Breaking Barriers: Automating Real-Time Solutions for Inclusive Esports Broadcasting

***Khizer Khan***

**FINAL YEAR PROJECT**

**COMP60011**

*k025079m*

*Supervisor: Joshua Lindsey*

**

2nd May 2025

# *Table Of Contents*

Table of Contents

[Table Of Contents 2](#__RefHeading___Toc5512_3304059719)

[List Of Figures 3](#__RefHeading___Toc5514_3304059719)

[Abstract 4](#__RefHeading___Toc5516_3304059719)

[Acknowledgements 5](#__RefHeading___Toc5518_3304059719)

[1. Introduction 6](#__RefHeading___Toc5520_3304059719)

[1.1. Research Background 6](#__RefHeading___Toc5522_3304059719)

[1.2. Aim 6](#__RefHeading___Toc5524_3304059719)

[1.3. Objectives 6](#__RefHeading___Toc5526_3304059719)

[1.4. Deliverables 7](#__RefHeading___Toc5528_3304059719)

[1.5. Structure of the Report 7](#__RefHeading___Toc5530_3304059719)

[2. Research Methods 8](#__RefHeading___Toc5532_3304059719)

[2.1. Research Methodologies 8](#__RefHeading___Toc5534_3304059719)

[2.1.1. Research Philosophy 8](#__RefHeading___Toc5536_3304059719)

[2.1.2. Research Approach 8](#__RefHeading___Toc5538_3304059719)

[2.1.3. Research Strategy 8](#__RefHeading___Toc5540_3304059719)

[2.1.4. Methodological Choices 8](#__RefHeading___Toc5542_3304059719)

[2.2. Research Methods 9](#__RefHeading___Toc5544_3304059719)

[2.3. Ethical Considerations 10](#__RefHeading___Toc5546_3304059719)

[3. Literature Review 11](#__RefHeading___Toc5548_3304059719)

[3.1. Introduction 11](#__RefHeading___Toc5550_3304059719)

[3.2. Analyses 11](#__RefHeading___Toc5552_3304059719)

[3.2.1. Theme 1: Esports 11](#__RefHeading___Toc5554_3304059719)

[3.2.2. Theme 2: accessibility in broadcast 13](#__RefHeading___Toc5556_3304059719)

[3.2.3. Gap Analysis: Under explored areas. 15](#__RefHeading___Toc5558_3304059719)

[3.3. Conclusion and Future Research 16](#__RefHeading___Toc5560_3304059719)

[4. Design and Implementation 19](#__RefHeading___Toc5562_3304059719)

[5.0 Testing and Validation of the Artefact 22](#__RefHeading___Toc5564_3304059719)

[6. Conclusion 26](#__RefHeading___Toc5566_3304059719)

[6.1. Key Findings & Contribution to Objectives 26](#__RefHeading___Toc5568_3304059719)

[6.1.1. Accessibility Barriers Identified (Objective 1) 26](#__RefHeading___Toc5570_3304059719)

[6.1.2. Solutions for Disabled or New Audiences (Objective 2,3&4) 27](#__RefHeading___Toc5572_3304059719)

[6.1.3. Prototype Validation (Objective 5) 28](#__RefHeading___Toc5574_3304059719)

[6.2. Implications for Esports Accessibility 28](#__RefHeading___Toc5576_3304059719)

[6.3. Limitations and Future Work 28](#__RefHeading___Toc5578_3304059719)

[6.4 Final Statement 29](#__RefHeading___Toc5580_3304059719)

[7. Refrence List 30](#__RefHeading___Toc5582_3304059719)

[Appendix A: Ethics Form 33](#__RefHeading___Toc5584_3304059719)

[Research Ethics 33](#__RefHeading___Toc5586_3304059719)

[Disclaimer Form 33](#__RefHeading___Toc5588_3304059719)

[Appendix B : Health and Safety Risk Assessment 36](#__RefHeading___Toc5590_3304059719)

# *List Of Figures*

Table of Figures

[Figure 1: ChatGPT giving context on an esports game 18](#Figure!0|sequence)

[Figure 2: Output of "nvtop" displaying server's GPU usage 24](#Figure!1|sequence)

# *Abstract*

The rising popularity of esports has contributed to the demand for more accessible broadcasts to reach a wider audience. This project investigates how different artificial intelligence technologies can be used to tackle accessibility barriers within live esports broadcasting. Through a mixed-methods approach, including both a literature review and the development of a prototype, the project focused on finding solutions that could be implemented in real-time broadcasting scenarios, such as automatic subtitles, explaining esports jargon and translation. A functional prototype of a streaming platform that uses OpenAi’s Whisper Model for captioning and translation, paired with a definition system that used a fuzzy-matching algorithm to detect esports specific terminology was developed. This system was tested across simulated professional, grassroots and non-English broadcasts to evaluate its performance under varying conditions. This paper also discusses other technologies not used in the prototype and how they could be used to improve accessibility within esports broadcasting.

# *Acknowledgements*

*I would like to thank my friends who helped and supported throughout this project.*

*I would also like to thank my supervisor Joshua Lindsey for his advice and feedback throughout this project.*

# 1. Introduction

## 1.1. Research Background

Esports is a fast growing industry that is increasingly becoming embedded in mainstream culture, with major events such as the *League of Legends Worlds 2024* finals proving this trend by selling out the O2 arena in London and reaching more then 6.9 million viewers at its peak (*Worlds 2024 Finals: Faker Breaks Wins and Viewership Records*, 2024).As esports viewership continue to expand there is an ever increasing demand to make esports broadcasts accessible to ensure you can appeal to a wide range of people and attract more viewers.

This paper examines the accessibility barriers present within esports broadcasts and investigates how technology can be utilised to remove or reduce the affect of these obstacles. In this context accessibility issues rage from challenges new viewers may face when they are not familiar with the game being played or the competitive scene, to more critical barriers experienced by people with disabilities who may require assistive technologies when consuming video content.

There is a wide variety of technologies such as artificial intelligence and real-time overlays that can be used to potentially help boost understanding and accessibility when implemented within esports broadcasts and streaming services. In this paper I research and evaluate these technologies usefulness in boosting accessibility in both professional and grassroots esports environments.

As part of this paper a prototype for a streaming platform was developed. It interoperates some of the technologies discussed within the literature review, to evaluate their functionality and potential to improve the accessibility of esports broadcasts.

## 1.2. Aim

The aim of this paper is to discuss methods technology can be used to improve the accessibility of esports broadcasts and create a streaming site to demonstrate how some of these methods can be implemented.

## 1.3. Objectives

I have five primary objectives in this paper:

1. Identify the main accessibility barriers in esports broadcasting.

2. Investigate ways esports broadcasts can be more accessible for people unfamiliar with the game being played or esports as a whole.

