

Automatic Playtesting and Quality Assurance of Games

Andreas Jonasson

Andreas.Jonasson@mau.se

Faculty of Technology and Society, Malmö University

Malmö, Sweden

ABSTRACT

In this part of my application, I present my PhD focus, conducted within an Interreg project, Game Tech Academy, with multiple partners in Denmark and Sweden. The aim of the research is the theory creation and development of a tool for automatic play testing, aimed at helping and empowering indie game developers, hobbyists, and small-scale studios in their game development journeys. As such, I have chosen to focus on a combination of design science research and collaborative research – which I also hope can become a way to bridge the industry-academia divide that can make it hard to conduct research with the industry. I present an abbreviated version of preliminary findings related to tools in game development within the communities existing in Skövde and Malmö. Lastly, I briefly discuss how I hope that the Game Tech Academy project can be the start of further collaborations.

CCS CONCEPTS

• **Software and its engineering** → *Interactive games; Application specific development environments; Software testing and debugging;*
• **Human-centered computing** → *User models;* • **Applied computing** → *Computer games.*

KEYWORDS

Automatic Gameplay Testing, Automatic Quality Assurance, Explainable AI, Collaborative Research

ACM Reference Format:

Andreas Jonasson. 2025. Automatic Playtesting and Quality Assurance of Games. In . ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Game Tech Academy (GTA) is an EU backed Interreg project, co-funded by the European Union, for the Öresund-Kattegat-Skagerrak region [1]. It is a collaborative project between departments within *Aalborg Municipality (AAM)*, *Aalborg University (AAU)*, *Malmö University (MAU)*, *Science Park Skövde (SPS)*, *University of Skövde (HiS)*, and *Dania Academy (Dania)*. The project aims to not only look at game technology and what areas outside of entertainment they can be used within, but also to strengthen research and education in games development across the partners. This strengthening is done through the development of a tool for gameplay testing. The aim is for this tool to be developed in collaboration with small scale companies, indie companies, and hobbyists who may not be able to

conduct testing at scale. As the sole PhD student within this project, the research and development of this tool will be my focus. For the development part, I will work together with a research engineer at HiS.

The goals¹ of the tool is to be *usable by non-technical developers, easily integrated, and align with the needs of specific testing solutions or games*. Additional goals for the tools is that it should work with different *player personas*, transparent and understandable in line with what is expected out of *explainable AI (XAI)* and *informed by AAA-practices and academia*. All of this in an attempt to democratize the game development process of gameplay testing and make it more accessible for small scale developers, indie-developers, hobbyists, and students.

The rest of this submission for the Doctoral Consortium at FDG'25 will be used to clarify the above topics, discuss how I propose to not only develop a tool for a complex task like this, but also how I hope that this research project can help bridge the academia-industry collaboration problems that I commonly see in the field – a personal goal for this PhD.

1.1 Gameplay testing

Gameplay testing can be divided into two different activities; testing as part of *quality assurance* or testing as part of *player experience testing* [11, 12]. In the former activity, the testing is of a more systematic kind where protocols are followed to make sure that different levels, mechanics and/or quests in the game works as intended [11, 12]. If something does not work as intended a bug report is carefully filed wherein the bug is described, the actions leading to the bug are accounted for, and a severity rating is given [11, 12]. In some systems, the bug can be categorized and automatically assigned to someone for handling. When it comes to the other part of play testing, with a focus on the *player experience testing*, this is generally done with “real” players, the target group of the game [11, 12]. How these test are conducted depends on where in the development process it is conducted, but it can be everything from a Friday afternoon in the office lunge to a full-scale public beta weekend where the game is distributed to thousands of players around the globe [11]. Regardless of methodology used, the aim of this test is to gauge the players’ experience with the game and, often, to collect data on their interactions with the game [11, 12]. However, these tests will often also shed light on bugs that either have not been handled yet or were not found by the quality assurance driven play testing conducted [11, 12].

