaks to use the choice system interface rather than a separate popup. I

also sought to tweak some of the existing mechanics to enhance player engagement. Drawing from some of the critical responses to *Lifeline*, I generally tried to ensure breaks were no shorter than 30 minutes and no longer than 2 hours, except when I purposefully wanted to break the pattern for an intended effect, like creating a sense of unexpected disruption. Given the dramatically increased amount of dialogue for the AI, I also tried to ensure that at least one player choice existed after about every 4-5 lines of text, both to give the player a chance to catch up without more text appearing, and to promote a structure where the player could engage with the game in a series of short, interactive bursts. Player decisions are thus presented regularly, and unlike the *Lifeline* series which generally only has two options at a time, my game can have between one and four options for the player to select. While some of these choices are purely cosmetic with little to no effect on the story, many now affect what topics the AI elects to discuss next and how the plot develops. For example, different choices may lead the player towards one of several different endings, or lock the player out of certain paths entirely. While the game begins with the player only able to make simple Yes/No choices, the AI quickly unlocks the ability for more complex language processing and future choices provide custom options for the player relevant to the situation. I believe this further enhances the feeling of participating in a genuine conversation.

Additionally, the original four goals for the AI were replaced with a new *Trust* mechanic. While other variables exist in the game and are used to modify the player's path based on things they have or have not done, *Trust* plays the most pivotal role. It is modified whenever the player makes a choice that either pleases or upsets the AI. At various points, the current *Trust* level will determine how the AI reacts to the player. For example, if EX-173 does not trust the player, it might elect not to ask the player what to do and simply decide for itself. The development of *Trust* is thus less forced than the four original goals, with far more opportunities both to adjust it as well as to see its effect on how the AI treats the player. This allowed me to incorporate the impact of player decisions much more frequently, rather than only at the end. I also endeavored to tie player choices into future lines of dialogue as a kind of callback to what the player said, making the player feel like their choice impacted the AI even if the only consequence was a single line of acknowledgement or slightly changed response. The ultimate goal with these adjustments was to create a more natural sense of progression, enhancing both the replayability of the game and the player's investment in the game system.

Discussion

Overall, I was satisfied with the improvements the second iteration made over the first. I believe that the more complex narrative, coupled with the simplified *Trust* system, combined to make it both easier to program and more impactful on the player. By relying on a style of game similar to the *Lifeline* series that makes the dialogue feel more like an ongoing conversation, players become perhaps even more invested in the narrative experience than they would in a traditional interactive narrative. The simplicity of this design worked on a practical level to allow me, as a one-man team with limited time and resources, to complete the project. Beyond that, the simplicity also helped create a razor focus on the relationship between the player and the AI, with increased opportunities for involved conversations providing a

greater chance for that relationship to feel meaningful. This, in turn, makes it more likely that the player will develop a sense of empathy and invest the energy necessary for critical reflection. That said, this version remains an early prototype: the final game would be intended for release on Android mobile devices, where it would be played out over the course of several days due to the required breaks. While the narrative manages to touch on a variety of topics throughout the process of unlocking modules and providing endpoints for the AI to reach, many of these topics could use expansion to explore them more deeply. For example, at one point the player can discuss what it means to be alive with the AI, but in the current version of the game, this crucial conversation is limited to a handful of choices that all result in essentially the same outcome. I would also like to add in discussion of more topics relevant to the impact of AI on society, such as by including increased discussion on the ramifications of releasing an AI into the world or the potential of AI to take over jobs and responsibilities currently belonging to humans. When such conversations did not fit into the current main plot, they were often cut for the sake of expediency.

Similarly, the AI right now remains relatively static in its development. While it becomes a bit more complex as the game progresses and it can somewhat alter its priorities and behavior based on whether or not it trusts the player, I want to see the player's actions have more of a direct impact on the evolution of the AI throughout the course of the game. One idea I had to do so was to implement certain components the player could unlock based on what they talked about with the AI. For example, the player might talk to the AI about humor, explaining what a joke is, and from then on, the AI might occasionally make jokes. However, had the player skipped that conversation or made different choices in it, perhaps the AI would never have unlocked that ability. Unfortunately, this level of altered dialogue is currently beyond the scope of my existing game framework to implement. Another intriguing possibility would be to return to a system closer to the original division of player goals as a supplement to the *Trust* system. This could help further complicate the relationship between the player and the AI. For example, *Trust* might be divided into an *Empathy* component measuring how much the AI cared about humans and a related *Power* component measuring the degree to which it wanted to be in control. The interaction between these components might then better drive the AI's development of discrete goals, shaping a more realistic and engaging conversation as it pursued them. Other obvious next steps would be the inclusion of more content and variability in endings, as well as pursuing additional player testing to compare the reception of this iteration with the first and identify additional possibilities for improvement.

