UTRECHT UNIVERSITY

Department of Information and Computing Sciences

Computing Science master thesis

Building Interpersonal Skill Profiles for Players in Competitive Team-Based Online Games

First examiner:

Dr. Julian Frommel

Second examiner:

Dr. Johannes Pfau

Candidate:

Jin Kai Huang

In cooperation with:

Patterns GG Limited

September 10, 2024

Abstract

The world of esports has seen rapid growth in the past decade. Despite this rapid growth, non-technical skills, such as communication, in a competitive team-based environment are an understudied topic, even though professional players, experts, and the esports community alike, do agree that non-technical skills and team dynamics affect the performance and potential of a professional esports team. Other research fields, where high-pressure teams play a central role, have realised the importance of non-technical skills, which led to the implementation of various methods to measure these skills. These methods can be applied to the esports industry as well due to the similarities between teams in competitive team-based online games and high-pressure teams in sectors such as the military, aviation, and emergency services. In this study, interpersonal skill profiles were defined based on methods used in these other fields. Participants were recruited whose data were collected. These data were then cleaned and manually labelled. Afterwards, interpersonal skill profile scores were defined based on literature from related fields and domain knowledge. These scores were then calculated using quantitative measures derived from the input data and visualised to allow the comparison of players' non-technical skills. Lastly, an automatic labelling model was trained using the manually labelled data to automate the entire implementation. The results showed that the profiles were able to differentiate between players based on their non-technical skills, and that there is potential in applying team dynamics related methods used in other research fields to the world of esports.

Contents

1 Introduction **4**

2 Related Work **7**
2.1 VALORANT 7
2.2 Analysing Data in Online Games 8
2.3 Team Dynamics in Online Games 10
2.4 Team Dynamics in High-Pressure Teams 13
2.5 Research Motivation 15

3 Methodology **17**
3.1 Data Collection 17
3.2 Interpersonal Skill Profiles 18
3.3 Mapping Data to Interpersonal Skill Profiles 19

4 Results **27**
4.1 Automatic Labelling Model 27
4.2 Interpersonal Skill Profiles Algorithm 29

5 Discussion **34**
5.1 Research Goals 34
5.2 Implications 37
5.3 Limitations 37

6 Conclusion **39**

Appendix

A Ethics and Privacy Scan Application **40**
A.1 Coordinator Approval 40
A.2 Submitted Application 40

B Interpersonal Skill Profile Components **51**
B.1 Table of Adapted Components (1/2) 51

B.2 Table of Adapted Components (2/2) 52

C Per-Match Basis Profile Scores 53

C.1 Per-Match Basis Profile Scores of Team 1 53

C.2 Per-Match Basis Profile Scores of Team 2 54

Bibliography 61

1. Introduction

The world of esports has seen rapid growth in the past decade, with increasingly more game developers, potential sponsors and investors, and of course fans and gamers becoming more interested [1]–[3]. Despite the increasingly higher stakes and prize money professional esports teams compete for at the highest level, team dynamics and non-technical skills in teams in competitive team-based online games have been an understudied topic [3], [4]. Even though there is a lack of documented research, professional players, experts, and the esports community alike, do agree that having a team that consists of the most technically capable players on paper does not necessarily lead to success, with many notable examples across a wide variety of esports titles in the past few years [5]–[14]. This suggests that team dynamics and non-technical skills play a large role in competitive team-based online games.

Meanwhile, non-technical skills have become increasingly more important across various fields and industries where high-pressure teams play a central role. For example, in the aviation community, considerable emphasis has been put on crew members' non-technical skills as a crucial factor for enhanced safety [15]. This development led to the introduction of behavioural marker systems in these fields to evaluate the non-technical skills of the high-pressure team members, e.g., *NOTECHS* in the aviation sector to evaluate non-technical skills of pilots [16]. Tan *et al.* showed that principles and methods that are being applied in other fields can be applied to esports as well due to the similarities between teams in competitive team-based online games and high-pressure teams in sectors such as the military, aviation, and emergency services [17]. This approach could solve the problem of building competitive teams that are bound to fail due to its members having clashing interpersonal skills (meaning how they interact and communicate with others), and thus, will not get along.

This research project investigates team dynamics within teams that compete at a high level in competitive team-based online games. It aims to define interpersonal skill profiles based on categories inspired by behavioural marker systems used in other research fields, communication metrics, and in-game data such as match metadata. Furthermore, it aims to solve the largest downside of behavioural marker systems, which is the need for an expert rater to manually evaluate the categories of these systems [15], [16]. After defining the interpersonal skill profiles, they can be automatically assigned to players by mapping communication data and other in-game data to the interpersonal skill profile components. The entire process requires a great deal of data and analysis. Therefore, the esports scene is a perfect candidate for this study due to its size and growth, meaning there is a large amount of potential data to be collected [1], [3]. There exists a wide variety of esports titles across various video game genres and platforms, with the most popular titles of 2023 being *League of Legends*, *Mobile Legends: Bang Bang*, *Counter-Strike*, *VALORANT*, and *Dota 2* [18]. This research project will focus on *VALORANT*, which is a 5-versus-5 character-based tactical first-person shooter developed by Riot Games [19]. A more in-depth explanation about the game will be given in Section 2.1.

In this study, interpersonal skill profiles were defined based on behavioural marker systems employed in other research fields and an automatic approach was implemented to assign the profiles to players in teams in competitive team-based online games. Firstly, participants were recruited whose voice and in-game data were collected. The voice data were then transcribed, cleaned, and manually labelled. The manually labelled data were later used to train the automatic labelling model. The approach in this study consisted of two parts. The first part consisted of the aforementioned automatic labelling model that assigns labels to voice chat transcriptions. These labels were defined based on research done in other fields and domain knowledge. The second part of the approach consisted of an algorithm that calculates interpersonal skill profile scores based on quantitative measures that include labels assigned in the first part. Four scores were defined based on literature from other fields, domain knowledge, and qualitative observations made during the data collection and cleaning process. After calculating the scores, the algorithm can visualise the interpersonal skill profile of a player in various charts. The results showed that the automatic labelling model had a validation accuracy of 82.1% and that the algorithm was able to differentiate between players based on the four interpersonal skill profile scores that make up their non-technical skills.

The main contributions of this study are: (1) defining interpersonal skill profiles based on team dynamics research done in other fields, (2) implementing an automated approach that assigns interpersonal skill profiles to players in teams in competitive team-based online games, (3) showing that the approach is able to differentiate between players based on their non-technical skills, and (4) showing that there is potential in applying methods used in related research fields to the world of esports.

2. Related Work

In order to properly motivate how this research project contributes to the research areas it is related to, it is important to grasp the main concepts and relevant recent advancements in these areas and identify where the gaps lie. Section 2.1 will give an in-depth explanation about *VALORANT* and explain why it is a good fit for the project. Sections 2.2, 2.3, and 2.4 will touch upon the aforementioned research areas and show how this research project differs and takes inspiration from previous literature.

2.1 VALORANT

VALORANT is a 5-versus-5 character-based tactical first-person shooter developed by Riot Games [19]. A match of *VALORANT* consists of two 5-member teams: the attackers and defenders. The goal of the attackers is to plant the spike and let it detonate, which happens after a set amount of time, or to eliminate all the defenders. The goal of the defenders is to stop the spike from being planted, defuse the spike before it detonates, or to eliminate all the attackers (assuming the spike has not been planted yet, else it must still be defused after eliminating all the attackers). Before the match starts and every player enters the playable map, they must select an agent. Every agent has different abilities and fulfil different roles, namely: Duelist, Initiator, Sentinel, and Controller. The different abilities and roles impact the team composition, which affects how a team plays. A round of *VALORANT* consists of three phases: Pre-round, round, and post-round. The pre-round phase is a short phase where players can devise an approach to the round and spend credits on weapons, armour, and abilities. The round itself lasts 100 seconds and may be extended by 45 seconds by planting the spike. Due to how short a round is, the game is very fast-paced. Every second matters, and teamwork and communication are very important. The post-round phase is a very short phase where players that are still alive may pick up weapons dropped by other players. Other factors and elements impact the dynamics of the game as well, such as ultimate, utility, and economy management. A team has won the match if it is the first to win 13 rounds (or the first to win both overtime rounds in case the match goes into overtime) [20].

Due to how fast-paced and dynamic the game is, and how well-established the esports scene of the game is, it is well-suited for this research project [21]. At the highest competitive level of the game, where every player possesses similar technical skills, team dynamics and communication make the difference [22], [23]. This makes it especially interesting to investigate team dynamics within high-level competitive *VALORANT* teams.

2.2 Analysing Data in Online Games

1) Analysing in-game data: In order to create interpersonal skill profiles for every player, voice and in-game data, such as data about a match or the rounds played within a match, are required. The former can be directly collected from a video game if no third-party voice chat application is used, but due to its personal nature, it might raise privacy concerns. The latter can be obtained non-intrusively, as the data can often be directly collected from a video game or its application public interface (API) [24]. Nascimento Junior *et al.* solely used the API provided by game developer Riot Games to identify and characterise team behaviour based on historical matches from online game *League of Legends* [24]. The authors used K-means clustering to cluster teams' performance and investigated for each cluster how and to what extent these features had an influence on the success and failure of the teams [25]. A similar approach could potentially be used to cluster players, i.e., create interpersonal skill profiles that are able to differentiate between players. Tan *et al.* researched how short-lived ad hoc teams of strangers communicate and investigated its effect on team performance using online game *Portal 2*. In their research, they only analysed text messages extracted from the game but stated that the lack of in-game communications and voice chat in their analysis is a limitation. By including this data, a more comprehensive understanding of the relationship between communication and performance could be gained [26].