Gameplay testing has been identified as a possibly important factor for a game’s success[11, 20, 21], yet is still commonly underutilized and done to late in the development process [14]. Even so, gameplay testing – together with quality assurance over all – is of

¹as defined by the GTA project description and early discussions regarding the scope of the tool

Conference'17, July 2017, Washington, DC, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in , <https://doi.org/10.1145/nnnnnnn.nnnnnnn>.

seemingly great importance to the AAA-industry. Research papers from R&D departments within the industry are often related to testing and quality assurance, see for instance [3, 25, 26], and it is not uncommon that this is a topic at GDC, see for instance [6, 10, 13]. From personal discussions with industry representatives in Skövde, Sweden, it seems like the local indie scene are not able to conduct play testing at scale or as often as recommended. Furthermore, they seem to undervalue play testing – claiming that it is “unnecessary”, “do not return investment”, and that “we already know what we must work on forward”. If this is generalisable to other indie or hobbyist development communities are currently uncertain, but there are indicators that indie game studios might not look at play testing as prioritized when developing before presenting for funding or publicity [20].

1.2 Player Personas

Player personas are models designed after some “ideal user” of a game [7]. These models can be used to look at how different players could traverse the world, interact with the world and the mechanics within, as well as what order they progress through the game in. Different personas can then encompass different ideas, play-styles and “clusters” of the target audience. Once play testing is conducted with representative players, the data collected can be used to update and verify these personas to more closely align with the players of the game. While the original approach were to hand-tune these personas [7], later work has looked into the usage of evolutionary or machine learning algorithms to automatically create personas [15, 16, 22].

1.3 Explainable AI

Explainable AI is an attempt to make AI models and systems more transparent, understandable, and interpretable [2]. It can be seen as attempt to change the common “black box” approach of AI, especially different kind of machine learning models, to more transparent ones. Adadi and Berrada identifies four reasons for the importance of explainability: Explain to Justify; Explain to Control; Explain to Improve; and Explain to Discover [2]. Explain to Justify is about justifying a certain outcome, giving the user the opportunity to not only make sure that there are no errors leading up to the outcome but also to understand how a certain conclusion was drawn. Explain to Control focuses more on preventing errors during development, with a focus on debugging and finding flaws. Explain to Improve is closely related to control, but builds on the idea that a system that is understood is also easier to develop. Lastly, Explain to Discover focuses on the ability to use the system to get new knowledge, either through gathering information through the system or by learning new facts discovered and presented by the system [2].

1.4 The Tool

The aim of the research is the theory creation and development of a tool for automatic play testing and quality assurance, aimed at helping and empowering indie game developers, hobbyists, and small-scale studios in their game development journeys. At the time of writing this text, little work has been conducted towards the actual production of the tool. The next step should perhaps

be to combine knowledge and theories found within the above defined areas, develop a small game to use as proof-of-concept, before turning to professionals within the industry and discuss the usability of the tool. However, I do not wish for the result of my PhD to become yet another case of research and development done within an ivory tower, neither connected to nor usable by those I aim to empower with it. As such, I need to not only understand the problems and challenges faced by those I aim to empower when it comes to gameplay testing, but I must also connect it to research conducted both within academia and the industry. Therefore, I think it is time to look towards the proposed research of my PhD.

2 PROPOSED RESEARCH

The GTA project description that guides my PhD contains an important but easily overlooked text which defines that it is important to “verify that the developed tools align with the needs of specific testing solutions or games.”. This is to be done through punctual contact with startups, to provide them with opportunities to define their testing needs. This has made me not only deciding to conduct research according to *design science research* paradigms, but also to look into *collaborative research* to more closely incorporate the actual needs of companies that may not have the know-how or resources for play testing. Through this section, I will go through these concepts and, lastly, present some challenges and opportunities identified with this approach.

2.1 Design Science Research

Design science research (DSR) is “...the scientific study and creation of artefacts as they are developed and used by people with the goal of solving practical problems of general interest” [17]. Framing the entirety of the doctoral project as an iterative DSR project – with smaller, more focused DSR projects informing the overarching project – will hopefully allow me to conduct research in a methodologically sound way, although the project’s goal is of a more practical nature. DSR can be seen as five steps, or activities: Explicate problem; Define requirements; Design and develop artefact; Demonstrate artefact; Evaluate artefact [17]. In short, explicate problem is to formulate and justify the problem for which the artefact is to be developed; Define requirements divides the problem into different requirements, both functional, structure, and context, the artefact is to be used in; Design and develop artefact is, as the name suggests, the design and development of the artefact; Demonstrate artefact is to perform a “proof of concept” test to show that the artefact is able to solve the problem proposed; and lastly evaluate artefact is to determine how well the developed artefact solved the problem [17].

