The FarmQuest Player Telemetry Dataset: Playthrough Data of a Cozy Farming Game

Kristen K. Yu, Matthew Guzdial, Nathan R. Sturtevant
Computing Science Department, University of Alberta
Alberta Machine Intelligence Institute
{kkyu, guzdial, nathanst}@ualberta.ca

Abstract

Open datasets are the foundation of many types of academic research. One field that requires datasets is data-driven player modeling, but there is a lack of variety in existing datasets. We introduce the FarmQuest Player Telemetry (FPT) dataset, a new playthrough dataset of a cozy farming game. We envision this dataset will be used for reproducing prior results, evaluating initial ideas, and evaluating the generalizability of new and existing algorithms.

Introduction

Open datasets are valuable for many types of research. Image classification is a classic example of a problem that has many datasets to facilitate research (Tian et al. 2020; Deng 2012). Data-driven player modeling is a similar field, where datasets are required to iterate and evaluate research projects. Finding a suitable dataset can sometimes be difficult, but open datasets uniquely offer ways to facilitate experimental reproducability and support future research.

Reproducability is a core principle of gaining knowledge and modern science (Popper 1959; Drummond 2009). This serves as a key way to verify the integrity of results published in a particular paper by allowing for other researchers to repeat the experiment. To do this, all key components must be made publicly available. To this end, we introduce the FarmQuest Player Telemetry (FPT) Dataset. In part to allow for reproducibility of our results in "Evaluating the Effects of AI Directors for Quest Selection" (Yu, Sturtevant, and Guzdial 2024), but also to support future research in data-driven player modeling.

To further support the reproducibility of our results, this dataset was built using publically available software. FarmQuest is a research testbed for procedural content generation and AI directors. All of the code for this testbed is publically available. FarmQuest is a coherent cozy game (Waszkiewicz and Bakun 2020), which we discuss more in the Background Section.

Beyond experimental transparency, the publication of this dataset has the potential to benefit other researchers interested in data-driven player modeling. Data-driven player modeling is contrasted with other forms of player modeling such as Bartle's taxonomy (Bartle 1996). A 2018 survey paper of data-driven player modeling approaches found forty-six different papers using datasets from thirty-four different games (Hooshyar, Yousefi, and Lim 2018). These datasets were predominantly acquired through two ways - the first was partnering with a video game studio and receiving their player data (Drachen, Canossa, and Yannakakis 2009; Harpstead et al. 2015), and the second was collecting the data themselves (Falakmasir et al. 2016; Zook et al. 2012). Researchers without ties to the industry often rely on the second method. These datasets are not always publicly available, which hinders the ability to reproduce results or to build upon prior research.

One key aspect of data-driven player modeling research is understanding the context in which the player model is effective. Is the player model intended to generalize to many types of games, or is it limited to only certain genres? One straightforward way to address these questions is to evaluate an algorithm on datasets from a variety of games. However, as we discuss in the Background section, most playthrough datasets only feature platformer games, where a playthrough dataset includes the telemetry of one or more players playing a game or games.

In our FPT dataset, thirty-two players played FarmQuest twice, resulting in sixty-four playthroughs with recorded telemetry events. Additionally, we collected survey data so the player's sentiments can be associated with their playthrough data. The FPT dataset can be found at the github repository https://github.com/kristenYu/FarmQuest-Player-Telemetry-Dataset.

We envision the FPT dataset will be used in three ways. The first is to reproduce and verify our previous experiments. The second is a starting point for researchers who are interested in player modeling for cozy games. The third is a test for generalizability of existing player modeling methods. We perform a few experiments to showcase the potential of uses two and three, specifically in the topics of churn prediction, player clustering, and next action prediction.

Background

In this section we cover the definition of cozy games, some relevant datasets, and data-driven player modeling techniques.

Cozy Games

In the original FarmQuest paper, we introduced the system as a platform to support research but did not categorize the genre of game. In order to fully characterize FarmQuest, we propose it is a coherent cozy farming game. Waszkiewicz and Bakun builds upon industry ideas of "cozy" as a descriptor for games (Short et al. 2018), and they maintain that a cozy game meets the criteria of "safety", "abundance" and "softness". Safety is defined as the absence of danger with no impending loss or threats. Abundance is defined as providing the space to work on higher level needs such as self expression, appreciation of beauty, or belonging. Softness refers to the visuals in a game, which convey to the player that nothing is pressing and players are in a lower state of arousal. Waszkiewicz and Bakun further defines cozy games into three categories: coherent, dissonant and situational. A coherent game is one where the cozy aesthetic matches the intended messaging of the game. Dissonant is where the cozy aesthetic juxtaposes the messaging of the game. Situational is where the game features cozy moments, but the game is not cozy in its entirety.