2) Analysing voice data: Voice chat is a feature that is incorporated in many online games and is an integral part of competitive team-based online games due to how important teamwork and communication are [27], [28]. Hence, the inclusion of voice data is essential to create interpersonal skill profiles. Reid *et al.* presented a modelling approach in their research that uses features derived from verbal communication data and minimal game metadata from the online game *Overwatch* to predict if a match was toxic or not. The minimal game metadata they used consisted of three features which were directly available in the game client. These features included the match outcome, the match rank, and the number of players in the party of the player whose data were collected. The authors evaluated three classification models, namely random forest (RF), support vector machine (SVM), and logistic regression, but state that other approaches that work directly with audio features, such as self-attentive convolutional neural networks (CNNs), may be promising for future work [29]. Even in the case of a model that is almost fully reliant in voice data, minimal game metadata is still collected and used. This points towards the usage of both voice and other in-game data, such as match metadata and player performance data, for modelling purposes. Other studies seem to have a similar approach, and some studies even include self-report data such as data obtained via questionnaires. For example, Tan *et al.* investigated team communication as a potential measure in a mixed methods study in the online game *League of Legends*. In their study, they used a combination of behavioural (observable behaviours via team communication) data, in-game data, and self-report data. They conducted an in-depth exploratory qualitative analysis and found which communication sequences low and high cohesion are associated with respectively [17]. Players' observable behaviours via team communication and patterns in their communication sequences can contribute to how they should be profiled. Another example is a study conducted by Frommel *et* *al.* In their study, they collected audio, video (webcam), in-game, and self-report data, from which they extracted 75 features to use in their RF and SVM models. With these models, they tried to predict affiliation between dyadic strangers, facilitated through their social interactions in a collaborative online game called *Labyrinth*. One of their findings is that a player's interpersonal affiliation towards a co-player in an online game setting can be predicted from unobtrusively gathered behavioural data [30]. This interpersonal affiliation towards a co-player could be an interesting feature to look at for the interpersonal skill profiles. Furthermore, their findings show that this could potentially be predicted by a model using data that is non-intrusively collected from a game's API. This means that the assignment of interpersonal skill profiles could perhaps be fully automated.

Based on previous research done in the field and recommendations from other authors, a combination of voice and other in-game data, and maybe even self-report data, should be collected to create interpersonal skill profiles. A model that might be suitable for handling this data is a (self-attentive) CNN. It is also critical to know how these interpersonal skill profiles should be defined, e.g., what metrics and factors should be used or considered. For that, it is important to look at team dynamics in teams in competitive team-based online games but also team dynamics in real-world high-pressure teams, such as military, aviation, emergency, and surgical teams, due to the similarities these two share [17].

2.3 Team Dynamics in Online Games

1) Team dynamics: Team dynamics describe the interpersonal interactions, communication patterns, and behaviours exhibited among members of a team, and they are a key component of the interpersonal skill profiles. Furthermore, they play an important role in the success of a team. At the highest competitive level, where competitors possess similar technical skills (often referred to as mechanics or micro gameplay), which encompass aim, movement, and ability usage in *VALORANT*, achieving optimal performance can potentially be attributed to team dynamics and communication [22],

[23]. The latter plays a big role in the macro gameplay, which refers to higher level elements such as strategic considerations, ultimate/utility/economy management, and map awareness. It is important to communicate and make correct decisions based on the available information [23]. Wanyi provided an overview of studies concerning the impact of team dynamics on performance in the esports industry. In her paper, she identified the characteristics of successful esports teams. These characteristics include the variability of the team composition, team communication, team cohesion, and trust in other team members [31]–[33]. The characteristic that is the easiest to measure, given the aforementioned types of in-game data, is team communication. Therefore, the study will put more focus on this characteristic.

2) Non-verbal communication: Team communication in competitive team-based online games consists of two types of communication: Non-verbal communication and verbal communication. The former often consists of text chat, pings (alerts that are easy to active and provide auditory and visual cues for team members), pre-defined messages and responses which a player can easily select, and annotations [17], [34], [35]. Wuertz *et al.* found that players of online game *Dota 2* view pings an essential tool for planning and coordination [35]. Unfortunately, the API provided by Riot Games for *VALORANT* does not provide data regarding text messages, pings, and pre-defined messages and responses [24]. Furthermore, at the time of writing, unlike most other esports titles, *VALORANT* does not have an official match demo or replay system implementation, which makes collecting non-verbal communication data impossible [36]. Thus, a focus will be put on verbal communication.

3) Verbal communication: The second type of team communication is verbal communication. Several studies have shown that verbal communication has a positive impact on a team or online community [28], [31], [33]. Williams *et al.* performed a controlled field experiment where they introduced verbal communication in an existing online community in online game *World of Warcraft*. By comparing data collected from a community that used both verbal and non-verbal communication to data collected from a community that kept using non-verbal communication only, the authors found that liking and trust among the community members increased due to the addition of voice chat [37]. These findings are in line with the findings of De Jong *et al.* [33]. Overall, these studies put further emphasis on the importance of verbal communication within a team environment. In a game such as *VALORANT*, verbal communication can roughly be divided into two main categories: Call-outs and strategies. Call-outs are short tactical phrases that occur during a round in the game. These phrases are brief, highly spatially and temporally contextual, and encoded with game state information. Furthermore, it is not uncommon to repeat call-outs to stress the urgency of the information, as information is often outdated after a few seconds due to the high-paced nature of the game. Call-outs are important because every member in a team has their own view into a 3-dimensional world. It is therefore important to convey information to team members which they are not able to gather from their own point of view [38]. Strategies are longer segments of verbal communication that can occur during any game phase and encompass a tactic, game plan, or idea. Often, a certain tactic or game plan is devised during the pre-round phase based on the available information, resources, and potentially even predictions (also known as reads) based on how the enemy team has behaved in previous rounds or even games. During the round, the team might adjust their tactic or game plan based on events that have already occurred during the round, newly obtained information, e.g., via call-outs, and potential reads. This is also known as mid-rounding [39]. These strategic verbal communication segments play a large role in the previously mentioned macro gameplay [23].

In the context of this research paper, understanding the logic behind call-outs and strategies, and how they impact team dynamics is essential. Every player is different. One player might give numerous precise call-outs and devise a game plan every round, but another player might give less call-outs and use the verbal communications channel as a medium to vent their frustration instead. The differences in how both verbal and non-verbal communication channels are being used and how players react and reply to their team members may help define part of the interpersonal skill profiles.

2.4 Team Dynamics in High-Pressure Teams

1) Non-technical skills and behavioural marker systems: In other fields, such as the aviation industry, emphasis on non-technical skills has increased over the past few decades [15]. In the late 1990s, *NOTECHS* was introduced as a behavioural marker system to evaluate the non-technical skills of pilots [16]. Over the next few years, the system, or a variety thereof, became widely adapted by several airlines [15], [40]. The *NOTECHS* system consists of four categories: Co-operation, Leadership and Managerial Skills, Situation Awareness, and Decision Making. Every category is subdivided into elements and behavioural markers. The former are related to and derived from the categories they are a part of. The latter are examples of good and poor practice and belong to an element. Given these elements and behavioural markers, a domain expert can assess the non-technical skills of a pilot [15]. It is worth mentioning that communication is not a separate category in the *NOTECHS* system. According to Birch *et al.*, this is because communication skills are inherent in all four categories and all the listed behaviours in the system include communication [41]. Ceschi *et al.* presented a system based on *NOTECHS* called *NOTECHS+*. It has two additional categories (Resilience and Emotion Regulation) and was developed for a sector that comprises both aviation and emergency personnel: the Helicopter Emergency Medical Service (HEMS). The system offers a systematic approach for the assessment of non-technical skills of professionals involved in the aviation and emergency sectors. According to the authors, their system is also applicable in other fields where unforeseen events might happen regularly and where it is important to be trained to cope effectively and maintain situational awareness under stressful conditions. Examples include the military and healthcare [42]. To a lesser extent, this might be applicable to teams in games such as *VALORANT*. Sometimes, despite playing perfectly on the macro level, a team might still lose a round due to bad luck or other factors outside its control. When such unfortunate events occur more frequently, a player or team might experience tilt. This term is often used to describe a phenomenon associated with frustration and a subsequent deterioration in performance [43]. In order to succeed at the highest competitive level, it is important to stay strong mentally and overcome tilt [22]. Therefore, the additional categories of *NOTECHS+* might contribute to a component in the interpersonal skill profiles. Neither *NOTECHS* nor *NOTECHS+* have communication as a separate category; however, the same does not hold for all behavioural marker systems. Hamlet *et al.* constructed two behavioural marker systems: HeliNOTS (O) for offshore transport pilots and HeliNOTS (SAR) for search and rescue crews; each with domain-specific behavioural markers. All focus groups in the study reported a strong preference for a standalone communication category due to the importance of this skill to helicopter crews. Similar findings were reported by Yule *et al.* and Irwin *et al.* Yule *et al.* developed a behavioural marker system for surgeons called *Non-Technical Skills for Surgeon (NOTSS)* which includes four categories: Situation Awareness, Decision Making, Communication, and Teamwork and Leadership. Similar to *NOTECHS*, each category is subdivided into elements [45], [47]. Irwin *et al.* developed a behavioural marker system for farmers called *FLINTS*. Its prototype taxonomy consists of five categories (Situation Awareness, Decision Making, Teamwork and Communication, Leadership, and Task Management) which are subdivided into elements [46]. Overall, a large overlap in categories can be noticed between the behavioural marker systems across different fields, with the main difference being the inclusion or exclusion of a separate communication category. Since communication information gained from both verbal and non-verbal communication channels is essential in this research project, it is important to look at the role communication plays in the aforementioned behavioural marker systems and how this can be properly modelled in the interpersonal skill profiles.