3. Explore how esports broadcasts can be made more accessible for people with disabilities.

4. Examine different technologies and how they can be used to improve accessibility in esports live broadcast.

5. Design and implement a prototype designed to tackle some of the accessibility barriers within esports broadcast.

## 1.4. Deliverables

The deliverables for this project are:

* A literature review exploring where esports can be more accessible and different technologies and how they can be Implemented to boost accessibility.
* A prototype streaming site that demonstrates selected accessibility features in a simulated broadcast.

## 1.5. Structure of the Report

This report is organised into the following chapters:

* **Chapter 1 – Introduction:** Outlines the background, aim, objectives, deliverables, and structure of the research.

* **Chapter 2 – Research Methods:** Outlines the methodology I used when researching and methods I used to create the literature review and artefact

* **Chapter 3 – Literature Review:** Reviews existing academic literature on esports broadcasting and accessibility technologies along with discussing how these could be implemented to improve accessibility

* **Chapter 4 – Design and Implementation:** Details the design decisions and development of the prototype.

* **Chapter 5 – Testing and Validation:** Evaluates the prototype through testing, discussing functionality and potential impact.

* **Chapter 6 – Conclusion:** Summarises key findings and suggests areas for future work.
* **Chapter 7 – Reference List:** Provides a full list of sources referenced throughout the dissertation.

# 2. Research Methods

## 2.1. Research Methodologies

### 2.1.1. Research Philosophy

This study is based on a pragmatist research philosophy which prioritised piratical solutions to accessibility challenges in esports broadcasting. This philosophy helped inform the research methods to identify accessibility barriers and test technological interventions to ensure a focus of actionable outcomes rather then just relying on theoretical solutions.

### 2.1.2. Research Approach

For this study a mixed inductive–deductive approach was used. Inductive reasoning was used during the literature review to find accessibility barriers within esports and potential technologies to address them. Deductive reasoning guided the development and testing of various solutions, which were Whisper based live captioning, jargon detection and explanation and translation to implement them within a working prototype.

### 2.1.3. Research Strategy

A design science research (DSR) strategy was used to develop the streaming platform prototype as an artefact addressing accessibility barriers identified in the literature review. The prototype was tested across three recordings of broadcasts to simulate different broadcast scenarios.

1. EVO 2023 Grand Finals — This was picked to simulate a professional high production value event
2. Meltdown x FightLab Monthly January 2025 Guilty Gear Strive — This was chosen to simulate a grassroots low production value broadcast. The broadcast was also 4 hours meaning it could be used to test how the prototype performs after being under load for an extended period of time.
3. EVO Japan 2023 Grand Finals — A non-English broadcast for the purpose of testing real-time translation.

### 2.1.4. Methodological Choices

A multi-method approach was employed. A Qualitative analysis for my literature review to analyse existing research and identify accessibility barriers and technologies along with discussing ways the technologies could be implemented within esports broadcast. The artefact was tested to ensure all aspects of the artefact worked along with an analysis of the computer resource usage and how that affects its usability along with an informal test of the accuracy, usefulness and user experience.

## 2.2. Research Methods

Firstly a comprehensive literature review covered literature on, the accessibility challenges faced and solutions to them in traditional sports broadcasting, accessibility challenges specific to esports broadcasting, viewer motivation in esports and existing and emerging technologies for broadcast accessibility. In the literature review i also discussed different technological approaches to address accessibility challenges in esports broadcasting.

Next I developed a functional artefact which was tested and hosted on a Linux server running NGINX-RTMP. Other key technical details include

* Hardware: NVIDIA RTX 4060 GPU for high-performance AI tasks
* Software: Python for back-end development for its compatibility with AI technologies. FFMPEG was used for simple capture from the RTMP server.
* Captioning: Whisper's "turbo" model was chosen for use with English transcription due to its balance of speed and size (809 million parameters and ~8× speed relative to the baseline). CUDA was used for GPU acceleration greatly improving performance but restricting memory usage to the GPU's VRAM restricting the use of the largest model due to more memory requirements then the GPU has.
* Jargon Detection: This was implemented using a fuzzy matching algorithm built with the RapidFuzz python module, to match Whisper’s transcriptions with a dataset of 26 Guilty Gear -Strive- specific terms from *The Fighting Game Glossary*.
* Translation: Whisper's "medium" model was chosen over "turbo" for real time translation due to the fact that turbo cannot translate. Jargon detection was not used for the translated output, as different languages would have different terminology and thus would require a similar dataset to \*The Fighting Game Glossary\* in Japanese.
* Front end: Built with the Next.js framework. HLS.js was used to allow for HLS input from NGINX-RTMP and subtitles and definitions were received through a websocket.

The prototype was tested with recordings of three broadcasts mentioned above. These broadcasts were downloaded and re-streamed to simulate real broadcast scenarios. Testing involved evaluating if speech was being transcribed or translated and if jargon was being correctly identified and definitions displayed. Iterative adjustments were made when issues were found to improve accuracy and performance.

## 2.3. Ethical Considerations

* All test streams were publicly available content; no personal data collected.
* Open-source tools (Whisper, FFmpeg, NGINX, NGINX-RTMP) were used to avoid licensing concerns.
* Permission was granted by the creator of *The Fighting Game Glossary* to use glossary content for this research.

# 3. Literature Review

## 3.1. Introduction

Esports is a rapidly growing industry that is only expected to grow even further. The value of the global esports industry is expected to grow from 2.17 Billion US Dollars in 2024 to 6.82 billion US Dollars in 2030 (Yadav and Utreja, 2024).As the industry grows it will become ever more important to make viewership of esports events more accessible for all people. In this literature review i will examine existing research to find out how we can use technology such as artificial Intelligence to improve accessibility for new and disabled esports viewers. I first explored the reasons and motivations for viewers of esports then looked into different existing technologies that could be used to improve the viewership experience and accessibility of esports for new and disabled viewers.

## 3.2. Analyses

### 3.2.1. Theme 1: Esports

Hamari and Sjöblom, (2017) investigated the motivations people had for consuming esports media. They did this through a large scale study of people who watch esports. In their study they define esports as “a form of sports where the primary aspects of the sport are facilitated by electronic systems”. Their results indicate that escapism, knowledge acquisition, novelty and aggressiveness were found to positively predict esports spectating frequency. they also state that "Socialising with peers has been shown to be of great importance in traditional sports and also within eSports".

Goncu and Finnegan, (2021) discuss some of the challenges blind people face with traditional sports broadcast for example while regular broadcast provides a good experience for sighted people "blind spectators must tune in to radio broadcasts to receive information which is normally captured in the video stream, for example team formation provided via info graphics". They also demonstrate that augmenting traditional sports broadcast with features such as audio descriptions and 3D audio to enhance accessibility for blind spectators. Applying this research to esports artificial intelligence powered tools could be developed to narrate the in game actions of esports players (e.g., 'Player 2 uses their ultimate on player 7') alongside using already existing in game spacial audio features present in games ,such as in first person games where you can hear the direction of other players footsteps, and preserving that audio data for the viewer. This should help reduce the visual dependency of esports broadcasts.

however Goncu and Finnegan also warn that having audio descriptions forces viewers to divide their attention between the main content of the broadcast and simultaneously pay attention to the audio descriptions and that this alongside 3D audio will cause the viewer to have a "higher cognitive load and even stress on the consumer".