We propose that FarmQuest meets the criteria of safety, abundance, and softness. Farmquest has safety because there are no enemies or other danger, and it is not possible to die in the game. FarmQuest has abundance because the player is not forced to worry about tasks like eating or shelter, and instead can focus on higher level needs such as self expression. FarmQuest has softness because the visual aesthetics of the game feature low arousal such as minimal animations.

To determine whether FarmQuest is coherent, dissonant or situational, we need to consider the intended messaging of the game. The message of the game to be the enjoyment of running a farm. Thus, we consider FarmQuest to be a coherent cozy game because the coziness of the visuals match the underlying message of the game.

Relevant Game Datasets

Researchers have introduced a variety of datasets that focus on video game research. These datasets cover a range of topics which demonstrates the breadth of interest in the research community. Researchers have published datasets focusing on dialogue (Rennick and Roberts 2023; Juraska, Bowden, and Walker 2019), music (Cardoso, O. Moraes, and N. Ferreira 2024), recorded game playing footage (Barman et al. 2018), and development problems (Politowski et al. 2020).

Data-driven player modeling uses playthrough datasets, though there is a limited variety in the genre of games. The Mario PCG dataset, Gwario, and platformer experience dataset all include playthrough data of Super Mario Bros levels (Guzdial and Riedl 2021; Siu, Guzdial, and Riedl 2017; Karpouzis et al. 2015), which is a platformer game. The AGAIN Dataset includes playthrough data from three platformers, three racing games, and three first person shooter games (Melhart, Liapis, and Yannakakis 2022). The Powerwash Simulator Dataset includes playthrough data from the research edition of Power Wash Simulator, a simulation game (Vuorre et al. 2023). Our dataset is intended to supplement the current body of playthrough datasets with a new genre - a cozy game.

Data-driven Player Modeling

There are many approaches to data-driven player modeling. Neural networks (Pedersen, Togelius, and Yannakakis 2010), clustering (Anagnostou and Maragoudakis 2009), and case-based modeling (van den Herik, Donkers, and Spronck 2005) have all been used to create player models, among other methods.

Player modeling can be used to learn different things about the player. Some common approaches to target are player clustering, churn modeling and next action prediction. Player Clustering is when the players are grouped to learn about the types of players in the game (Drachen et al. 2014; Ramirez-Cano, Colton, and Baumgarten 2010). Churn prediction refers to predicting how long a player will play a certain game, or the likelihood that a player will return to play the same game again (Bertens, Guitart, and Periáñez 2017; Mustač et al. 2022). Next action prediction approximates the player's next action or goal (Asayama et al. 2015; Bisson, Larochelle, and Kabanza 2015).

FPT Dataset

The FPT dataset is a collection of survey data and playthrough data for the game FarmQuest (Yu, Guzdial, and Sturtevant 2022).

FarmQuest

FarmQuest features a single game loop, where players can plant vegetables, harvest mushrooms and berries, cook food, decorate their house, buy and sell things in a shop, and complete quests for money. The goal of the player is to pay off their mortgage for the farm. The player starts with 300 coins, and the goal is to earn 1000 total coins in order to pay off their mortgage. Players complete quests which gives 100 coins, or sell items which gives various amounts of coins.

There are four major types of gameplay in the game. The first major gameplay type is placing furniture, where players customize their house. The second type is planting, where players plant crops in order to use them in recipes. The third type is cooking, where players cook recipes, and the fourth is harvesting, where players can harvest berries and mushrooms that spawn every in-game day.

There is a day-night cycle in the game, which tracks the passage of time. Each day is thirty seconds, twilight is ten seconds, and nighttime is twenty seconds. In total, an in game day is one minute.

Playthrough Data

There are thirty different types of player telemetry events in this dataset. These events are grouped into classes, and each individual event is labeled with an event type and optional value. The timestamp of each event is also recorded, however the timestamp did not record correctly and should not be used. For example, one event has the class "Event", the type "SessionStart" and the meaning is "Player started the game". A full list of all event classes, and their corresponding event types and meanings are available on the github.