2) Team cognition: Team cognition encompasses the organised structures that support the members of a team to acquire, distribute, store, and retrieve critical knowledge [48], [49]. It plays an important role in both esports and real-world high-pressure teams. Cooke *et al.* analysed team cognition in experienced versus inexperienced command-and-control teams and found some key differences. When focusing on team communication and coordination, which is more applicable to this research, they found that experienced teams have a smaller coordination/communication ratio. This ratio can be defined as the number of coordination events divided by the number of total communication events through voice chat. According to the authors, a smaller ratio might suggest that experienced teams have more efficient communication, freeing up cognitive resources for off-task communication. Furthermore, it might suggest that experienced teams interact and think about interactions differently [50]. Therefore, it is also worth considering the number of call-outs, the number of total voice chat messages sent, and the ratio between these two as potential metrics in the interpersonal skill profiles.

Research done in other areas that share similarities with the research area of this work can give a large amount of insight and ideas about the direction of this research project. behavioural marker systems such as *NOTECHS* require expert raters to manually evaluate the categories of the systems [15], [16]. Furthermore, these experts have access to more data and resources than just voice communication data and their equivalent of other in-game data. In contrast to these systems, this research project aims to automate the evaluation of the interpersonal skill profile categories. Due to the differences in available data compared to the other fields, communication and other in-game data will have to be mapped against categories and metrics such as decision making, teamwork, and leadership. It will also be essential to properly define the metrics and factors that will make up the interpersonal skill profiles.

2.5 Research Motivation

Based on the research gaps described in the previous sections, this research project aims to define interpersonal skill profiles, which are inspired by behavioural marker systems employed in other research fields. Due to the similarities between teams in competitive team-based online games and real-world high-pressure teams, it is possible to apply some findings in this research project [17]. As mentioned before, data availability might pose a problem, but solving this problem would mean that perhaps in the future, assessment using behavioural marker systems in other fields could be automated as well, or the assignment of interpersonal skill profiles to high-pressure team members might also become possible. Furthermore, due to the abundance of potential data available to researchers in this field, it makes for a great test bed to study team dynamics and team cognition. This project aims to achieve the following research goals:

RG1: Define interpersonal skill profiles in competitive team-based online games based on team dynamics research done in related fields.

RG2: Build and evaluate a model that takes voice and other in-game data, such as game metadata, and constructs interpersonal skill profiles.

3. Methodology

Inspired by previous literature, voice and other in-game data had been collected for this study. Firstly, participants were recruited whose voice and in-game data were collected. Then, the interpersonal skill profiles were defined based on categories used in behavioural marker systems employed in other fields, communication metrics, and other in-game data (RG1). Lastly, a mapping was created from the collected data to interpersonal skill profile scores. This process was then automated using an approach that first labels input data and then calculates the interpersonal skill profile scores. In order to automate the first part of this approach, labels were defined, and input data were manually labelled in order to train an automatic labelling model. After the model had been fine-tuned and trained, the entire approach could be automated. Lastly, the results of the model were evaluated to prove that it works (RG2).

3.1 Data Collection

The data for this study were gathered through high-level *VALORANT* matches recorded using a tool provided by Patterns GG. The tool collects voice chat data, in-game data, and gameplay footage. Furthermore, it transcribes the voice data to text and adds timestamps to the text segments. The transcribed data and the in-game data are saved in *JSON* files, which allows easy accessibility. The transcription is done by OpenAI's *Whisper* model [51]. A local instance of the *Whisper* model was run on Pattern GG's servers to ensure the input data would not get shared with external parties. This is essential due to the privacy-sensitive nature of voice data.

The study had to pass the Ethics and Privacy scan of the institution first before the participant recruitment phase could start. The approved application can be viewed in Appendix A. Participants were recruited via various means and were recruited in teams of five with optional substitute slots. Participation requirements were as follows: participants should at least be 18 years old and their *VALORANT* rank should be at least *Diamond 1*. Furthermore, participating teams were required to communicate in English. During the recruitment phase, two teams were recruited. One team consisted of five high-ranked players and two substitutes. The other team consisted of five high-ranked players, a coach, and two substitutes, that compete together in the *VALORANT Game Changers* open circuit, which is part of Riot Games' esports circuit exclusively for marginalised genders [52]. Both teams were asked to play a minimum of ten matches to collect game and voice data for the study. Before all the participants confirmed their participation, they had given their consent and knew that their data would be used for a search project. Furthermore, they were made aware of the fact that they could back out of the study at any time they wanted.

The first team played a total of fourteen matches and the second team played a total of ten matches. After the voice data of these matches had been transcribed, they were manually cleaned to assure correctness. Unfortunately, due to time constraints, errors encountered by the participants while using the tool, and the low accuracy of the *Whisper* model, only five matches per team were fully cleaned. Then, the cleaned data could be manually labelled, after which a model would map the labelled data to interpersonal skill profile components, but first, the interpersonal skill profiles had to be defined.

3.2 Interpersonal Skill Profiles

Interpersonal skill profiles take inspiration from behavioural marker systems such as *NOTECHS*, but it is not possible to fully copy the categories used in those systems due to the differences in main objectives and how the teams operate. Therefore, only two categories from *NOTECHS* were adapted for the interpersonal skill profiles, namely *Leadership and Managerial Skills*, and *Situation Awareness* [16]. These two categories contain elements and behaviours that could be adapted into *VALORANT* equivalents and could also be measured given the available input data. *Resilience and Emotion Regulation*, as introduced by Ceschi *et al.* in their system *NOTECHS+*, also became a part of the interpersonal skill profiles due to the importance of mental resilience at the highest level of *VALORANT* [22], [42]. Unfortunately, the system lacks behavioural markers in the style of the original *NOTECHS*. Therefore, the interpersonal skill profile components also took inspiration from other literature [17], [53]. A more in-depth explanation will be given in Section 3.3. Lastly, *Communication* was added as a category due to its overall importance. To emphasise the focus on task-related communication, the category was renamed to *Communication Utility*. The communication metrics that were measured in this project are quantitative and are based on the findings of Cooke *et al.* [50]. All the adapted components can be viewed in Appendix B.

3.3 Mapping Data to Interpersonal Skill Profiles

After defining the interpersonal skill profiles, an approach that takes the input data and generates an interpersonal skill profile could be defined. This approach consisted of two parts:

1. A model that takes the input data and automatically labels the data accordingly.
2. An algorithm that takes the labelled data and calculates the interpersonal skill profile scores.

In order to create a model that labels the input data correctly, the labels have to be properly defined first. These labels are based on the components shown in the previous section, definitions of game or domain-specific terms, descriptors from the coding scheme of Zijlstra *et al.*, and descriptors added to the aforementioned coding scheme by Tan *et al.* [17], [38], [53]. The labels can be seen in Table 3.1.

Label	Description	Examples
Call-out: Directive	Telling the team or a specific team member to do something.	"Peek him, peek him, peek him."; "Watch out."
Call-out: About self	Notifying the team about their own status.	"I smoke."; "Darting backsite."
Call-out: About enemy	Notifying the team about the enemy team's status.	"There's two site."; "Omen vents."; "Guys, they have eco."
Call-out: Environment	Notifying the team about the environment.	"There's a wall."; "Market door shot."
Strategy	One or multiple communication segments that encompass a tactic, plan, or idea.	"I think we play Cypher A maybe and we fight B main."
Question/Inquiry	Requesting for information or asking for confirmation. Questions do not contain recommendations and are game-related.	"Do you want to fight them?"; "Where is Brim?"; "Was the spike mid or B?"
Answer/Confirmation	Any response to a question. Includes "yes/no" responses as confirmation to questions.	"Okay, sure."; "Yep."; "Okay."; "Yeah.", "No, I couldn't."
Encouragement	Statements that encourage, celebrate, or comfort the team or a team member.	"Nice, nice shot."; "Nice try."; "Nice, insane guys."
Anger and Frustration	Statements that exhibit feelings of anger or frustration. These can be directed towards the team, teammate, enemy, or self.	"Oh, are you kidding me?"; "I can't, this game is so bad."; "What the hell."
Apologies and Remorse	A remark that expresses sorrow or regret for prior action.	"My bad, my bad."; "Wait, I'm sorry."; "Sorry for the flash."
Miscellaneous	All segments that do not fall under the categories listed above.	"Tum tum tum tum tum."; "Whack-a-mole shit."

Table 3.1: Table containing the labels, their descriptions, and various examples taken from the collected data. Game-specific terms such as agent names or location names of areas in the playable maps are underlined.

3.3.1 Automatic Labelling Model

After the labels had been defined, they were manually assigned to every text segment in the ten matches that were cleaned during the data collection part of the study. This process led to a grand total of 10,880 manually labelled text segments. After removing duplicates, 8,384 unique text segments remained. Other than cleaning the data, common pre-processing steps in natural language processing, such as lemmatisation and the conversion of uppercase to lowercase, were not taken. This is due to the fact that some important information might be lost in this process. For example, "smoke B" and "smokes B" have very different meanings. Furthermore, there is a large difference between 'a' and 'A' as the latter can refer to a site in *VALORANT*.