Hassan, Baltzar and Kämäräinen, (2024) discuss how esports broadcasts remain inaccessible to disabled audiences due to how many visual and audio cues there are and that this overabundance of fast paced information during esports matches can become overwhelming to neurodiverse people. They state how it is important for esports to be accessible for disabled people as up to 20% of people who play games have disabilities that restrict how they can play games. While the chapter is mostly focused on in person events they do suggest some accommodations that can be applied to online esports broadcast, namely "Captioning of commentary (with speaker, colour, emotions)"

Yu et al., (2022) investigates the spectator motivations of esports viewers. they did this through a qualitative survey where they found five primary reasons for spectators to watch esports; entertainment, to support a team or player, learning more about the game from pro players, interest in the casters or analysts and social aspects such as through a chat on the streaming site or other external platforms. This is supported by Qian et al., (2020) where they found that things such as "skill improvement and vicarious sensation" as motives for watching esports. Going off this it stands to reason that using technology to help people understand and learn about what they are watching would in turn boost the accessibility of esports.

Charleer et al., (2018) explored the challenges of real -time esports, with a specific focus on League of Legends and Counter-Strike: Global Offensive, and how they can better empower viewers to follow the fast paced nature of esports broadcasts. They did this by creating an interactive real-time dashboard to provide players about information about the game currently being played. Their research indicates that dashboards contribute to the spectator's insight and experience by clarifying complex game play dynamics, however they must be made with careful consideration such as avoiding oversimplification and managing screen space while adapting to different spectator motivations (e.g., learning vs. entertainment).

The Royal National Institute of Blind People outlines 7 best practices for making commentaries accessible for viewers with vision impairment. Many of these can also translate to esports such as important text on screen being read out loud, if there is important information on screen such as the name of the level or map being played on it should be read out loud. Another guideline important for esports is explaining any jargon or terminology and to not assume audience familiarity with niche sporting language (*Best practice for making TV sports commentaries inclusive and accessible for viewers with vision impairments*, 2023). It is also stated that commentators should describe the position of the ball and players in team sports such as football to provide spacial context which is similar to Goncu and Finnegan (2021) approach of providing blind and low vision audiences with 3D audio.

### 3.2.2. Theme 2: accessibility in broadcast

Simon, Torcoli and Paulus, (2019) examine object based audio , specifically MPEG-H, and how it can be used to remedy barriers to accessibility in broadcast and streaming. They begin by identifying barriers to accessibility in today's broadcast and streaming environment. one area to note is hearing, on top of people with hearing loss they state that "One of the most common complaints to broadcasters is about the low intelligibility of the speech due to the loud background music and noise" and that users would benefit from being able to control the levels of different audio sources such as background music and vocals. They state another barrier as sight where many people with sight loss require and that while people with sight loss typically use audio descriptions to understand context that would typically be given through visual clues. On top of physical disabilities they also state barriers such as literacy cognition and language.

One of the features of object based audio is the ability to implement immersive sound which Goncu and Finnegan explored as a way to improve accessibility in sports broadcast media for blind people. MPEG-H also provides Multichannel audio which can even be transferred over SDI which allows it to be compatible with the majority of current broadcast setups, they do not however mention whether this is compatible with NDI a new protocol for transferring audio and video over Ethernet which is a technology that has more recently seen an uptake in usage in new broadcast spaces. It also has support for the transmission of multiple alternative audio streams which could be used for different languages or different versions of the same language. This can be used to improve accessibility for visually impaired people as they can have the option to select between the regular audio channel and an audio description version. Similarly people with hearing impairments or are simply viewing in conditions where hearing the dialog may be difficult have to select a "Dialog+" version of the audio where the volume of the dialog is boosted and any background noises lowered.

While having audio descriptions may seem like a continent solution it is difficult to implement in live video as audio descriptions are best when made by trained professionals who have vision impairments or work directly with people who do however this process is very expensive and time consuming. Because of the high time and cost AI solutions have been developed to automate this process however they face challenges when it comes to accuracy and bias (Wang *et al.*, 2021; Soldan *et al.*, 2022). Gao et al., (2024) explain how Dense Video Captioning (DVC) can LLM/VLM's to generate an initial script of "formative descriptions" which is then given to a human for AD Post-Editing (APE). in this phase the initial AD are refined according to specific standards. They also note that if the performance of the DVC step was high enough this step would not be needed however as of the time of the paper this is not the case.

Current technology for creating audio descriptions for digital visual media has the problem of not being designed for a live video context. World Scribe seeks to solve this by providing live descriptions of what the user is pointing their camera at while providing dynamic descriptions providing the user with more detail when looking at a static image while sticking to brief descriptions for dynamic scenes. i believe this unique approach would be best suited for live esports as this system trades off detail of its descriptions for a lower latency providing more value in a real-time environment (Chang, Liu and Guo, 2024).

On the other side people who are deaf or suffer from hearing loss typically rely on captions while watching video content. For live content captions are typically generated using automatic speech recognition(ASR). Current ASR systems typically use long fragments of speech to more accurately identify what word is being said by looking at the surrounding words and using that context to pick the output with the highest probability. While this increases the accuracy it means that the output typically comes several seconds after the dialog has been said which can cause confusion however there is research into synchronising the subtitles with live video (de Castro *et al.*, 2011). Whisper is an ASR model developed by OpenAI which was trained by large-scale weak supervision on multilingual training data giving it the ability to not only transcribe speech in languages other than English but translate non-English speech into English (Radford *et al.*, 2022). Macháček, Dabre and Bojar, (2023) demonstrate how the whisper model can be converted into a real-time transcription system using the LocalAgreement (Liu, Spanakis and Niehues, 2020) algorithm. This system allowed them to reach a 3.3 second average latency.

### 3.2.3. Gap Analysis: Under explored areas.

While there is a lot of research on the accessibility challenges of traditional sports the same cannot be said for esports however we can take ideas designed for improving accessibility for traditional esports and adapt them to esports for example Goncu and Finnegan's approach of providing 3D audio is something that could be recreated in esports using the already pre-existing spacial audio features many genre of games such as first person shooter games, which have a multitude of titles popular within esports such as Counter Strike 2, Valorant and Overwatch 2, this 3D audio if preserved in the live capture software/hardware could then be transferred to the spectator using the object based audio system described by Simon, Torcoli and Paulus, (2019). This would allow blind and low-vision spectators to have a better understanding of where every character is in the game world while also allowing them to control the levels of each audio source separately giving the viewer more control to customise the audio to their needs. Having 3D audio may also improve the immersion of regular viewers.