While a single DSR project using this methodology do not need to contain all steps, I plan for my overarching PhD research to not only contain all five, but also to iterate on them as the tool is developed, tested, and recontextualised. I do, however, not wish to do this research in a way that perhaps would traditionally be done. Far too often I see research conducted within the field of games research done through surveys, interviews, post-mortem analysis, or other methodologies where the researcher is not an active participant of the development process. Instead, I would want to conduct this research with games industry companies and representatives –

especially from the Swedish and Danish indie-game development scene – as active collaboration partners. This would not only allow for a DSR project that is well-informed through all five steps, but if it is successful it would also bring forth a possibility to create new ways of working with the industry. This, I hope, can be done through the usage of *Collaborative Research*.

2.2 Collaborative Research

Collaborative research (CR) is in and on itself not necessarily a single methodology [9, 18, 19]. However, within this short presentation, I will talk about it as a singularity and give a broad presentation of this way of conducting research and why I think it is worth pursuing through my PhD. My reason for looking towards CR is in and on itself a simple one; it is a paradigm for conducting research with stakeholders from both academia and non-academia with the end goal of co-producing knowledge that is transferable to stakeholders regardless of affiliation [9, 19]. While not all research within this paradigm will produce generalisable results [9, 18], a desirable outcome from a CR project is the learning process itself [24] – leading to tacit knowledge development for both academia and non-academia stakeholders.

Another important part of CR, and why I think it is of great usage for me, is that the joint research endeavour needs to be grounded inside the context of the non-academic stakeholder(s) [9, 18, 19]. This does not only open for the possibility of the conducted research being of direct applicability or relevance for the non-academic stakeholder, but it may also work as an incitement for different non-academic stakeholders to join into the research itself.

2.3 Challenges and Opportunities

One thing that DSR, CR, and game development have in common is the iterative process. While academia and industry works in different speed cycles [12, 23], working iteratively and in projects together with non-academic stakeholders could prove beneficial to attempt collapsing these speed cycles. As neither DSR nor CR focuses on specific methodologies [9, 17] and both these methodologies focuses on refining both understanding of the problem, context and processes wherein they are used [9, 17, 19], it is my hope that they will be fitting processes for the quick turns of game development [12, 23]. In the best of worlds, the development of the tool would fit into the development circle of the companies involved in the project in such a way that it becomes usable “here and now”. Of course, collaborating with companies are in and on itself a problem that must be handled. Many game development companies are already operating on tight schedules and with little room for extra work [12], and this is seemingly extra true for indie game companies [20]. This, in turn, can make it hard to start the work of building trust in that the research conducted with the industry is worth the time of the companies.

Another challenge that can come to light during a CR project is that of language. It is not uncommon that academia uses a language that can be hard to comprehend for those outside of academia [9, 18]. In fact, this is something that has also been presented as problematic within the field of games development – to the point of the research conducted becomes inaccessible to the developers [23]. However, as the field of CR has worked with this problem for decades [9],

it seemingly exists multiple strategies to pick and chose from to handle the problem.

One last challenge to bring up here is the case of Non-disclosure agreements (NDA) [12, 27, 28]. These makes it hard to disseminate findings and can make it downright impossible for the researcher to report any findings. If a NDA is not signed, it is easy to see how the researcher can become excluded from discussions throughout the project. However, by agreeing to an NDA, the researcher’s objectivity can suddenly be questioned. Can these researchers be neutral, objective, or, for that matter, a voice of reason? Does it truly matter? As NDA are often used for secrecy [12], these can also be used to silence any problematic findings from being reported.

3 CURRENT RESULTS

At the time of writing, I do not have any published results for my research. I have conducted a survey regarding AI tools for game development aimed at the local cluster of game developers in Skövde², with the primary finding that the discourse of “AI tool” today is owned almost exclusively by generative AI. Interestingly enough, multiple answers hinted at an understanding of what an AI tool is, yet a failure to identify AI tools that are currently in use.

I have also just finished two courses at doctoral level, one arranged by the Council of Swedish Games Researchers and one arranged by Malmö University. In both these courses, my focus became to position myself within the field of games research by highlighting what I find problematic – the lack of understanding for the development processes within the industry – and what I believe could be a partial solution to develop this knowledge; the aforementioned collaborative research.