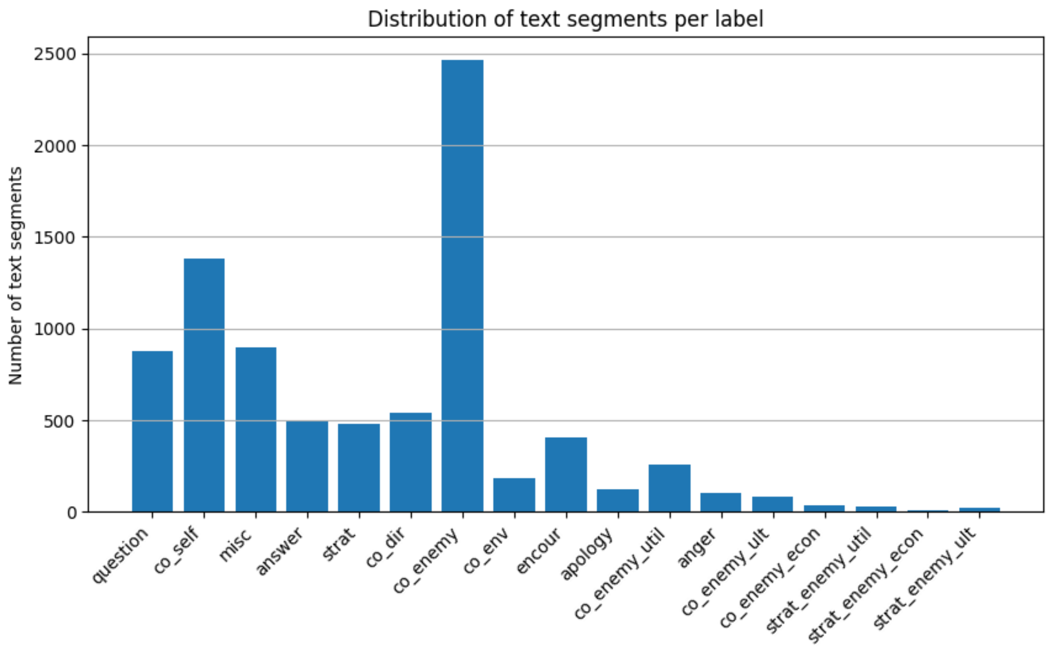

Figure 3.1: Bar chart showing the distribution of text segments per label in the data set. The data set consists of 8,384 unique text segments and 17 unique labels.

Data set analysis showed that the data set was very imbalanced, as can be seen in Figure 3.1. 2,468 text segments (29.44% of the data set) were labelled 'co_enemy', while only 9 text segments were labelled 'strat_enemy_econ'. Even though it makes sense that call-outs about the enemy occur most frequently in *VALORANT* matches, it would pose a challenge for the automatic labelling model. To combat this imbalance, balanced class weights were created using the *scikit-learn* library, which uses a heuristic inspired by King and Zeng [54], [55].

The automatic labelling model is an implementation of the *RoBERTa* model by Liu *et al.*, which is a transformer-based language model built on the *BERT* model by Devlin *et al.* [56], [57]. It is pre-trained on five large English-language corpora, which significantly reduced the time and computational power needed to train the automatic labelling model for the English text segments in the data set. The model was implemented in *Python* using the *Keras*, *scikit-learn*, *TensorFlow*, and *HuggingFace Transformers* libraries [55], [58]–[61]. Hyperparameter tuning was performed in order to maximise the model's performance. Several combinations of hyperparameter values were tested, with the best combination leading to a validation accuracy of 82.1%. The results of the hyperparameter tuning process are shown in Table 3.2. Some hyperparameters remained unchanged throughout the tuning process, namely the max token length (256), batch size (32), loss function (sparse categorical cross-entropy loss), and optimiser (Adam) [59], [62].

Test split	Epochs	Learning rate	Class weights	Accuracy
0.1	5	0.00001	Yes	77.8%
0.1	5	0.00001	No	80.2%
0.1	5	0.0001	Yes	30.6%
0.1	**5**	**0.0001**	**No**	**82.1%**
0.2	5	0.0001	Yes	70.0%
0.2	5	0.0001	No	77.6%
0.2	10	0.0001	Yes	30.1%
0.2	10	0.0001	No	79.3%
0.3	5	0.00001	Yes	74.8%
0.3	5	0.00001	No	78.5%
0.3	5	0.00005	Yes	69.3%
0.3	5	0.00005	No	78.1%
0.3	5	0.0001	Yes	63.1%
0.3	5	0.0001	No	79.4%
0.3	10	0.00001	Yes	77.3%
0.3	10	0.00001	No	79.1%

Table 3.2: Table containing the validation accuracy of the automatic labelling model along with the tested hyperparameter values. The best combination of hyperparameter values is highlighted in bold.

A wide variety of hyperparameter values were tested. During the tuning process, certain hyperparameter values were adjusted based on observations made during the training phase. For example, it quickly became apparent that training for 10 epochs was unnecessary considering how small the data set is. Most of the time, the model converged after 4 to 5 epochs. It also became apparent that the validation accuracy was higher on average when the test split was smaller. This might be due to the size of the data set. A 10-20% increase to the size of the training split could lead to a noticeable improvement in the model's performance. It is worth noting that the accuracy values shown in Table 3.2 were achieved on validation sets of the size indicated by the test split hyperparameter, e.g., a validation accuracy of 82.1% was achieved on a validation set of size 0.1 (839 text segments) after being trained on a training set of size 0.9 (7,545 text segments). The models that utilised class weights always performed worse than its counterpart with no class weights, which was unexpected. Perhaps, using balanced class weights was not the most effective way to combat the imbalance in the data set. An alternative approach, namely stratified K-Fold cross-validation, was considered but not implemented due to computational limitations. Lastly, two of the models had a validation accuracy of around 30%. This was due to the models converging in a local maximum, where the models would simply label every text segment 'co_enemy', which is the label that occurs the most in the data set.

3.3.2 Interpersonal Skill Profile Algorithm

The second part of the approach consists of an algorithm that takes the labelled data and creates the interpersonal skill profiles. In order to create such an algorithm, a link had to be established between the interpersonal skill profile components shown in Appendix B and the labelled data plus game metadata. This connection can be seen in Table 3.3.

Component	Metric
Number of first call-outs or strategies of the round	Number of ’Call-out’ (any type) or ’Strategy’ labelled text segments that occurred first in the round (meaning that at most one player on a team can get this in each round)
Number of mid-round strategies (mid-rounding)	Number of ’Strategy’ labelled text segments that occurred after the buy phase of a round
Number of encouraging voice segments	Number of ’Encouragement’ labelled text segments
Number of overall strategies and number of call-outs about themselves	Number of ’Strategy’ and ’Call-out: About self’ labelled text segments
Number of call-outs regarding enemy team’s ultimate, utility, and economy management	Number of ’Call-out: About enemy’ labelled text segments that also include terms such as ’ult’, ’eco’, and ’low buy’
Number of strategies based on enemy team’s ultimate, utility, and economy management	Number of ’Strategy’ labelled text segments that also include terms such as ’ult’, ’eco’, and ’low buy’
Number of voice segments indicating anger and frustration	Number of ’Anger and Frustration’ labelled text segments
Number of voice segments indicating remorse or an apology	Number of ’Apologies and Remorse’ labelled text segments
Number of call-outs	Number of ’Call-out’ (any type) labelled text segments
Number of strategies	Number of ’Strategy’ labelled text segments
Number of voice segments	Total number of labelled text segments

Table 3.3: Table containing the interpersonal skill profile components and their corresponding metrics.

The algorithm imports transcript *CSV* and *JSON* game metadata files and saves the metrics to a *Python* dictionary and the labels to a *pandas* DataFrame [58], [63], [64]. In order to calculate the scores for the four adapted categories using the saved metrics, they had to be defined first.
The *Leadership and Managerial Skills* score is defined as follows:

$$\begin{aligned}\text{LMS} = {} & \text{\#first call-outs and strategies of the round} + \\ & \text{\#strategies mid-round} + \alpha \times \text{\#encouraging voice segments} + \\ & \text{\#overall strategies} + \text{\#call-outs about themselves}\end{aligned}$$

where α is a weight value between 0 and 1. This weight value is assigned to the number of encouraging voice segments due to its lower importance and higher frequency compared to the other metrics. The score is clamped between 0 and 100 and scales linearly.
The *Situation Awareness* score is defined as follows:

$$\begin{aligned}\text{SIT} = {} & \alpha \times (\text{\#call-outs regarding enemy team's ult/util/econ management} + \\ & \text{\#strategies based on enemy team's ult/util/econ management})\end{aligned}$$

where α is a weight value larger than 1. This weight value is assigned to the score to scale it to a range between 0 and 100 due to the low occurrence count of the metrics in the score. Similarly to the *Leadership and Managerial Skills* score, the *Situation Awareness* score is clamped between 0 and 100 and scales linearly.
The *Resilience and Emotion Regulation* score is defined as follows:

$$\text{RES} = \frac{\text{\#voice segments indicating remorse or an apology}}{\text{\#voice segments indicating anger and frustration}}$$

The score is log-normalised to a range of 0 to 100 and scales logarithmically.

In the edge case where the number of voice segments indicating anger and frustration is 0, a score of 100 is assigned. The score is calculated in this manner because of some qualitative observations that were made during the data collection and cleaning process. It became apparent that players could roughly be assigned to three categories:

1. Players that get frustrated frequently and do not apologise (tilted players); these would score below 50.
2. Players that get frustrated but do apologise, maybe because they realise they are started to become tilted; these would score close to 50.
3. Players that do not get frustrated (frequently) but still take the blame if something goes wrong, even if they are not necessarily the culprit; these would score above 50.

These three categories become easily distinguishable when the score is calculated in this manner.
Lastly, the *Communication Utility* score is defined as follows:

$$\text{COM} = \frac{\text{\#call-outs} + \text{\#strategies} + \text{\#questions} + \text{\#answers}}{\text{\#voice segments}}$$

Similarly to the first two scores, this score is also clamped between 0 and 100 and scales linearly. It shows what percentage of a player's total communication was task-related.
All four scores are calculated by the algorithm, after which a visualisation can be made, either in the form of a radar chart that shows a player's interpersonal skills over one or multiple matches, or in the form of a line chart that shows how a player's scores differ on a per-match basis. These visualisations give an overview of a player's interpersonal skills and can be used to quickly compare players.