Current AI-driven solutions for audio descriptions focus on pre-recorded content and are generally focused on pre-recorded content as demonstrated by Gao et al. 2024. If a solution which took a similar approach to world scribe (Chang, Liu and Guo, 2024) were to be implemented in esports it would allow blind viewers to gain a better understanding of what is happening on the screen with minimal delay. Another approach would be to export game states such as player location, stats or in game events like a team capturing an objective, this data could then be used For example in a game such as League of Legends "Lee sin is standing under the tower" or "Gragas used his ultimate ability". Work needs to be done into researching how audio descriptions should be used in esports broadcast and what information needs to be conveyed, for example when should you use a players name over the character they are playing?

Competitive esports often has its own "jargon" which can make esports unaccsessible to those unfamiliar with this and necessitated the creation of online glossaries from organisations such as the British Esports Federation and Esports Insider (Bryony-Hope, 2023; Nordland, 2023). This could be tackled by having on-screen or spoken explanations of a word new viewers may not be familiar with when spoken by commentators or other talent and could be powered by automatic speech recognition. This would also align with guidelines from the Royal National Institute of Blind People (RNIB) to minimise unexplained jargon.(*Best practice for making TV sports commentaries inclusive and accessible for viewers with vision impairments*, 2023) to minimise unexplained jargon.

The majority of player motivation studies focused primarily on the most popular esports which tend to fall within the "Multiplayer Online Battle Arena" (MOBA) or "First Person Shooter" (FPS) genre. Yu et al., (2022) only surveyed "League of Legends" viewers while Hamari and Sjöblom, (2017) did not mention fighting games as a genre of esports games. While fighting game events are generally more grass roots often held by local communities they have been growing in size and number. With the events being more grass roots the atmosphere tends to be more personal with many of the attendees and viewers having personal relationships with the players on stage. This along with other significant differences between smaller games and larger games with large corporate backings may cause a difference in spectator motivation which should be looked into.

## 3.3. Conclusion and Future Research

The literature review reveals several key insights into possible ways esports broadcasts could be made more accessible for disabled spectators. Several technological solutions show promise for improving accessibility. Audio descriptions generated in real time by AI could narrate player actions to give blind spectators context to what is happening on the screen. Building on this game data could be used to give blind spectators information that would normally be displayed on stream graphics though timing when this information is spoken carefully is crucial to avoid cognitive overload. Leveraging existing 3D spatial audio from games could enhance spatial awareness for blind viewers. For deaf audiences, real-time automatic speech recognition tools can be used to enhance accessibility however it is important that the latency is kept as low as possible as having too long of a latency can cause confusion. Object-based audio solutions like MPEG-H allow for the customisation of the levels of different audio sources to allow hard of hearing people to mix the audio to a level that is most comfortable for them while also providing the option for different audio tracks to allow for a audio description track for blind people (Simon, Torcoli and Paulus, 2019).

Beyond disabled people esports accessibility could be improved in a multitude of ways, both (Yu et al., (2022) and Hamari and Sjöblom, (2017) agree that skill improvement was a core motivator behind why people watched esports. This is further backed by Yu et al. (2022) who stated that learning more about the game was one of the primary motivators behind esports. Learning can be supplemented through the use of real time dashboards that viewers can interact with (Charleer et al., 2018). Building on this AI systems can be developed to detect when a whos definition may not be immediately clear to someone to familiar with esports is said and a explanation can be provided. Whisper can also be used to make non English broadcasts more accessible to English speakers due to its ability to translate audio in near-real-time (Radford et al., 2022). Building on this Large Language models such as OpenAi's GPT-4o can theoretically be used to help give viewers new to the scene context about players or teams. An example of this can be seen in *Figure 1* where ChatGPT is asked about when the last time two teams played were and is able to respond with the date, the tournament, the scores of each team and the players on the team at the time. Using another large language model trained to observe the subtitles and prompt another large language model trained to respond with useful information this process could be automated to give the viewer information without them needing to research it themselves. This could allow new watchers to understand any context they may be missing such as rivalries between teams and encourage them to support a team further motivating them to watch esports as supporting a player or team is a major motivator to watch esports (Yu et al., 2022).

Figure 1: ChatGPT giving context on an esports game

# 4. Design and Implementation

My primary goal when designing the artefact was to have a solution that could be easily implements in existing broadcast workflows with minimal effort required for broadcasters. The solution should also be able to be deconstructed, modified and reconfigured to match different broadcast workflows, esports games and accessibility needs. It was important that whatever solution was implemented it was as close to real-time as possible as to not disorientate or confuse viewers. Finally the solution should not be too resource intensive or expensive as to make it accessible to not only professional broadcasts but grassroots broadcasts too.

My initial plan was to create a system that provided blind and low vision spectators with audio descriptions giving more context to the players actions as they happen on the live broadcast however as I further researched this it quickly became clear that this was out of the scope of the project. I also plan on creating an AI system that detects when esports "jargon" is said and provide the watcher with a definition of the said jargon. To demonstrate this i chose the game "Guilty Gear -Strive-" (GGST) as it has many terms that only players experienced with the game will know while also having online resources, such as *The Fighting Game Glossary by Infil*

*To make sure this is as easy to implement into existing workflows I decided to develop a solution that would be implemented either on the streaming platform's side or act as a middle man in between the streaming platform and the broadcaster. This approach lessens the requirement on the broadcaster to have and operate more equipment. To do this I designated a server with a NVIDIA RTX 4060 graphics card running Linux to act as the streaming platform server for this prototype. I chose to host it on Linux as this is the standard for many web servers and for its greater compatibility with with AI systems. having a NVIDIA graphics card also allows me to run AI systems using CUDA drivers allowing for greater performance and speed which is vital when trying to get real time responses. This is also where the RTMP server to which the broadcaster live streams to is hosted. RTMP servers acts as a gateway where viewers can connect to watch live streams however in this case it is also being used by my artefact to process the stream. I chose the RTMP module for NGINX as it is free and open source. Another benefit of NGINX-RTMP is that it comes with support for HLS* (*What is HTTP Live Streaming? | HLS streaming*, no date) and allows for live streams to be easily viewed on a web browser.

Once I had prepared the RTMP server i first attempted to capture the video from the server using OpenCV's python module (CV2). I chose python for this implementation due to its compatibility with modern AI technologies and for its extensive library ecosystem which made developing the implementation much simpler. once the RTMP stream was captured by CV2 I overlayed a transparent image onto the stream and then output the stream to a sub-process running FFMPEG, a highly versatile and lightweight command line tool capable of encoding live video. FFMPEG then output the video back to the RTMP server for viewers. The purpose of this was to test how processing the video in python would affect the delay, this resulted in the video being heavily desynced with the audio and would also cause video artifacting making this approach unsuitable and after doing more research into creating audio descriptions I decided they were out of the scope of the project.