Lastly, I have started (continued) networking within the industry. Through previous work at the University of Skövde, I had the good fortune to come in contact with both developers and students from the indie game development scene of Skövde, some of whom I suspect will be interested in participating in the research I conduct going forward. Attending two conferences during the fall of 2024, Sweden Game Conference [5] in Skövde and AI and Games Conference [8] in London, where I met and connected with both developers and researchers working, and researching, in both the industry and within academia. I hope to see many of them at this year’s Artificial Intelligence and Games Summer School [4] which will be held in Malmö, Sweden, this summer where I will be an active participant.

4 PLANS FOR COMPLETION

Due to the early stage of my PhD, even seeing that it would ever come to completion is hard. I know that one day, not many years for now, it will be done and I will stand there with my dissertation; but the road there is windy and uncertain. What I do know is that for my current plan to bear fruit, I will need to find collaborators that believe in me. As the GTA project is conducted in collaboration with partners closely connected to the games-industry in three different cities (Aalborg, Malmö, and Skövde), I hope that I through them can find these collaborators.

²Skövde, Sweden, is known for its indie game development scene. With many small companies in the local incubator (SPS), the University of Skövde with six different game development programs at bachelor level and three at higher level, it has been the spawning ground for games like Goat Simulator, V Rising, Valheim, and Satisfactory.

Then it is just the problem of solving automatic gameplay testing and create an easily integrated tool for non-technical developers that can play their games, but who's keeping score?

ACKNOWLEDGMENTS

A big thanks to my supervisors, José María Font Fernández, Alberto Enrique Alvarez Uribe, and Raquel Breejon Robinson for their feedback and support during the writing of this application. An even bigger thank you to my wife for her support through this journey; without her, nothing of this would be possible.