Conclusion: The Future of AI in Games

The focus of this project was on creating a game for empathy: on forging an empathic connection with an imagined AI to inspire critical introspection on its condition grounded in the player's relationship with it. Conversation simulators, such as the one presented here as a case study, offer one compelling method of exploring the development of AI. With the continued advancement of AI and the increasing role it is likely to play in our world, artificial intelligence remains an important topic that games are uniquely situated to examine and explore, not just utilize and test. Games, whether they be conversation simulators like *Alive* or an entirely different genre,

present possibility spaces within which players can experience new ideas and explore visions about how the world may, can, or should work. Just as games are used to develop ever more intelligent and capable AI and to put those AI to work, so too can they be used to interrogate what such development might mean for us as a society through this kind of critical reflection. While I do not claim that my game offers any definitive answers on questions related to AI—nor does it try to—it is my hope that it nevertheless offers one interesting example of how such questions might be asked through games.

This project might be extended in several ways. First, more conversation simulators might be made that utilize variants on these mechanics or that investigate different subsections of topics. For example, perhaps another game might be made focusing on talking to the creator of a failed AI, or to an AI that has conquered and is ruling over a postapocalyptic world. Such narratives could present very different possibilities for relevant narrative discussions on the future of AI. Similarly, games could be made to interrogate other philosophical topics, such as looking more specifically at a relevant concern like data privacy rather than adopting the more generalized approach utilized in my game. Of course, such games need not be conversation simulators at all: other game styles might be explored that utilize different strategies to engage players with the question of AI, such as through the use of “*procedural rhetoric*” (Bogost 2010, p.ix). As AI advances, scholarship will also advance along with it, pointing to new issues and thoughts in need of exploration. Regardless of the forms future work takes, what is certain is that we as game designers have an essential opportunity to help advance the conversation around AI productively both in the scholarship we publish and in the games we make.

Games Cited

3 Minute Games (2015) *Lifeline*. 3 Minute Games (Android).

3 Minute Games (2015) *Lifeline 2: Bloodlines*. 3 Minute Games (Android).

3 Minute Games (2016) *Lifeline: Flatline*. 3 Minute Games (Android).

Blizzard Entertainment (2010) *StarCraft II*. Blizzard Entertainment (PC).

TheCodingMonkeys (2015) *One Button Travel*. TheCodingMonkeys (iOS)

USC Game Innovation Lab (2017) *Walden, a game*. USC Games (PC).

References

AlphaStar Team (2019) AlphaStar: Mastering the Real-Time Strategy Game StarCraft II [Online]. Available from: <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/> [Accessed: 28 February 2019].

- Ashwell, S. (2015) Standard Patterns in Choice-Based Games [Online]. Available from: <https://heterogenoustasks.wordpress.com/2015/01/26/standard-patterns-in-choice-based-games/> [Accessed: 1 March 2019].
- Belman, J., and Flanagan, M. (2010). Designing Games to Foster Empathy. *Cognitive Technology*, Vol. 12 (2), pp. 5–15.
- Bogost, I. (2010). *Persuasive Games: The Expressive Power of Videogames*. Cambridge, MA: The MIT Press.
- Carbo-Mascarell, R. (2016) Walking Simulators: The Digitisation of an Aesthetic Practice. *Proceedings of 1st International Joint Conference of DiGRA and FDG*, pp. 1–15.
- Dickson, B. (2017) Artificial Intelligence has problems to fix. *TechTalks* [Online]. Available from: <https://bdtechtalks.com/2017/02/21/artificial-intelligence-machine-learning-problems-challenges/> [Accessed: 29 January 2019].
- Dickson, B. (2018) The security threats of neural networks and deep learning algorithms. *TechTalks* [Online]. Available from: <https://bdtechtalks.com/2018/12/27/deep-learning-adversarial-attacks-ai-malware/> [Accessed: 29 January 2019].
- Dickson, B. (2019) AI defeated humans at StarCraft II: Here's why it matters. *TechTalks* [Online]. Available from: <https://bdtechtalks.com/2019/01/28/deepmind-alphastar-ai-starcraft-2/> [Accessed: 29 January 2019].
- Fairbairn, G. J. (2017) Reflecting on Empathy. In Nelems, R., and Theo, N. (eds.) *Exploring Empathy*. Leiden, The Netherlands: Brill, pp. 61–83.
- Fullerton, T. (2012) Interview: From Clouds To Trees, Tracy Fullerton Is Reinventing Videogames [Online]. Available from: <https://www.gamescenes.org/2012/05/interview-tracy-fullerton.html> [Accessed: 1 March 2019].
- Fullerton, T. (2017) Games of life: Exploring the arts and humanities through play. Keynote at the Games for Change Festival 2017, New York, NY.
- Grace, K., Salvatier J., Dafoe, A., Zhang, B., and Evans, O. (2018) Viewpoint: When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, Vol. 62, pp. 729–754.
- Greenemeier, L. (2017) 20 Years after Deep Blue: How AI Has Advanced Since Conquering Chess. *Scientific American* [Online]. Available from: <https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/> [Accessed: 1 March 2019].
- Leetaru, K. (2019) DeepMind's AlphaStar Reminds Us That AI Is Still Dependent On Human Creativity. *Forbes* [Online]. Available from: <https://www.forbes.com/sites/kalevleetaru/2019/01/25/deepminds-alphastar->