When developing the AI definition provider the first step was to transcribe the audio being spoken. I did this by first using python to create another FFMPEG sub process to capture the audio of the RTMP stream as a ".wav" file that is added to in length as more audio is captured. Next the program loads the "turbo" Whisper model, the turbo model was chosen for its 809 million parameters and ~8× relative speed making it the second largest and fastest Whisper model, with the fastest model "tiny" being the smallest and the largest model being the slowest. After loading the model runs a loop where the wav file is given to Whisper and the outputed transcription is writen to a text file. This process works to provide transcriptions but slows down considerably when you exceed Whisper's recomended input length of 30 seconds, continuously re transcribing audio that has already been transcribed is also an inefficient use of resources. To solve this i used "pydub", a python module for manipulating audio, to create a new wav file consisting of only the most recent 30 seconds of audio. This process was quickly changed as to cut down on unnecessary SSD read write operations and minimise latency, this was done by instead writing the last 30 seconds into a memory buffer skipping the need to write to the disk.

Whisper also frequently suffers from "hallucinations" where it will try and detect speech in segments of audio where there is none, I implemented two solutions to combat this. Firstly i padded the start and end of each segment with a second of completely blank audio and reduced the segment length to 28 seconds to compensate. I did this as whisper often guessed that any sound at the start or the end of the audio was a word that was cut of and makes a guess at what it might of been. padding with a second of audio with no sound makes it clear that there is no speech. The second thing I did was wait to see if a sentence transcribed by whisper appears after the segment has been transcribed twice if it has it is less likely it is a hallucination.

Next I developed a front end website where the streams would be viewable. The website was developed using Next.js a React based framework for creating web applications. The website consists of two components, a live stream player powered by "hls.js" and a box for the subtitles setup to receive subtitles from a websocket. The websocket server would act as the bridge allowing me to send subtitles from the backend where they are being transcribed to all browsers currently running the front end. I then configured the websocket server and had my backend send subtitles to the server once it knew they were not hallucinations.

Finally to create the definition system I first made a "target words" json file that contained all the words the program should give definitions for. The json has two lists "words" and "phrases" every object in the json had 3 values in common: a word or phrase to match to, a brief definition of the word and a link to its entry in *The Fighting Game Glossary where viewers could see more information about the word and related topics. The difference between words and phrases is that some words are used in phrases such as "burst" and "burst safe" to avoid triggering both the word and the phrase when the program matches a word it will first check if the phrase stater value is false, if it is the program will just output the words definition if it is true the word object will have a list of phrases it appears in. The program then iterates through this list and looks at each phrase objects "search term" which in the example of "burst safe" is "safe". The program then looks before and after the word for the search term and if it appears the definition of the phrase is returned. Finally if no phrase is found the program will default back to the original word.*

*One problem I faced is that whisper will not always transcribe a word as it is written in the target words file and this causes the program to miss some words. To solve this I used "RapidFuzz" a fuzzy matching module for python. Fuzzy matching is used for identifying if two strings are the same value without being exact matches. It dose this by using an algorithm to score how close two strings are* (Chaudhuri *et al.*, 2003). Finally I simply updated the subtitles box component on the front end website and it now gives definitions for "fighting game jargon" as it is spoken on stream.

Due to the subtitles being powered by Whisper this system can easily be adapted to translate non English speech into English. While in transcription mode it can also transcribe up to 99 different languages although the target words file will need to be updated for the new language. It is important to note however that Whisper may not be as accurate for less spoken languages where there is less training data available.

For the translation system the same process was used with some notable changes. Firstly the whisper model used was "medium", this is due to the fact that the "turbo" model dose not have access to translation. The other change was was there was no jargon detection as different languages would have different terminology it would require a glossary of the jargon along with testing how whisper translates the terms when spoken. It would also require someone fluent in the language and esports game to meaningfully test.

# 5.0 Testing and Validation of the Artefact

The aim of this prototype was to create a proof of concept demonstrating various solutions for enhancing accessibility in esports broadcasting. The primary goal was to show that such improvements are both feasible and valuable, warranting further research and development. To this end I tested it to ensure its functionality.

To test its subtitling capability alongside its ability to correctly identify esports jargon. I did this by downloading 2 recordings of previous *Guilty Gear -Strive-* (GGST) live streams, I would then re-stream these recordings to the RTMP server to simulate a live broadcast.

The initial recording used to test the system was the *2023 Evo Guilty Gear -Strive- Grand Finals*. This event was selected due to its professional-level production quality and the presence of experienced casters, ensuring that Whisper would have an optimal environment for accurate transcription. This choice was intended to establish a baseline for assessing system performance under ideal conditions.

One of the first issues discovered was Whisper's tendency to merge two or more sentences into a one segment when there is no apparent pause between them. While this may be something that can be corrected with the implementation of Whisper-Live, as discussed by Macháček, Dabre and Bojar, (2023), it would consist of a complete rewrite of the existing codebase with significant alterations to the definition system and how subtitles are currently being passed to the front end website to accommodate differences between Whisper and Whisper-Live. While this issue negatively affects readability and the overall user experience, it does not hinder the core functionality of the prototype. Thus, i decided to leave the system as is for the purposes of this proof of concept. There were many false positives, all giving a fuzzy match score of "80.00". This issue was fixed by incrementally increasing the fuzzy threshold to "81", as the most accurate matches all consistently scored 81 or more. This tiny adjustment greatly improved the detection system's precision without negatively impacting its performance.

The second recording I used was *Meltdown x FightLab's* January monthly Guilty Gear Strive tournament. This was chosen for three reasons, the first was the fact that FightLab is a smaller grassroots organisation and the event was held in a busy gaming bar. These factors mean that the stream was done using consumer-grade equipment with nobody monitoring the audio levels in an environment with lots of background noise, all these factors together would help test how Whisper and the prototype has a whole perform in less then ideal conditions. Building on this FightLab has an "open mic" system for their casters meaning anyone in attendance can try casting meaning the casting will not be by people who are trained to speak clearly into a microphone, further testing the limits of the system. Finally to test how the system performs after running for a long period similar to regular live broadcasts, this length also has the added benefit of giving the prototype more opportunities to detect words.

One thing I discovered from this second test is that the prototype struggles when phrases are rephrased. An example of this was during the test the phrase "break the wall" was said but did not trigger the definition of "wall break". A similar issue arises when a word that like "risk" is said as it is a very common word in the English language but has a different meaning when referring to the game systems of GGST, this caused the definition of "risk" in GGST to be displayed when talking about a players risky game play. Both these problems are caused by the matching system not looking at the greater context of the sentence being said. One way this could be tackled would be to use a fine tuned large language model (LLM) to identify the words, this would work as the LLM could look at the greater context of the sentence and better understand if a word being said is used in the context that is given in the target words list or if the word is being used in a context different to how it is defined in the target words file. However this would have the downside of consuming a large amount of GPU resources in an already GPU intensive solution. This would have the affect of making any real world implementation of this solution more expensive to operate

Transcribed Text: Wild as all.