This dataset was collected as part of an experiment on AI Directors, or systems that dynamically change a game to suit an individual player. There were three AI Directors present in the experiment, which selected quests to present to the player. The first was random, which randomly assigns quests to show to the player. The second was PaSSAGE, which uses the player's previous actions to predict which quests a player might prefer (Thue et al. 2007). The third was a combinatorial multi-armed bandit (CMAB) algorithm, which uses which quests the player previously selected to inform which quests to present to the player (Yu et al. 2022). Each player experienced two of the AI directors, which resulted in each player having two playthroughs of the game.

Survey Data

The dataset includes short survey data, where each player completed a short survey after a playthrough, and comparison survey data, where players were asked to compare their experience. The short survey questions were a series of Likert questions. The comparison survey questions were a combination of Likert questions and short answer questions. Both the short survey and the comparison survey questions are available in "Evaluating the Effects of AI Directors for Quest Selection" (Yu, Sturtevant, and Guzdial 2024).

Experiments

In addition to reproducing our previous results, we illustrate some other potential uses of our dataset by performing experiments relevant to data-driven player modeling. We chose three problems commonly found in player modeling literature: player clustering, player churn modeling, and next action prediction (Hooshyar, Yousefi, and Lim 2018).

Player Clustering

We construct a player type discovery algorithm by using k-means clustering. To perform the player clustering, we created player vectors by counting the number of times that a player placed something, the number of times that a player planted something, the number of times a player cooked something, and the number of times a player harvested berries or mushrooms. These represent the four major gameplay mechanics within the game. The feature vector is ⟨*place*, *plant*, *cook*, *harvest*⟩.

Based on Kruskal-Wallis statistical testing, there is a difference between each session and each AI Directo. According to each player vector, we perform player clustering on six individual populations. We used the elbow method approach and a computational knee locator to determine the number of clusters. Players were sorted into either three or four major clusters depending on session and AI director. We performed principle cluster analysis (PCA) to reduce the space to two axes and graph the results of the clustering processes. Each 'x' marks the center of a cluster, and each cluster is color coded, shown in Figure 1.

In session 2 Passage, there are four distinct clusters. Two of these clusters are close to each other, while there are two that are farther away from the main body of clusters. We attribute these clusters to the four main action types, and we assume the larger clusters represent players who were interested in harvesting or placing furniture, as those were the most popular ways to complete a quest.

Sample	Predicted Value	Actual Value
1	10.24	16
2	8.71	9
3	3.37	4

Table 1: Churn Prediction using Linear Regression

Player Churn Prediction

We consider player churn modeling. Because the timestamp data is unusable, we use the in-game day as a proxy for time. The current day is logged at the start of each day, and each in-game day is one minute. Thus, we can approximate the time spent in game to the nearest minute.

We drew on all available features when modeling churn. To model the player, we tracked all of the interactions in the game. The feature vector tracks the number of times the player completes each interaction. The feature vector is

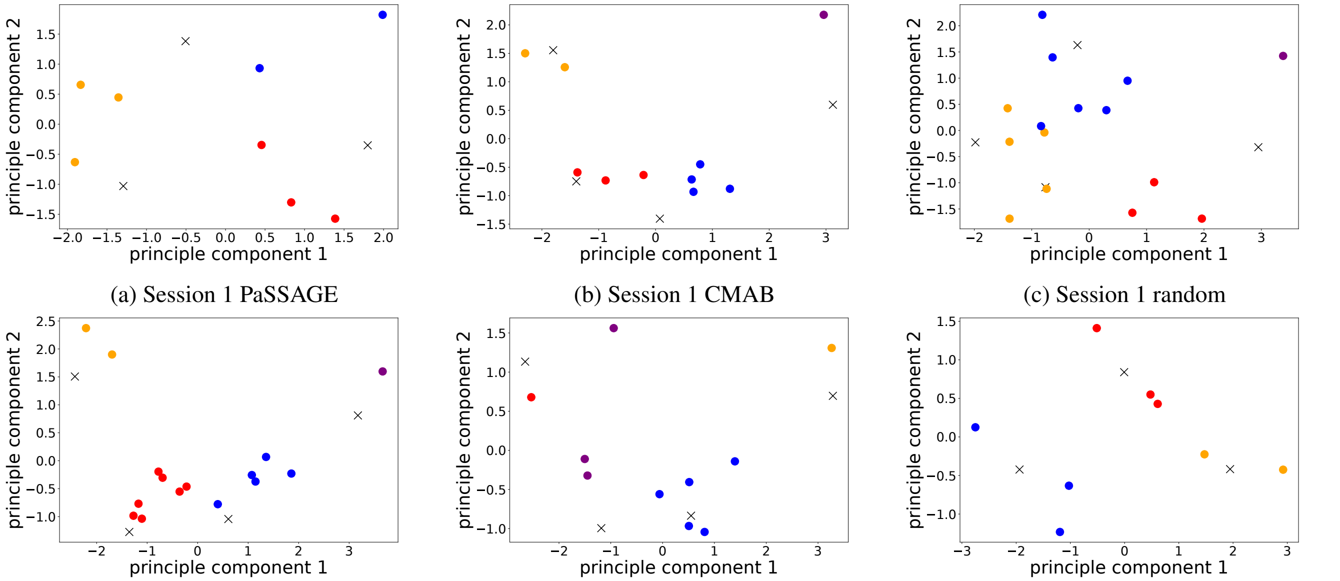
⟨ plant, cook, harvest berry or mushroom, place furniture,bought in shop, sold in shop, harvest crop, rotate furniture, pickup furniture, tried to buy something in shop, transitioned between scenes, payed mortgage ⟩.