4. Results

In order to prove that the model works, the results of the model have to be evaluated (RG2). The results of the study can be divided into two sections corresponding with the two parts mentioned in the approach in Section 3.3. Firstly, the results of the automatic labelling model will be shown. Then, the results of the interpersonal skill profile algorithm will be shown.

4.1 Automatic Labelling Model

As briefly mentioned in Section 3.3.1, the automatic labelling model managed to reach a validation accuracy of 82.1% using the hyperparameter values shown in Table 4.1, meaning that it achieved an accuracy of 82.1% on a validation set containing 839 text segments after being trained on 7,545 text segments. The confusion matrix of the model can be observed in Figure 4.1. The confusion matrix shows the IDs of the 17 labels. The conversion table can be found in Table 4.2. Lastly, the precision, recall, F1 scores, and number of occurrences of each label in the validation set can be observed in Table 4.3.

Hyperparameter	Value
Max token length	256
Batch size	32
Epochs	5
Test split	0.1
Learning rate	0.0001
Loss function	Sparse categorical cross-entropy loss
Optimiser	Adam
Class weights	No

Table 4.1: Table containing the hyperparameter values of the automatic labelling model.

Label ID	Label name
0	question
1	co_self
2	misc
3	answer
4	strat
5	co_dir
6	co_enemy
7	co_env
8	encour
9	apology
10	co_enemy_util
11	anger
12	co_enemy_ult
13	co_enemy_econ
14	strat_enemy_util
15	strat_enemy_econ
16	strat_enemy_ult

Table 4.2: Table containing the IDs of the labels in the confusion matrix (Figure 4.1) along with their corresponding names.

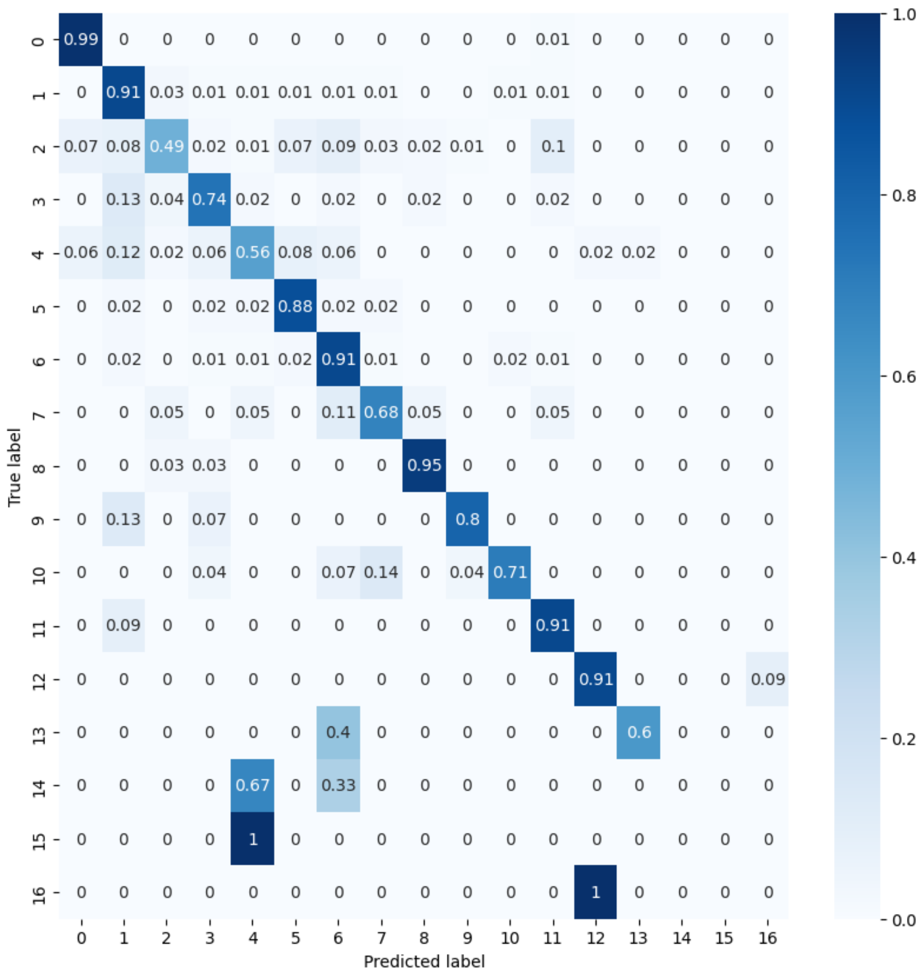

Figure 4.1: Confusion matrix of the automatic labelling model.

Label	Precision	Recall	F1 score	Occurrences
question	0.90	0.99	0.94	83
co_self	0.81	0.91	0.86	140
misc	0.81	0.49	0.62	89
answer	0.73	0.74	0.74	47
strat	0.74	0.56	0.64	50
co_dir	0.70	0.88	0.78	42
co_enemy	0.92	0.91	0.91	257
co_env	0.54	0.68	0.60	19
encour	0.90	0.95	0.92	37
apology	0.86	0.80	0.83	15
co_enemy_util	0.80	0.71	0.75	28
anger	0.40	0.91	0.56	11
co_enemy_ult	0.77	0.91	0.83	11
co_enemy_econ	0.75	0.60	0.67	5
strat_enemy_util	0.00	0.00	0.00	3
strat_enemy_econ	0.00	0.00	0.00	1
strat_enemy_ult	0.00	0.00	0.00	1
Accuracy			0.82	839

Table 4.3: Table containing the precision, recall, F1 scores, and the number of occurrences of each label in the validation set. At the bottom, the validation accuracy and the total number of labels in the validation set are shown.

4.2 Interpersonal Skill Profiles Algorithm

As mentioned in Section 3.3.2, the algorithm is able to calculate the interpersonal skill profile scores of a player. As the approach taken in this study is novel, there is a lack of relevant literature to compare it to. Therefore, a different approach was taken to show that the algorithm works. As mentioned in Section 3.1, data was collected from two teams. Even though the players and substitutes from both teams are of similar ranks, indicating that they have similar technical skills, the first team consists of players that play the game more casually while the second team is actively competing in the *VALORANT Game Changers* open circuit. The difference in mindsets between the teams should lead to a clear difference in the interpersonal skill profiles as well. Figures 4.2 and 4.3 show the interpersonal skill profiles of all the players and substitutes of both teams. For a clearer overview, the exact scores can be found in Table 4.4 and the scores on a per-match basis per player can be viewed in Appendix C. Furthermore, these per-match basis scores were used to calculate the mean interpersonal skill profile scores of both teams, which can be observed in Figure 4.4.

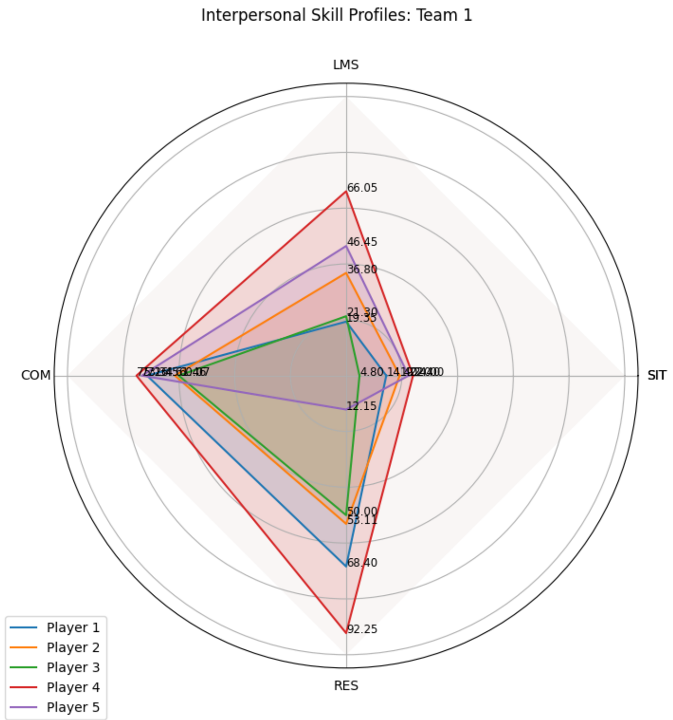

Figure 4.2: Radar chart showing the interpersonal skill profiles of the players of team 1.

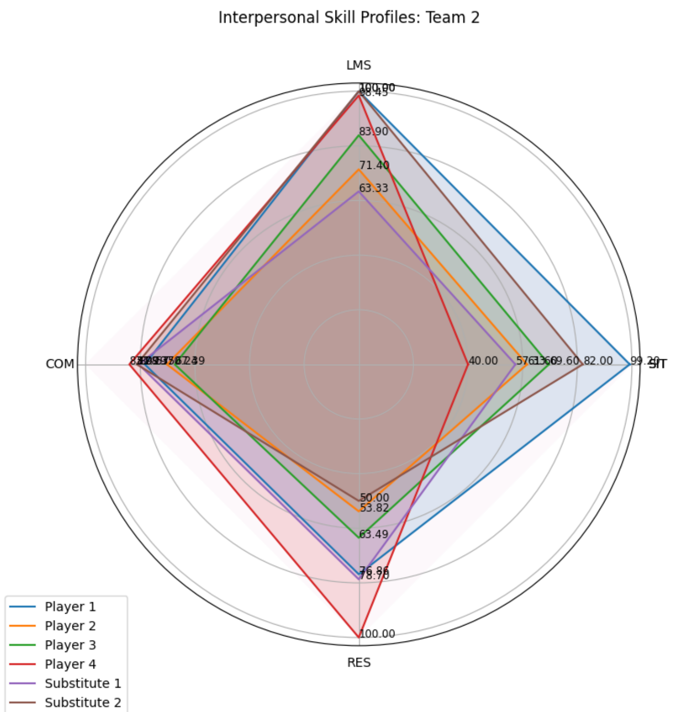

Figure 4.3: Radar chart showing the interpersonal skill profiles of the players and substitutes of team 2.