New Detected Words/Phrases:

 - wild assault (score: 81.82)

On the other hand the fuzzy matching system is able to correctly identify words and phrases that are misheard by Whisper. Above you can observe Whisper mishearing someone saying "Wild assault" as "Wild as all" however it is still detected through fuzzy matching and the definition is able to be returned. Despite these issues Whisper's transcriptions were not noticeably worse and handled less then optimal conditions very well.

To test whispers ability to translate for a live esports setting I used a segment of Japanese GGST broadcast and watched it using the translated subtitles. Using the translated subtitles I was able understand basic sentences and gain insight from the casters however many things such as names of characters, their attacks or player names were miss translated causing mild confusion.

When it came to resource usage Whisper would draw around 30W of power on average and consistently use upwards of 90% of the GPU's compute power. Furthermore the GPU's internal temperature rose to an average of 65° from a idle temp of 40°. It's worth noting that testing was conducted on a very warm day in a poorly ventilated room which may have contributed to higher observed GPU temperatures.

Figure 2: Output of "nvtop" displaying server's GPU usage

This prototype was able to successfully demonstrate the feasibility of enhancing accessibility in live esports broadcasting through the use of near real-time subtitling, jargon detection and translation. By simulating both high quality and grassroots broadcast environments, it displayed the systems strengths and limitations. Whisper provided a foundation for the system by effectively transcribing the audio, although it did suffer from being unable to effectively segment the audio during fast segments of speech and is an area the system could be improved in the future. The fuzzy matching algorithm, while initially giving many false positives improved significantly after making a small adjustment to the match threshold and thus, was able to detect misheard words and phrases with promising accuracy. While the implementation of a fine-tuned large language model could be used to address words being matched in the wrong context, it would result in a greater computational cost, raising concerns for its practicality for deployment in grass roots broadcasts with limited space and budget. If implemented into a full solution such as a streaming site this could increase esports viewership as the barrier to entry would be lowered meaning new viewers are more likely to continue watching in the future. Live translation would also allow English viewers to follow major esports events not in their language therefore expending the markets of their broadcasts. Despite the challenges it faced the prototype proved that using this technologies to improve the accessibility of esports is possible, and could be further improved with more research into the topic.

# 6. Conclusion

This paper set out to address gaps in the accessibility of live esports broadcasts by investigating how different emerging technologies could be utilised to reduce barriers for both disabled viewers and newcomers to the esports scene. Through a combination of a literature review and a prototype implementation, this project was able to explore different solutions possible with current technology and implement some of them proving their feasibility while also revealing the limitations of the current technology and how they could be improved in the future. In this chapter i will give a summery of my finding and results, their implications and recommendations for further improving accessibility in esports live broadcasting using methods described in this paper but not implemented during this project.

## 6.1. Key Findings & Contribution to Objectives

### 6.1.1. Accessibility Barriers Identified (Objective 1)

The literature review revealed a variety of accessibility barriers of which the following were chosen as a focus for the implementation:

* **Jargon-Heavy commentary:** Terms such as "burst safe" or acronyms like "QRC" in *Guilty Gear -Strive-* can create confusion for newcomers unfamiliar with the terms. it also negatively impacts blind people as it causes misalignment with the Royal National Institute of Blind People's best practice guidelines about unexplained niche terminology (*Best practice for making TV sports commentaries inclusive and accessible for viewers with vision impairments*, 2023).
* **Lack of real-time solutions:** traditional accessibility solutions are typically built on pipelines that include the involvement of trained professionals and are not conducive for being used in real time scenarios.

### 6.1.2. Solutions for Disabled or New Audiences (Objective 2,3&4)

From the literature review I came up with three solutions to boost accessibility, these were:

* **Subtitles:** Real time subtitles using Whisper's "turbo" model to provide high accuracy and speed subtitles. It was able to provide accurate results in even suboptimal conditions. These subtitles would improve accessibility for people who are deaf or hard of hearing.
* **Jargon explanation:** By feeding the subtitles into an algorithm designed to detect words from a curated glossary, definitions could be provided to the viewer whenever "jargon" was detected. The algorithm was made to be robust by using fuzzy matching as to still understand when words even when they are misheard by Whisper. However, limitations emerged in handling rephrased phrases (e.g., "break the wall" vs. "wall break") and contextual ambiguity (e.g., "risk" as a game play mechanic vs. as word to denote danger). These gaps could be solved by an AI system such as an LLM which is able understand contextual nuance and is tuned to understand esports terminology and identify them when spoken at the cost of higher computational resources required.
* This would benefit not only newcomers but blind audiences if the definitions are accessible by a screen reader
* **Translation:** Using Whisper's "medium" model enables non-English to English translation at the cost of the accuracy and speed of the "turbo" model. Whisper's translation usual struggles with things such as proper nouns as they will typically not translate into names English audiences are familiar with causing confusion.

These solutions were all designed to work either together or separate from each other so they can be used and implemented in a variety of broadcast workflows.

### 6.1.3. Prototype Validation (Objective 5)

Testing was conducted across three different broadcast environments which were professional (EVO 2023), grassroots (FightLab), and non-English (EVO Japan 2023). The testing demonstrated that the systems core functionality worked while allowing me to iteratively improve on it, however it also exposed the limitations of the design.

* The transcription accuracy was not noticeably different between professional and grassroots scenarios.
* The system was able to maintain a stable performance for over 4 consecutive hours without degradation in performance.
* Whisper did not separate sentences into separate segments when sentences were spoken close together causing subtitles to be multiple sentences long.
* GPU utilisation reached 90% (30W power draw, 65°C temps), raising concerns about scalability for low-budget broadcasts.

## 6.2. Implications for Esports Accessibility

One implication this project has on esports accessibility is that it can serve to lower the entry barriers of watching esports. By explaining jargon in real time the prototype enables new viewers to engage with and understand the content they are watching. This aligns with the viewer motivations "skill improvement and vicarious sensation" outlined by Qian *et al.*, (2020). Having real time translations of non English broadcasts to English would lead to a larger number of people consuming non-English esports broadcasts and increase their market. It would also mean a greater level of inclusive for disabled viewers as having jargon explained benefits blind viewers (*Best practice for making TV sports commentaries inclusive and accessible for viewers with vision impairments*, 2023) while also benefiting deaf audiences through subtitles. This solution while beneficial is highly GPU intensive making it less viable for grassroots adoption without assistance though a cloud based service, as many grass roots broadcasts work on a tight budget.

## 6.3. Limitations and Future Work

One of the biggest ways this system could be improved would be to implement the system using Whisper-Live (Macháček, Dabre and Bojar, 2023) as this would reduce the time it takes for the transcriptions to be processed while also solving the segmentation issues. Subtitles with lower delay would go a long way to reducing confusion from delayed subtitles for viewers who rely on them. it would also save on computing resources as the same audio is not being processed multiple times.