We performed linear regression using this large feature vector to predict the day data. The R^2 value of the regression is 0.82, which shows linear approximation to be able to learn a reasonable model for this dataset. We randomly selected three samples from the the training feature set and predicted the number of days using the trained linear regression model, shown in Table 1. The maximum coefficient is 0.82 and the minimum coefficient is -0.18, where the maximum coefficient is attributed to paying the mortgage, which stops the game. The action that stops the game is the clearest indicator of the number of days spent in game, and thus has the highest coefficient.

Next Action Prediction

We consider predicting the next action. Figure 2 shows the overall frequency of actions taken in the game. To predict the next action, we used n-grams and selected the next action based on the highest probability. We implemented unigrams, bigrams, and trigams on the data. We built the trigram and bigram data from the occurrences within the dataset rather than creating the entire combinatorial space. This is due to the fact that there are illegal sequences of actions that cannot occur in game, to ensure that illegal sequences were not possible to be predicted. For example, the sequence "Transition:Shop", "Transition:Home" is not possible because it is not possible to enter the home from the shop.

Table 2 shows a random sampling of five trigrams and the results from the n-gram prediction. In one case there were different predicted next actions based on the type of n-gram used, but in most cases the next action was the same. For the first sample, we believe the variety is due to the fact that after placing an object, there was the most choice in what the player decided to do next. In comparison to the second sample, players seemingly mostly entered the home in order to cook if they had previously harvested a crop. This makes



(a) Session 1 PaSSAGE (b) Session 1 CMAB (c) Session 1 random

(d) Session 2 PaSSAGE (e) Session 2 CMAB (f) Session 2 random

Figure 1: K-Means Clustering Results

Sample	Trigram Next Action	Trigram Probability	Bigram Next Action	Bigram Probability	Unigram Next Action	Unigram Probability
Transition:Home;Place;Place	Place	0.45	Transition:Main	0.46	Pickup	0.50
Transition:Main;HarvestCrop;TransitionHome	Cook	0.57	Cook	0.67	Cook	0.57
Plant;Plant;Harvest	Harvest	0.91	Harvest	0.91	Harvest	0.75
Plant;Harvest;Transition:QuestBoard	Submit	1.00	Submit	0.79	Submit	0.53
Submit;Transition:Main;Harvest	Harvest	0.80	Harvest	0.85	Harvest	0.75

Table 2: Predicted next moves using trigrams, bigrams, and unigrams from a random sampling of trigrams