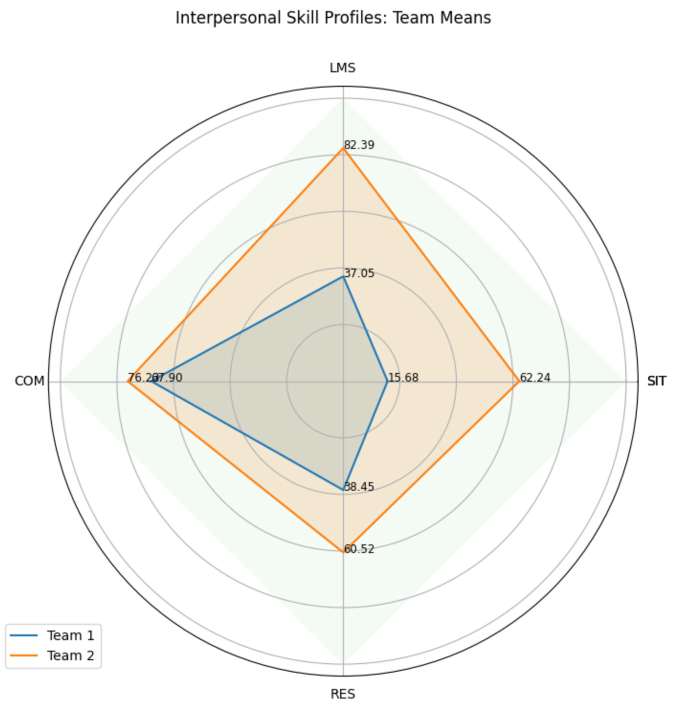

Figure 4.4: Radar chart showing the mean interpersonal skill profiles of both teams.

Team	Player	Matches played	LMS	SIT	RES	COM
1	Player 1	5	19.35	14.40	68.40	71.45
1	Player 2	5	36.80	19.20	53.11	61.46
1	Player 3	5	21.30	4.80	50.00	60.07
1	Player 4	5	66.05	24.00	92.25	75.26
1	Player 5	5	46.45	22.40	12.15	73.34
2	Player 1	5	100.00	99.20	76.86	78.35
2	Player 2	5	71.40	61.60	53.82	70.24
2	Player 3	5	83.90	69.60	63.49	67.39
2	Player 4	5	98.45	40.00	100.00	83.99
2	Substitute 1	3	63.33	57.33	78.70	80.57
2	Substitute 2	2	100.00	82.00	50.00	81.29

Table 4.4: Table containing all players and substitutes from both teams along with the number of matches played and their interpersonal skill profile scores.

At a quick glance, there are notable differences between the interpersonal skill profiles of team 1 and 2, but it had to be proved formally as well. This was done by conducting a statistical test on the four interpersonal skill profile scores. In order to choose the correct statistical test to conduct, several assumptions had to be checked first [65].

Firstly, the observations must be independent of each other. This is the case as the interpersonal skill profile scores of one player do not impact the scores of another player, although some scores are calculated using various labels which may have an impact on other labels, e.g., more 'question' text segments from fellow players will lead to more 'answer' text segments. The impact of this is not very large though, as call-outs still take up most of the communication within a team as can be observed in Figure 3.1, and whether a player makes a call-out or not is not dependent on other players. Secondly, the data were checked if they meet the assumption of normality. The Shapiro-Wilk test showed that the distributions of LMS ($W = 0.98, p = .45$), SIT ($W = 0.97, p = .28$), and RES scores ($W = 0.96, p = .06$) met the assumption of normality. On the other hand, the test showed that the distribution of COM scores ($W = 0.94, p = 0.02$) did show evidence of non-normality [66]. Lastly, the data were checked if they meet the assumption of homogeneity of variance. Levene's test showed that there was homogeneity of variances between the LMS ($F(1, 48) = 1.00, p = .32$), RES ($F(1, 48) = 1.36, p = .25$), and COM scores ($F(1, 48) = 0.26, p = .61$), but also indicated unequal variances for the SIT scores ($F(1, 48) = 12.53, p < .001$) [67].

Based on the results of the assumption tests, an analysis of variance (ANOVA) was conducted for the LMS and RES scores, Welch's ANOVA for the SIT scores, and the Kruskal-Wallis test for the COM scores [65], [68]–[71]. An ANOVA showed that the differences in LMS ($F(1, 48) = 71.11, p < .001, \eta^2 = 0.60$) and RES scores ($F(1, 48) = 8.95, p = .004, \eta^2 = 0.16$) between the two teams were significant. Welch's ANOVA showed that the difference in SIT scores between the two teams was significant as well ($F(1, 48) = 91.96, p < .001, \eta^2 G = 0.66$). Lastly, the Kruskal-Wallis test showed that the difference in COM scores between the two teams was also significant ($H(1) = 13.38, p < .001, \eta^2 = 0.26$). The results of all the statistical tests can be found in Tables 4.5, 4.6, and 4.7.

Score	Shapiro-Wilk test	Levene's test	ANOVA
LMS	$W = 0.98, p = .45$	$F(1, 48) = 1.00, p = .32$	$\boldsymbol{F(1, 48) = 71.11, p < .001, \eta^2 = 0.60}$
RES	$W = 0.96, p = .06$	$F(1, 48) = 1.36, p = .25$	$\boldsymbol{F(1, 48) = 8.95, p = .004, \eta^2 = 0.16}$

Table 4.5: Table containing the results of the statistical tests comparing the LMS and RES scores between team 1 (LMS: $M = 37.05, SD = 20.52$; RES: $M = 38.45, SD = 27.62$) and team 2 (LMS: $M = 82.39, SD = 16.52$; RES: $M = 60.52, SD = 23.30$).

Score	Shapiro-Wilk test	Levene's test	Welch's ANOVA
SIT	$W = 0.97, p = .28$	$F(1, 48) = 12.53, p < .001$	$\boldsymbol{F(1, 48) = 91.96, p < .001, \eta^2 G = 0.66}$

Table 4.6: Table containing the results of the statistical tests comparing the SIT scores between team 1 ($M = 15.68, SD = 8.97$) and team 2 ($M = 62.24, SD = 22.03$).

Score	Shapiro-Wilk test	Levene's test	Kruskal-Wallis test
COM	$W = 0.94, p = .02$	$F(1, 48) = 0.26, p = .61$	$\boldsymbol{H(1) = 13.38, p < .001, \eta^2 = 0.26}$

Table 4.7: Table containing the results of the statistical tests comparing the COM scores between team 1 ($M = 67.90, SD = 7.21$) and team 2 ($M = 76.23, SD = 7.03$).

5. Discussion

In this section, the results of the study will be discussed and analysed to check if the research goals defined in Section 2.5 were achieved or not. Then, the implications of this study will be discussed. Lastly, the limitations of the study will be discussed.

5.1 Research Goals

RG1: Define interpersonal skill profiles in competitive team-based online games based on team dynamics research done in related fields.

In Section 3.2, interpersonal skill profiles were defined based on behavioural marker systems employed in other fields, such as *NOTECHS* and *NOTECHS+* [16], [42]. Then, in Section 3.3, this definition was further refined by defining labels based on behavioural marker system metrics, definitions of game or domain-specific terms, and descriptors from Zijlstra *et al.* and Tan *et al.* [17], [38], [53]. To complete the definition and to achieve the research goal, a concrete approach was defined which describes the implementation and automation of the generation of interpersonal skill profiles. Using the complete definition, the second research goal could be tackled.

RG2: Build and evaluate a model that takes voice and other in-game data, such as game metadata, and constructs interpersonal skill profiles.

Based on the definition of interpersonal skill profiles defined in the first research goal, an approach was implemented to automate the generation of interpersonal skill profiles. An automatic labelling model was implemented and scores were defined based on metrics and labels defined in the first research goal. Now, the results of the model have to be evaluated to prove that it works.

The automatic labelling model achieved a validation accuracy of 82.1% after the hyperparameters had been tuned as described in Section 3.3.1, but this number alone does not tell much. A few observations can be made when looking at the confusion matrix in Figure 4.1 and the classification report in Table 4.3. The model has a high recall for the 'question', 'co_self', 'encour', 'anger' and 'co_enemy_ult' labels. The precision is relatively high for these labels as well, except for 'anger', which is very low at 0.40. As can be seen in the confusion matrix, a relatively large group of 'misc' text segments (0.1) was incorrectly labelled as 'anger'. The model might not be able to tell these apart, as the 'misc' label is used for all text segments that do not fit in any other category. Thus, the definition of the 'misc' label is very broad and the model might not be able to pick that up considering the small size of the data set as well.
Furthermore, 'strat_enemy_util' gets incorrectly labelled as 'strat' and 'co_-enemy'. 'strat_enemy_econ' also gets incorrectly labelled as 'strat' and 'strat_enemy_ult' gets incorrectly labelled as 'co_enemy_ult'. This could be due to the low number of occurrences of the more specific strategy labels in the data set. The model might not be able to pick up the subtle differences between specific strategies and general strategies due to the lack of data. The model also does not have access to domain-specific vocabulary. The lack thereof in combination with the size of the data set might have led to the model not being able to tell apart more game-specific terms, call-outs, and strategies.
The recall of 'strat' is low (0.56) and a relatively large percentage of negative 'strat' predictions are predicted as 'co_self' (27.27%). This might be due to strategy text segments often containing a part where the person calling the strategy explains what they are contributing to the strategy. Take the following example from the data set: "Yeah, um I'm gonna do two smokes, I'm gonna flash this side when we push up." followed by "So then you can flash like the other side.". A strategy involving smokes and flashes is being called here, but as the caller is also telling the team what they are doing themselves in this strategy, the model might pick it up as a call-out about themselves.
A few other labels, namely 'answer', 'apology', and 'anger' also get mis-

labelled as 'co_self' relatively often. A possible explanation could be that some text segments that are labelled as the former three also contain first-person singular pronouns, such as "I". Take these three examples from the data set: "Yeah they will I think." (answer), "Oh my, he can't see me, I'm done." (anger), and "I'm done, I'm sorry." (apology). These types of text segments might be confusing the model, as it might just predict numerous text segments containing these pronouns as 'co_self'. This would also explain why the precision (0.81) is a bit lower than the recall (0.91).
Overall, the F1 scores of the labels are not bad, taking into account what is discussed above. It shows that the model is decent at telling apart the various labels, even though the data set is rather small and highly imbalanced. Given more input data, one can only imagine how much better the model would perform, which would benefit the automated process of generating interpersonal skill profiles as a whole.