The Jargon detection system could also be improved by implementing a system that could understand the context of where the word is being said. Future research could experiment with fine-tuning a lightweight LLM to distinguish game play specific terms from their synonyms in common dialect and could understand when words have been rephrased. Multilingual Glossaries could also be developed to allow for the system to work across languages and when speech is being translated by whisper. Whisper's translation is also only available for non-English to English translations meaning the translation system cannot be used by non-English viewers to watch English broadcasts or other non-English broadcasts (e.g., Japanese to Portuguese).

Audio description systems could also be developed in a multitude of different ways. One way would be to have an AI system describe what is happening on the screen however this approach may struggle with latency issues. Another way would be to use developer support and APIs to Leverage the live game state to gain information such as individual player locations, what abilities they are using or kill feeds and develop a system that decides what information is relevant to a blind viewer and describe it to them in the form of audio descriptions. Finally utilising the 3D audio many first person games have and preserving that data for viewer would help blind viewers visualise where things are happening in the game world, however this approach would likely require a large overhaul of any broadcasting systems from game capture all the way to the streaming site itself to ensure the audio is preserved

Finally formal usability studies should be conducted with disabled esports viewers to test the usefulness of features of MPEG-H streaming such as audio mixing in the context of live esports broadcast and if they noticeably improve the experience for disabled audiences.

## 6.4 Final Statement

This project conclusively demonstrates that AI-driven technologies can meaningfully enhance esports accessibility today. Although there are still contextual and computational issues, the prototype provides a basic framework on which broadcasters and developers can expand. By prioritising inclusive design and leveraging advancements in real-time AI, the esports industry can transform from its origins as a niche entertainment form into a universally accessible global phenomenon. The next frontier lies in bridging the gap between technological possibility and practical implementation, ensuring accessibility becomes a cornerstone of esports’ meteoric growth.

# 7. Refrence List

*Best practice for making TV sports commentaries inclusive and accessible for viewers with vision impairments* (2023) *RNIB*. Available at: https://www.rnib.org.uk/professionals/research-and-data/reports-and-insight/best-practice-for-making-tv-sports-commentaries-inclusive-and-accessible-for-viewers-with-vi/ (Accessed: 2 May 2025).

Bryony-Hope (2023) *A-Z of esports & competitive gaming jargon*, *British Esports Federation*. Available at: https://britishesports.org/the-hub/about-esports/competitive-game-jargon/ (Accessed: 2 May 2025).

de Castro, M. *et al.* (2011) ‘Real-time subtitle synchronization in live television programs’, in *2011 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. *2011 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–6. Available at: https://doi.org/10.1109/BMSB.2011.5954889.

Chang, R.-C., Liu, Y. and Guo, A. (2024) ‘WorldScribe: Towards Context-Aware Live Visual Descriptions’, in *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. *UIST ’24: The 37th Annual ACM Symposium on User Interface Software and Technology*, Pittsburgh PA USA: ACM, pp. 1–18. Available at: https://doi.org/10.1145/3654777.3676375.

Charleer, S. *et al.* (2018) ‘Real-Time Dashboards to Support eSports Spectating’, in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. *CHI PLAY ’18: The annual symposium on Computer-Human Interaction in Play*, Melbourne VIC Australia: ACM, pp. 59–71. Available at: https://doi.org/10.1145/3242671.3242680.

Chaudhuri, S. *et al.* (2003) ‘Robust and efficient fuzzy match for online data cleaning’, in *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*. New York, NY, USA: Association for Computing Machinery (SIGMOD ’03), pp. 313–324. Available at: https://doi.org/10.1145/872757.872796.

Gao, Y. *et al.* (2024) ‘Audio Description Generation in the Era of LLMs and VLMs: A Review of Transferable Generative AI Technologies’. arXiv. Available at: https://doi.org/10.48550/arXiv.2410.08860.

Goncu, C. and Finnegan, D.J. (2021) ‘“Did You See That!?” Enhancing the Experience of Sports Media Broadcast for Blind People’, in C. Ardito et al. (eds) *Human-Computer Interaction – INTERACT 2021*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 396–417. Available at: https://doi.org/10.1007/978-3-030-85623-6\_24.

Hamari, J. and Sjöblom, M. (2017) ‘What is eSports and why do people watch it?’, *Internet Research*, 27(2), pp. 211–232. Available at: https://doi.org/10.1108/IntR-04-2016-0085.

Hassan, L., Baltzar, P. and Kämäräinen, M. (2024) ‘Esports for People With Disabilities’, in *Routledge Handbook of Esports*. Routledge.

Liu, D., Spanakis, G. and Niehues, J. (2020) ‘Low-Latency Sequence-to-Sequence Speech Recognition and Translation by Partial Hypothesis Selection’, in *Interspeech 2020*. *Interspeech 2020*, ISCA, pp. 3620–3624. Available at: https://doi.org/10.21437/Interspeech.2020-2897.

Macháček, D., Dabre, R. and Bojar, O. (2023) ‘Turning Whisper into Real-Time Transcription System’, in S. Saha and H. Sujaini (eds) *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: System Demonstrations*. Bali, Indonesia: Association for Computational Linguistics, pp. 17–24. Available at: https://doi.org/10.18653/v1/2023.ijcnlp-demo.3.

Nordland, J. (2023) ‘Esports — a glossary of industry terms’, *Esports Insider*, 24 November. Available at: https://esportsinsider.com/2023/11/esports-glossary-industry-terms (Accessed: 2 May 2025).

Qian, T.Y. *et al.* (2020) ‘It is in the game: dimensions of esports online spectator motivation and development of a scale’, *European Sport Management Quarterly*, 20(4), pp. 458–479. Available at: https://doi.org/10.1080/16184742.2019.1630464.

Radford, A. *et al.* (2022) ‘Robust Speech Recognition via Large-Scale Weak Supervision’.

Simon, C., Torcoli, M. and Paulus, J. (2019) ‘MPEG-H Audio for Improving Accessibility in Broadcasting and Streaming’. arXiv. Available at: https://doi.org/10.48550/arXiv.1909.11549.

Soldan, M. *et al.* (2022) ‘MAD: A Scalable Dataset for Language Grounding in Videos from Movie Audio Descriptions’, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5016–5025. Available at: https://doi.org/10.1109/CVPR52688.2022.00497.

*The Fighting Game Glossary | infil.net* (no date). Available at: https://glossary.infil.net/index.html (Accessed: 2 May 2025).

Wang, Y. *et al.* (2021) ‘Toward Automatic Audio Description Generation for Accessible Videos’, in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. *CHI ’21: CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, pp. 1–12. Available at: https://doi.org/10.1145/3411764.3445347.