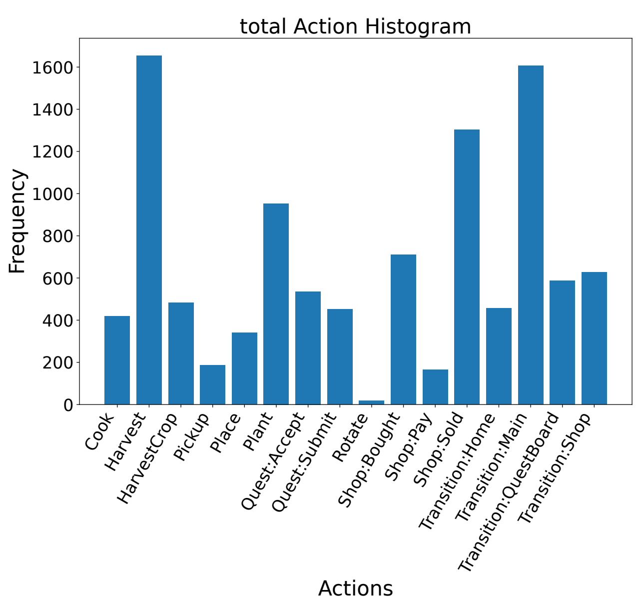


Figure 2: Histogram of the frequency of actions taken in game

sense within the gameplay loops, as a player needs to harvest crops to obtain ingredients to cook.

Conclusion

In this paper we presented the FPT dataset, a new playthrough dataset of a cozy farming game. We envision this dataset could be used for three main reasons. The first is to reproduce results from our experiments. The second is to support new data-driven player modeling algorithms for cozy games. The third is to expand the variety of genres of games present in existing playthrough datasets, so that existing player modeling algorithms can be evaluated for generalizability. To support uses two and three, we ran three relevant experiments for data-driven player modeling with the targets of player clustering, churn prediction, and next action prediction.

Acknowledgements

This work was funded by the Canada CIFAR AI Chairs Program, Alberta Machine Intelligence Institute, and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Anagnostou, K.; and Maragoudakis, M. 2009. Data Mining for Player Modeling in Videogames. In *2009 13th Panhellenic Conference on Informatics*, 30–34.

Asayama, K.; Moriyama, K.; Fukui, K.-i.; and Numao, M. 2015. Prediction as faster perception in a real-time fighting video game. In *2015 IEEE conference on computational intelligence and games (CIG)*, 517–522. IEEE.

Barman, N.; Zadtootaghaj, S.; Schmidt, S.; Martini, M. G.; and Möller, S. 2018. GamingVideoSET: A Dataset for Gaming Video Streaming Applications. In *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, 1–6.

Bartle, R. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research*, 1(1): 19.

Bertens, P.; Guitart, A.; and Periáñez, Á. 2017. Games and big data: A scalable multi-dimensional churn prediction model. In *2017 IEEE conference on computational intelligence and games (CIG)*, 33–36. IEEE.

Bisson, F.; Larochelle, H.; and Kabanza, F. 2015. Using a recursive neural network to learn an agent's decision model for plan recognition. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Cardoso, I.; O. Moraes, R.; and N. Ferreira, L. 2024. The NES Video-Music Database: A Dataset of Symbolic Video Game Music Paired with Gameplay Videos. In *Proceedings of the 19th International Conference on the Foundations of Digital Games*, FDG 2024. New York, NY, USA: Association for Computing Machinery. ISBN 9798400709555.

Deng, L. 2012. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine*, 29(6): 141–142.

Drachen, A.; Canossa, A.; and Yannakakis, G. N. 2009. Player modeling using self-organization in Tomb Raider: Underworld. In *2009 IEEE Symposium on Computational Intelligence and Games*, 1–8.

Drachen, A.; Thurau, C.; Sifa, R.; and Bauckhage, C. 2014. A Comparison of Methods for Player Clustering via Behavioral Telemetry. *Proceedings of the 8th International Conference on the Foundations of Digital Games*.

Drummond, C. 2009. Replicability is not reproducibility: nor is it good science. In *Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML*, volume 1. National Research Council of Canada Montreal, Canada.

Falakmasir, M. H.; Gonzalez-Brenes, J. P.; Gordon, G. J.; and DiCerbo, K. E. 2016. A Data-Driven Approach for Inferring Student Proficiency from Game Activity Logs. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, 341–349. New York, NY, USA: Association for Computing Machinery. ISBN 9781450337267.

Guzdial, M.; and Riedl, M. 2021. Game Level Generation from Gameplay Videos. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 12(1): 44–50.

Harpstead, E.; Zimmermann, T.; Nagapan, N.; Guajardo, J. J.; Cooper, R.; Solberg, T.; and Greenawalt, D. 2015. What drives people: Creating engagement profiles of players from game log data. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 369–379.

Hooshyar, D.; Yousefi, M.; and Lim, H. 2018. Data-Driven Approaches to Game Player Modeling: A Systematic Literature Review. *ACM Computing Surveys*, 50(6): 1–19.