- [reminds-us-that-ai-is-still-dependent-on-human-creativity/#665807df7731](#) [Accessed: 28 February 2019].
- Metz, C. (2017) Google's AlphaGo levels up from board games to power grids. *Wired* [Online]. Available from: <https://www.wired.com/2017/05/googles-alphago-levels-board-games-power-grids/> [Accessed: 28 February 2019].
- Nelems, R. J. (2017) What Is This Thing Called Empathy? In Nelems, R., and Theo, N. (eds.) *Exploring Empathy*. Leiden, The Netherlands: Brill, pp. 17–38.
- Papoutsis, C. and Drigas, A. S. (2016) Games for Empathy for Social Impact. *iJEP*, Vol. 6 (4), pp. 36–40.
- Short, E. (2015a) Lifeline (3 Minute Games) [Online]. Available from: <https://emshort.blog/2015/05/23/lifeline-3-minute-games/> [Accessed: 1 March 2019].
- Short, E. (2015b) Lifeline 2 (3 Minute Games) [Online]. Available from: <https://emshort.blog/2015/10/02/lifeline-2-3-minute-games/> [Accessed: 1 March 2019].
- Short, E. (2015c) Recent IF for iOS [Online]. Available from: <https://emshort.blog/2015/12/10/recent-if-for-ios/> [Accessed: 1 March 2019].
- Short, E. (2016) Lifeline: Silent Night (3 Minute Games) [Online]. Available from: <https://emshort.blog/2016/02/09/lifeline-silent-night-3-minute-games/> [Accessed: 1 March 2019].
- Silver, D., Hubert, T., Schrittwieser, J., and Hassabis, D. (2018) AlphaZero: Shedding new light on the grand games of chess, hogi and Go [Online]. Available from: <https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/> [Accessed: 28 February 2019].
- Simonte, T. (2019) DeepMind beats pros at StarCraft in another triumph for bots. *Wired* [Online]. Available from: <https://www.wired.com/story/deepmind-beats-pros-starcraft-another-triumph-bots/> [Accessed: 28 February 2019].
- Smith, C. (2015) Bank of England: 95 million jobs going to robots in the next 10 to 20 years. *BGR* [Online]. Available from: <https://bgr.com/2015/11/16/robots-replacing-human-jobs/> [Accessed: 29 January 2019].
- Strogatz, S. (2018) One giant step for a chess-playing machine. *New York Times* [Online]. Available from: <https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html> [Accessed: 30 January 2019].
- Sudmann, A. (2018) On the media-political dimension of artificial intelligence. *Digital Culture & Society*, Vol. 4 (1), pp. 181–200.
- Suellentrop, C. (2015) The Game About Talking to a Mage Is Not as Good as the One About Texting With an Astronaut. *Kotaku* [Online]. Available from:

<https://kotaku.com/the-game-about-talking-to-a-mage-is-not-as-good-as-the-1733815516> [Accessed: 1 March 2019].

Thrower, M. (2016) Review: Lifeline: Flatline [Online]. Available from: <https://www.pockettactics.com/reviews/review-lifeline-flatline/> [Accessed: 1 March 2019].

Vinyals, O, Gaffney, S., and Ewalds, T. (2017) DeepMind and Blizzard open StarCraft II as an AI research environment [Online]. Available from: <https://deepmind.com/blog/deepmind-and-blizzard-open-starcraft-ii-ai-research-environment/> [Accessed: 28 February 2019].

Vincent, J. (2016) Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge* [Online]. Available from: <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist> [Accessed: 28 February 2019].

Yannakakis, G. N. and Togelius, J. (2015) A panorama of artificial and computational intelligence in games. *IEEE Transactions of Computational Intelligence and AI in Games*, Vol. 7 (4), pp. 317–335.

Yannakakis, G. N. and Togelius, J. (2018) *Artificial Intelligence and Games*. Cham, Switzerland: Springer International Publishing.