The second part of the approach consists of an algorithm that calculates the interpersonal skill profile scores based on the labels assigned by the automatic labelling model. Figures 4.2 and 4.3 show the interpersonal skill profiles of the players and substitutes of the two teams that were recruited in the study. To prove that the interpersonal skill profiles actually hold any meaning and are able to tell apart players, e.g., casual players from players competing in the competitive circuit, statistical tests were performed to show that the differences in the interpersonal skill profiles of both teams were significant. As can be observed in Tables 4.5, 4.6, and 4.7, the differences in the four scores were statistically significant, showing that this approach of measuring non-technical skills in the field of competitive team-based online games has potential. This suggests that the second research goal was achieved.

5.2 Implications

All in all, the results of the novel approach taken in this study show that it has potential and might be worth expanding upon. The study has shown that approaches and frameworks commonly used in related fields can be adapted and applied to the esports industry as well. The approach is able to differentiate between players based on their non-technical skills, which could be a key factor for esports organisations who are looking to trial new players for their team, for example.

5.3 Limitations

It is also worth discussing the limitations of the study, which might also impact the validity and quality of the results. Firstly, the *Whisper* model performed very poorly. Some timestamps were off by up to roughly 30 seconds, some transcriptions were of bad quality, and some voice segments did not get transcribed at all. Although this issue was fixed by manually cleaning the data afterwards, this time-consuming process did impact the amount of data that could be cleaned for the study, which is the second limitation.
As can be observed in Appendix C, not all players had consistent interpersonal skill profile scores throughout the matches they played. Although players behave differently on a per-game basis, fluctuations should not be too large. These fluctuations are mostly noticeable in the *Resilience and Emotion Regulation* scores, although this is partly due to its logarithmic scaling, which is more sensitive to small changes in the measured metrics. If more data could have been collected and cleaned, the data might have been more consistent or might have shown underlying patterns that could be analysed.
Another limitation is the lack of variety in the data, as all the data were only collected from two teams (twelve players). Unfortunately, the participant recruitment process did not go as planned. At first, the goal was to recruit eight teams total and have them face each other in a round-robin tournament. This would create a competitive environment and allow the collec-

tion of a wide variety of data from various teams. In the end, only two teams signed up, which impacted how the approach would be evaluated.
The definitions of the interpersonal skill profiles, metrics, labels, and scores are heavily based on literature research and domain knowledge. This theoretical approach might not be the most optimal. For example, another approach would be to perform principal component analysis to define the scores instead.
Another point worth noting is that there was no second rater for the manual label assignment. Although the assignment of labels is based on domain knowledge and is mostly objective, some subjective cases are inevitable. The lack of inter-rater reliability tests affects the validity of the results.
The statistical tests comparing the scores of both teams were performed using an ANOVA, Welch's ANOVA, and the Kruskal-Wallis test. One could argue that the data were hierarchical due to occurrences of certain labels having an impact on the number of occurrences of other labels. For example, more 'question' text segments will lead to more 'answer' text segments, which indirectly impacts some scores as these are calculated using the labels. Furthermore, how the team around a player acts and communicates might impact how a player communicates. Therefore, a model that is able to deal with hierarchical data, such as a linear mixed model, would perhaps lead to more valid results.
Lastly, as briefly mentioned before, the automatic labelling model did not have access to a domain-specific dictionary, which might have negatively affected the results of the model. Even if such a dictionary were to be added, it would have to be updated over time as well as new agents, maps, and weapons are added to the game.

6. Conclusion

This study investigated team dynamics within teams that compete in competitive team-based online games. Interpersonal skill profiles that aim to differentiate between players based on their non-technical skills were defined based on approaches and methods used in other research fields. The profiles were implemented in two steps. The first step consisted of an automatic labelling model that labels text transcribed voice chat text segments, and the second step consisted of an algorithm that calculates the interpersonal skill profile scores. The results showed that the approach taken in the study has potential and that approaches and methods related to team dynamics commonly used in related fields can potentially be applied to the world of esports as well.

A. Ethics and Privacy Scan Application

A.1 Coordinator Approval

Coordinator COSC
coordinator.cosc@uu.nl

Reply | Forward | Archive | Junk | Delete | More

To Huang, J.K. (Jin Kai) <j.k.huang@students.uu.nl>

5-4-2024 01:07

Re: Status update regarding Ethics and Privacy Quick Scan

Dear Jin Kai,

Thank you for your message.

I contacted the Data Stewart concerning the result of your scan and his opinion, as well as mine, is that your thesis can continue as-is, without the need for a fuller privacy assessment.

Therefore I hereby approve your scan.

Good luck with your thesis. If you need any additional advice, please let us know.

Kind regards,
[redacted]

A.2 Submitted Application

Response Summary:

Section 1. Research projects involving human participants

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from Twitter, Reddit) without directly recruiting participants, please answer no.
Yes

Recruitment

P2. Does your project involve participants younger than 18 years of age?
No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?
No

P4. Is your project likely to involve participants engaging in illegal activities?
No

P5. Does your project involve patients?
No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?
No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?
No

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)
Yes

PC2. Will you tell participants that their participation is voluntary?
Yes

PC3. Will you obtain explicit consent for participation?
Yes

PC4. Will you obtain explicit consent for any sensor readings, eye tracking, photos, audio, and/or video recordings?
Yes

PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?
Yes

PC6. Will you give potential participants time to consider participation?
Yes

PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?
Yes

PC8. Does your project involve concealment or deliberate misleading of participants?
No

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of **personal data** (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?
Yes

High-risk data

DR1. Will you process personal data that would jeopardize the physical health or safety of individuals in the event of a personal data breach?
No

DR2. Will you combine, compare, or match personal data obtained from multiple sources, in a way that exceeds the reasonable expectations of the people whose data it is?
No

DR3. Will you use any personal data of children or vulnerable individuals for marketing, profiling, automated decision-making, or to offer online services to them?
No

DR4. Will you profile individuals on a large scale?
No

DR5. Will you systematically monitor individuals in a publicly accessible area on a large scale (or use the data of such monitoring)?
No

DR6. Will you use special category personal data, criminal offense personal data, or other sensitive personal data on a large scale?
No

DR7. Will you determine an individual's access to a product, service, opportunity, or benefit based on an automated decision or special category personal data?
No

DR8. Will you systematically and extensively monitor or profile individuals, with significant effects on them?
No

DR9. Will you use innovative technology to process sensitive personal data?
Yes

Privacy Warning. As high-risk data processing seems involved (yes to any of DR1-DR9), a fuller privacy assessment is required. Please provide more information on the DR1-DR9 questions with a yes here:
Voice data will be processed using OpenAI's Whisper to transcribe speech to text. To adhere to the university's ethics & privacy policy, a local instance of the model will be used so that none of the voice data, which is privacy sensitive data, will be shared with external parties.

Data minimization

DM1. Will you collect only personal data that is strictly necessary for the research?
Yes

DM4. Will you anonymize the data wherever possible?
Yes

DM5. Will you pseudonymize the data if you are not able to anonymize it, replacing personal details with an identifier, and keeping the key separate from the data set?
Yes

Using collaborators or contractors that process personal data securely

DC1. Will any organization external to Utrecht University be involved in processing personal data (e.g. for transcription, data analysis, data storage)?
Yes

DC2. Will this involve data that is not anonymized?
Yes

DC3. Are they capable of securely handling data?
Yes

DC4. Has been drawn up in a structured and generally agreed manner who is responsible for what concerning data in the collaboration?
Yes

DC5. Is a written contract covering this data processing in place for any organization which is not another university in a joint research project?
Yes

International personal data transfers

DI1. Will any personal data be transferred to another country (including to research collaborators in a joint project)?
Yes

DI2. Do all countries involved in this have an adequate data protection regime?
Yes

Fair use of personal data to recruit participants

DF1. Is personal data used to recruit participants?
No

Participants' data rights and privacy information

DP1. Will participants be provided with privacy information? (Recommended is to use as part of the information sheet: For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information at `www.uu.nl/en/organisation/privacy`**.)**
Yes

DP2. Will participants be aware of what their data is used for?
Yes

DP3. Can participants request that their personal data be deleted?
Yes

DP4. Can participants request that their personal data be rectified (in case it is incorrect)?
Yes

DP5. Can participants request access to their personal data?
Yes

DP6. Can participants request that personal data processing is restricted?
Yes

DP7. Will participants be subjected to automated decision-making based on their personal data with an impact on them beyond the research study to which they consented?
No

DP8. Will participants be aware of how long their data is being kept for, who it is being shared with, and any safeguards that apply in case of international sharing?

Yes

DP9. If data is provided by a third party, are people whose data is in the data set provided with (1) the privacy information and (2) what categories of data you will use?