REFERENCES

- [1] Game Tech Academy. 2024. Game Tech Academy. <https://gametechacademy.dk/>.
- [2] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6 (2018), 52138–52160.
- [3] William Ahlberg, Alessandro Sestini, Konrad Tollmar, and Linus Gisslén. 2023. Generating Personas for Games with Multimodal Adversarial Imitation Learning. In *2023 IEEE Conference on Games (CoG)*. IEEE, IEEE, 1–8.
- [4] AI and Games Summer School. 2024. Artificial Intelligence and Games. <https://school.gameaibook.org/>.
- [5] Sweden Game Arena. 2024. Sweden Game Conference. <https://swedengameconference.se/en/>.
- [6] Igor Borovikov. 2020. Imitation Learning: Building Practical Agents to Test and Explore a First-Person Shooter. <https://gdcvault.com/play/1026838/Imitation-Learning-Building-Practical-Agents>.
- [7] Alessandro Canossa and Anders Drachen. 2009. Patterns of play: Play-personas in user-centred game development. *Breaking New Ground: Innovation in Games, Play, Practice and Theory - Proceedings of DiGRA 2009* 4 (2009). <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84873361599&partnerID=40&md5=8ac7f79742a354a8fa06f84c35ed34cb>
- [8] Game AI Events CIC. 2024. AI And Games Conference. <https://www.aiandgamesconference.com/>.
- [9] Flora Cornish, Nancy Breton, Ulises Moreno-Tabarez, Jenna Delgado, Mohi Rua, Ama de-Graft Aikins, and Darrin Hodgetts. 2023. Participatory Action Research. *Nature Reviews Methods Primers* 3, 1 (2023), 1–14. <https://doi.org/10.1038/s43586-023-00214-1>
- [10] Olivier Delalleau. 2019. ML Tutorial Day: Smart Bots for Better Games: Reinforcement Learning in Production. <https://gdcvault.com/play/1025891/ML-Tutorial-Day-Smart-Bots>.
- [11] Anders Drachen, Pejman Mirza-Babaei, and Lennart E Nacke. 2018. *Games user research*. Oxford University Press.
- [12] Henrik Engström. 2020. *Game Development Research* (1 ed.). University of Skövde, Högskölevägen 1, 541 28 Skövde. <https://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-19248> This book was created within the Game Hub Scandinavia 2.0 project. Project id: NYP520201849, EU Interreg Öresund-Kattegat-Skagerrak..
- [13] Jonas Gillberg. 2019. AI for Testing: The Development of Bots that Play 'Battlefield V'. <https://gdcvault.com/play/1025905/AI-for-Testing-The-Development>.
- [14] Christoffer Holmgård. 2024. Shifting Games QA Left: Building Better Production Pipelines With Rob Sandberg.
- [15] Christoffer Holmgård, Michael Cerny Green, Antonios Liapis, and Julian Togelius. 2018. Automated playtesting with procedural personas through MCTS with evolved heuristics. *IEEE Transactions on Games* 11, 4 (2018), 352–362.
- [16] Christoffer Holmgård, Antonios Liapis, Julian Togelius, and Georgios N Yannakakis. 2014. Evolving personas for player decision modeling. In *2014 IEEE Conference on Computational Intelligence and Games*. IEEE, IEEE, 1–8.
- [17] Paul Johannesson and Erik Perjons. 2014. *An introduction to design science*. Vol. 10. Springer.
- [18] Alfred Kieser and Lars Leiner. 2012. Collaborate With Practitioners: But Beware of Collaborative Research. *Journal of Management Inquiry* 21, 1 (2012), 14–28. <https://doi.org/10.1177/1056492611411923>
- [19] Angela McCabe, Rachel Parker, Tom Osegowitsch, and Stephen Cox. 2023. Overcoming Barriers to Knowledge Co-Production in Academic-Practitioner Research Collaboration. *European Management Journal* 41, 2 (April 2023), 212–222. <https://doi.org/10.1016/j.emj.2021.11.009>
- [20] Pejman Mirza-Babaei, Naeem Moosajee, and Brandon Drenikow. 2016. Playtesting for Indie Studios. In *Proceedings of the 20th International Academic Mindtrek Conference* (New York, NY, USA, 2016-10-17) (*AcademicMindtrek '16*). Association for Computing Machinery, 366–374. <https://doi.org/10.1145/2994310.2994364>
- [21] Pejman Mirza-Babaei, Samantha Stahlke, Günter Wallner, and Atiya Nova. 2020. A Postmortem on Playtesting: Exploring the Impact of Playtesting on the Critical Reception of Video Games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2020-04-21). ACM, 1–12. <https://doi.org/10.1145/3313831.3376831>
- [22] Luvneesh Mugrai, Fernando Silva, Christoffer Holmgård, and Julian Togelius. 2019. Automated playtesting of matching tile games. In *2019 IEEE Conference on Games (CoG)*. IEEE, IEEE, 1–7.
- [23] Marcello Passarelli, Jeffrey Earp, Francesca Maria Dagnino, Flavio Manganello, Donatella Persico, Francesca Pozzi, Thomas Buijtenweg, Mata Haggis, Chris Bailey, and Carlo Perrotta. 2020. The Distant Horizon: Investigating the Relationship between Social Sciences Academic Research and Game Development. *Entertainment Computing* 34 (2020), 100339. <https://doi.org/10.1016/j.entcom.2020.100339>
- [24] Robert Earl Patterson, Byron J. Pierce, Herbert H. Bell, and Gary Klein. 2010. Implicit Learning, Tacit Knowledge, Expertise Development, and Naturalistic Decision Making. *Journal of Cognitive Engineering and Decision Making* 4, 4 (2010), 289–303. <https://doi.org/10.1177/155534341000400403>
- [25] Julien Roy, Roger Girgis, Joshua Romoff, Pierre-Luc Bacon, and Christopher J. Pal. 2022. Direct Behavior Specification via Constrained Reinforcement Learning. In *39th International Conference on Machine Learning (ICML 2022) (Proceedings of Machine Learning Research (PMLR))*. Jmlr-Learning Machine Learning Research, 18828–18843.
- [26] Alessandro Sestini, Linus Gisslén, Joakim Bergdahl, Konrad Tollmar, and Andrew David Bagdanov. 2022. Automated gameplay testing and validation with curiosity-conditioned proximal trajectories. *IEEE Transactions on Games* 16, 1 (2022), 113–126.
- [27] Leon Y. Xiao. 2023. Debate: Academics Should Collaborate with the Technology Industry, but Not in Lieu of Noncollaborative Research. *Child and Adolescent Mental Health* 28, 2 (2023), 338–340. <https://doi.org/10.1111/camh.12644>
- [28] David Zendle and Heather Wardle. 2023. Debate: We Need Data Infrastructure as Well as Data Sharing – Conflicts of Interest in Video Game Research. *Child and Adolescent Mental Health* 28, 1 (2023), 155–157. <https://doi.org/10.1111/camh.12629>