Juraska, J.; Bowden, K.; and Walker, M. 2019. ViGGO: A Video Game Corpus for Data-To-Text Generation in Open-Domain Conversation. In *Proceedings of the 12th International Conference on Natural Language Generation*, 164–172.

Karpouzis, K.; Yannakakis, G.; Shaker, N.; and Asteriadis, S. 2015. The Platformer Experience Dataset. In *Sixth Affective Computing and Intelligent Interaction (ACII) Conference*.

Melhart, D.; Liapis, A.; and Yannakakis, G. N. 2022. The Arousal Video Game AnnotatIoN (AGAIN) Dataset. *IEEE Transactions on Affective Computing*, 13(4): 2171–2184.

Mustač, K.; Bačić, K.; Skorin-Kapov, L.; and Sužnjević, M. 2022. Predicting player churn of a Free-to-Play mobile video game using supervised machine learning. *Applied Sciences*, 12(6): 2795.

Pedersen, C.; Togelius, J.; and Yannakakis, G. N. 2010. Modeling player experience for content creation. *IEEE Transactions on Computational Intelligence and AI in Games*, 2(1): 54–67.

Politowski, C.; Petrillo, F.; Ullmann, G. C.; de Andrade Werly, J.; and Guéhéneuc, Y.-G. 2020. Dataset of Video Game Development Problems. In *Proceedings of the 17th International Conference on Mining Software Repositories*, MSR '20, 553–557. New York, NY, USA: Association for Computing Machinery. ISBN 9781450375177.

Popper, K. 1959. *The logic of scientific discovery*. Hutchinson.

Ramirez-Cano, D.; Colton, S.; and Baumgarten, R. 2010. Player classification using a meta-clustering approach. In *Proceedings of the 3rd Annual International Conference Computer Games, Multimedia & Allied Technology*, 297–304. Citeseer.

Rennick, S.; and Roberts, S. G. 2023. The Video Game Dialogue Corpus.

Short, T. X.; Ordon, A.; Hurd, D.; Howe, C.; Forbes, J.; Eiserloh, S.; Diaz, J.; and Cook, D. 2018. Cozy Games.

Siu, K.; Guzdial, M.; and Riedl, M. O. 2017. Evaluating singleplayer and multiplayer in human computation games. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, 1–10. Hyannis Massachusetts: ACM. ISBN 978-1-4503-5319-9.

Thue, D.; Bulitko, V.; Spetch, M.; and Wasylishen, E. 2007. Interactive Storytelling: A Player Modelling Approach. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 3(1): 43–48.

Tian, Y.; Suzuki, C.; Clanuwat, T.; Bober-Irizar, M.; Lamb, A.; and Kitamoto, A. 2020. Kaokore: A pre-modern japanese art facial expression dataset. *arXiv preprint arXiv:2002.08595*.

van den Herik, H.; Donkers, H.; and Spronck, P. 2005. Opponent modelling and commercial games. In Kendall, G.; and Lucas, S., eds., *Proceedings of the IEEE 2005 symposium on computational intelligence and games*, 15–25. Essex University. Pagination: 11.

Vuorre, M.; Magnusson, K.; Johannes, N.; Butlin, J.; and Przybylski, A. 2023. An intensive longitudinal dataset of in-game player behaviour and well-being in PowerWash Simulator. *Scientific Data*, 10.

Waszkiewicz, A.; and Bakun, M. 2020. Towards the aesthetics of cozy video games. *Journal of Gaming an Virtual Worlds*, 12(3): 225–240.

Yu, K.; Sturtevant, N. R.; and Guzdial, M. 2024. Evaluating the Effects of AI Directors for Quest Selection. In *Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*.

Yu, K. K.; Guzdial, M.; and Sturtevant, N. R. 2022. PWR: A Demonstration of an AI Director Video Game Test Bed. In *Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*.

Yu, K. K.; Guzdial, M.; Sturtevant, N. R.; Cselinacz, M.; Corfe, C.; Lyall, I. H.; and Smith, C. 2022. Adventures of AI Directors Early in the Development of Nightingale. In *Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*.

Zook, A.; Lee-Urban, S.; Drinkwater, M. R.; and Riedl, M. O. 2012. Skill-based Mission Generation: A Data-driven Temporal Player Modeling Approach. In *Proceedings of the The Third Workshop on Procedural Content Generation in Games*, PCG'12, 1–8. New York, NY, USA: Association for Computing Machinery. ISBN 9781450314473.