*What is HTTP Live Streaming? | HLS streaming* (no date). Available at: https://www.cloudflare.com/learning/video/what-is-http-live-streaming/ (Accessed: 2 May 2025).

*Worlds 2024 Finals: Faker Breaks Wins and Viewership Records* (2024) *Esports Illustrated On SI*. Available at: https://www.si.com/esports/league-of-legends/worlds-2024-finals-faker-breaks-viewership-record (Accessed: 2 May 2025).

Yadav, M.K. and Utreja, K. (2024) ‘eSports on the Rise: A Statistical Odyssey Unveiling the Probability of Growth’, in *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*. *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, pp. 1–5. Available at: https://doi.org/10.1109/ICEECT61758.2024.10739039.

Yu, W.B. *et al.* (2022) ‘Exploring Esports Spectator Motivations’, in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI EA ’22), pp. 1–6. Available at: https://doi.org/10.1145/3491101.3519652.

# Appendix A: Ethics Form

## Block University LogoResearch Ethics

## Disclaimer Form

The following declaration should be made in cases where the researcher and the supervisor (where applicable) conclude that it is not necessary to apply for ethical approval for a specific research project.

**PART A: TO BE COMPLETED BY RESEARCHER**

|  |  |
| --- | --- |
| Name of Researcher: | Khizer Khan |
| School | School of Digital,Technology,Innovation andBusiness  |

|  |
| --- |
| **Student/Course Details (If Applicable)** |
| Student ID Number: | 22025079      |
| Name of Supervisor(s)/Module Tutor: | Joshua Lindsey      |
| PhD/MPhil project: |  |  |
| Taught Postgraduate Project/Assignment: |  | Award Title:Module Title: |            |
| Undergraduate Project/Assignment: |  |

|  |  |
| --- | --- |
| Project Title: | Boosting Accessibility of Live Broadcasting in Esports Using AI Audio Description and Captioning      |
| Project Outline: | In this project I will be creating a streaming platform witch will do three things to improve the accessibility of esports broadcasts. Firstly I will create audio descriptions for live esports broadcasts to help provide context blind viewers may otherwise miss. I will do this by using visual language models and Large Language Models. Secondly I will incorporate Open AI's open source speech recognition neural network Whisper to provide accurate subtitles and also translate non English broadcasts into English and vice versa. I also plan on testing other open source transcribing software and compare the difference Finally I will be providing definitions for terms specific to the current esport being played. For the purpose of this project I will be using a fighting game called Guilty Gear -Strive- and the definitions will be sourced from The Fighting Game Glossary a free online resource. When a term specific to the genre or game is said the site will then give the user the option to see the definition for the term |
| Give a brief description of research procedure (methods, tests etc.) | i will not be collecting any data. I will only be testing multiple solutions to see what gives the best performance      |
| Expected Start Date: | 11/11/24      | Expected End Date: |     02/05/25  |

**Declaration**

I/We confirm that the University’s Ethical Review Policy has been consulted and that all ethical issues and implications in relation to the above project have been considered. I/We confirm that ethical approval need not be sought. I/We confirm that:

|  |  |
| --- | --- |
| The research does not involve human or animal participants |  |
| The research does not present an indirect risk to non-participants (human or animal). |  |
| The research does not raise ethical issues due to the potential social or environmental implications of the study. |  |
| The research does not re-use previously collected personal data which is sensitive in nature, or enables the identification of individuals. |  |
| Has a risk assessment been completed for this project? |  Yes N/A |

|  |  |  |  |
| --- | --- | --- | --- |
| Signature of Researcher: |       | Date: |   10/11/24    |
| Signature(s) of Project Supervisor(s)(If student) ORSignature of Head of Department/ Senior researcher (if staff) |       | Date: |       |

**NB:** If the research departs from the protocol which provides the basis for this disclaimer then ethical review may be required and the applicant and supervisor (where applicable) should consider whether or not the disclaimer declaration remains appropriate. If it is no longer appropriate an application for ethical review **must** be submitted.

# Appendix B : Health and Safety Risk Assessment

| Staffordshire University Logo**GENERAL RISK ASSESSMENT FORM** | **Severity** multiplied by **Likelihood** equals **Risk Rate**.NB: Calculated after taking in to account existing precautions |
| --- | --- |
| Department: Department for Digital Technologies and Arts | Severity | Insignifi-cant (1) | Minor (2) | Moder-ate (3) | Serious (4) | Fatal / Critical (5) |
| Likelihood |
| Task/Activity/Area: Final Year Project | Almost Certain (5) | **5** | **10** | **15** | **20** | **25** |
| Likely (4) | **4** | **8** | **12** | **16** | **20** |
| Student ID: 22025079 | Signature:  | Possible (3) | **3** | **6** | **9** | **12** | **15** |
| Unlikely (2) | **2** | **4** | **6** | **8** | **10** |
| Date of Assessment: 10/11/24 | Review Date:10/11/24 | Rare (1) | **1** | **2** | **3** | **4** | **5** |

| **ID** | **Activity/Process/Machines** | **Hazard** | **Persons in Danger** | **Severity1-5** | **Likelihood1-5** | **Risk Rate** | **Measures/Comments** | **Result** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Using open source software | Cyber security risk | me | 3 | 1 | 3 | I will ensure that I take cyber security best practices into account when developing my solution | T |
|  | Having a public website | Cyber security risk | Me | 3 | 1 | 3 | I will ensure that I take cyber security best practices into account when developing my solution and only have the website operational during testing | T |
|  | Trip hazard from cables and furniture around the room |  Minor fall injury's from the trip | Anyone at the event | 1 | 2 | 2 | The room will be organised in a way to make sure there is plenty of space which should reduce the risk | T |
|  | Electonic equipment, could result in shock if liquid spilled on them | Electronic shock | Anyone using the equipment | 2 | 1 | 2 | People will be kept out of the control room. When at stations only drinks with lids will be allowed to reduce spill risk. | A |
|  | Fights between players | Altercations between players have the possibility to result in physical fights and incur injury's from fights such as broken nose. | Competitors | 2 | 1 | 2 | This is a casual event and this is unlikely in the first place but there will also be people around who can defuse the situation if there was to be an issue. | T |
|  | Screens and equipment | Staring at screens for a long time in poor conditions can result in dizziness, headaches and nausea. | Competitors | 1 | 2 | 2 | Ensure that lighting and set ups are comfortable to reduce eye stress. Include breaks to allow for eyes to recover | T |
|  | Fire | There is always a risk of fire and injury associated although this isn’t event specific. | Everyone attending | 1 | 4 | 4 | The room and building is already up to fire safety standards and the alarms are working. External competitors should be informed of fire drill and emergency exits. | A |
|  |  |  |  |  |  |  |  |  |

Key to result **T** = Trivial Risk **A** = Adequately Controlled **N** = Not Adequately Controlled **U** = Unable to decide (further information required).