Yes

Using data that you have not gathered directly from participants

DE1. Will you use any personal data that you have not gathered directly from participants (such as data from an existing data set, data gathered for you by a third party, data scraped from the internet)?

Yes

DE2. Will you use an existing dataset in your research?

Yes

DE3. Do you have permission to do so from the owners of the data set?

Yes

DE4. Have the people whose data is in the data set consented to their data being used by other researchers and/or for purposes other than that for which that data set was gathered?

Yes

DE5. Are there any contractual conditions attached to working with or storing the data from DE2?

Yes: NDA and WPA are signed already.

DE6. Does your project require access to personal data about participants from other parties (e.g., teachers, employers), databanks, or files?

No

DE9. Does the project involve collecting personal data from websites or social media (e.g., Facebook, Twitter, Reddit)?

No

Secure data storage

DS1. Will any data be stored (temporarily or permanently) anywhere other than on password-protected University authorized computers or servers?
Yes

DS2. Does this only involve data stored temporarily during a session with participants (e.g. data stored on a video/audio recorder/sensing device), which is immediately transferred (directly or with the use of an encrypted and password-protected data-carrier (such as a USB stick)) to a password-protected University authorized computer or server, and deleted from the data capture and data-carrier device immediately after transfer?
No

DS3. Does this only involve data stored with a collaborator or contractor?
Yes

DS4. Excluding (1) any international data transfers mentioned above and (2) any sharing of data with collaborators and contractors, will any personal data be stored, collected, or accessed from outside the EU?
No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

H1. Does your project give rise to a realistic risk to the national security of any country?
No

H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?
No

H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)
No

H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)
No

H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?
No

H6. Does your project give rise to a realistic risk of harm to the researchers?
No

H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?
No

H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?
No

H9. Is there a realistic risk of other types of negative externalities?
No

Section 4. Conflicts of interest

C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?
No

C2. Is there a direct hierarchical relationship between researchers and participants?
No

Section 5. Your information.

This last section collects data about you and your project so that we can register that you completed the Ethics and Privacy Quick Scan, sent you (and your supervisor/course coordinator) a summary of what you filled out, and follow up where a fuller ethics review and/or privacy assessment is needed. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information. Please see the guidance on the ICS Ethics and Privacy website on what happens on submission.

Z0. Which is your main department?
Information and Computing Science

Z1. Your full name:
Jin Kai Huang

Z2. Your email address:
`j.k.huang@students.uu.nl`

Z3. In what context will you conduct this research?
As a student for my master thesis, supervised by: Dr. Julian Frommel

Z5. Master programme for which you are doing the thesis
Computing Science

Z6. Email of the course coordinator or supervisor (so that we can inform them that you filled this out and provide them with a summary):
`j.frommel@uu.nl`

Z7. Email of the moderator (as provided by the coordinator of your thesis project):
`coordinator.cosc@uu.nl`

Z8. Title of the research project/study for which you filled out this Quick Scan:
Building Interpersonal Style Profiles for Players in Competitive Team-Based Games Based on Interactions Between and Communication With Other Players

Z9. Summary of what you intend to investigate and how you will investigate this (200 words max):
The study project investigates team dynamics within teams in competitive team-based online games. Due to the advancements regarding team dynamics and non-technical skills in other fields such as the military, aviation, and emergency services and due to the similarities between teams in competitive team-based online games and high-pressure teams in the aforementioned sectors, methods that are being used to evaluate non-technical skills in those sectors can be applied to teams in competitive team-based online games as well. This project aims to create so-called interpersonal skill profiles based on those methods and apply them to teams in esports. Furthermore, it aims to solve one of the biggest downsides of those methods in other sectors, which is the need for an expert rate who manually evaluates the non-technical skills. In order to achieve these goals, a large language model such as OpenAI's Whisper is needed to transcribe voice data. Afterwards, the data will be used to train a classification model to automate the evaluation process of non-technical skills.

Z10. In case you encountered warnings in the survey, does supervisor already have ethical approval for a research line that fully covers your project?
No

Scoring

- Privacy: 1
- Ethics: 0

B. Interpersonal Skill Profile Components

B.1 Table of Adapted Components (1/2)

Category	Element	Behaviour	Component
Leadership and Managerial Skills	Use of Authority/Assertiveness	Takes initiative to ensure involvement and task completion	Number of first call-outs or strategies of the round
		Takes command if situation requires	Number of mid-round strategies (mid-rounding)
		Motivates crew by appreciation and coaches when necessary	Number of encouraging voice segments
	Planning and Coordination	Encourages crew participation in planning and task completion	Number of encouraging voice segments
		Clearly states intentions and goals	Number of overall strategies and number of call-outs about themselves
Situation Awareness	Environmental Awareness	Collects information about the environment	Number of call-outs regarding enemy team's ultimate, utility, and economy management
		Shares information about the environment with others	Number of call-outs regarding enemy team's ultimate, utility, and economy management
	Anticipation	Identifies possible/future problems	Number of strategies based on enemy team's ultimate, utility, and economy management

Table B.1: Table containing the first half of the adapted interpersonal skill profile components. Where applicable, the elements and behaviours of *NOTECHS* categories are also shown [16].

B.2 Table of Adapted Components (2/2)

Category	Element	Behaviour	Component
Resilience and Emotion Regulation	N/A	N/A	Number of voice segments indicating anger and frustration
		N/A	Number of voice segments indicating remorse or an apology
Communication Utility	N/A	N/A	Number of call-outs
		N/A	Number of strategies
		N/A	Number of voice segments

Table B.2: Table containing the second half of the adapted interpersonal skill profile components. Where applicable, the elements and behaviours of *NOTECHS* categories are also shown [16].

C. Per-Match Basis Profile Scores

C.1 Per-Match Basis Profile Scores of Team 1

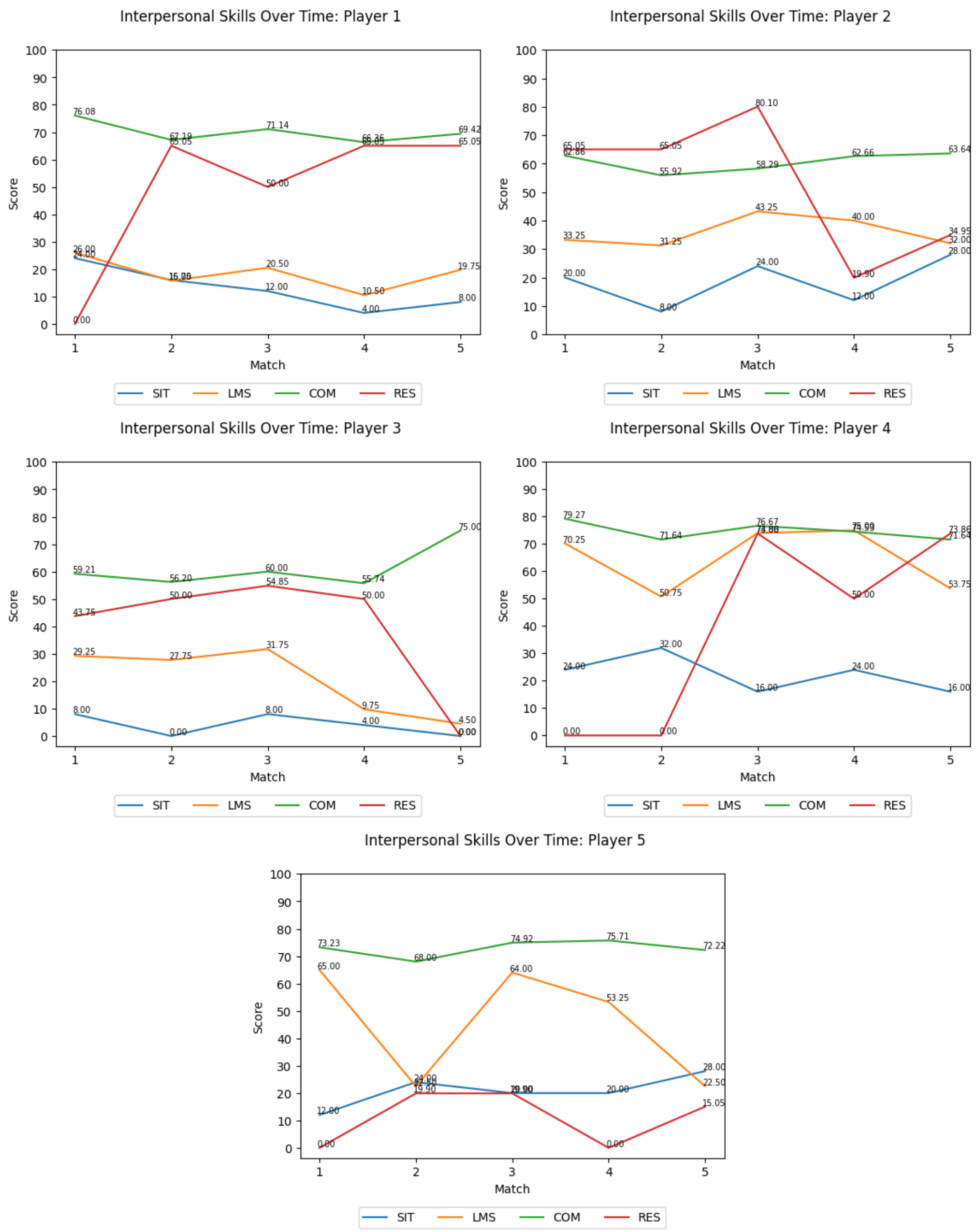

Figure C.1: Line charts showing the interpersonal skill profile scores of the players of team 1 on a per-match basis.

C.2 Per-Match Basis Profile Scores of Team 2

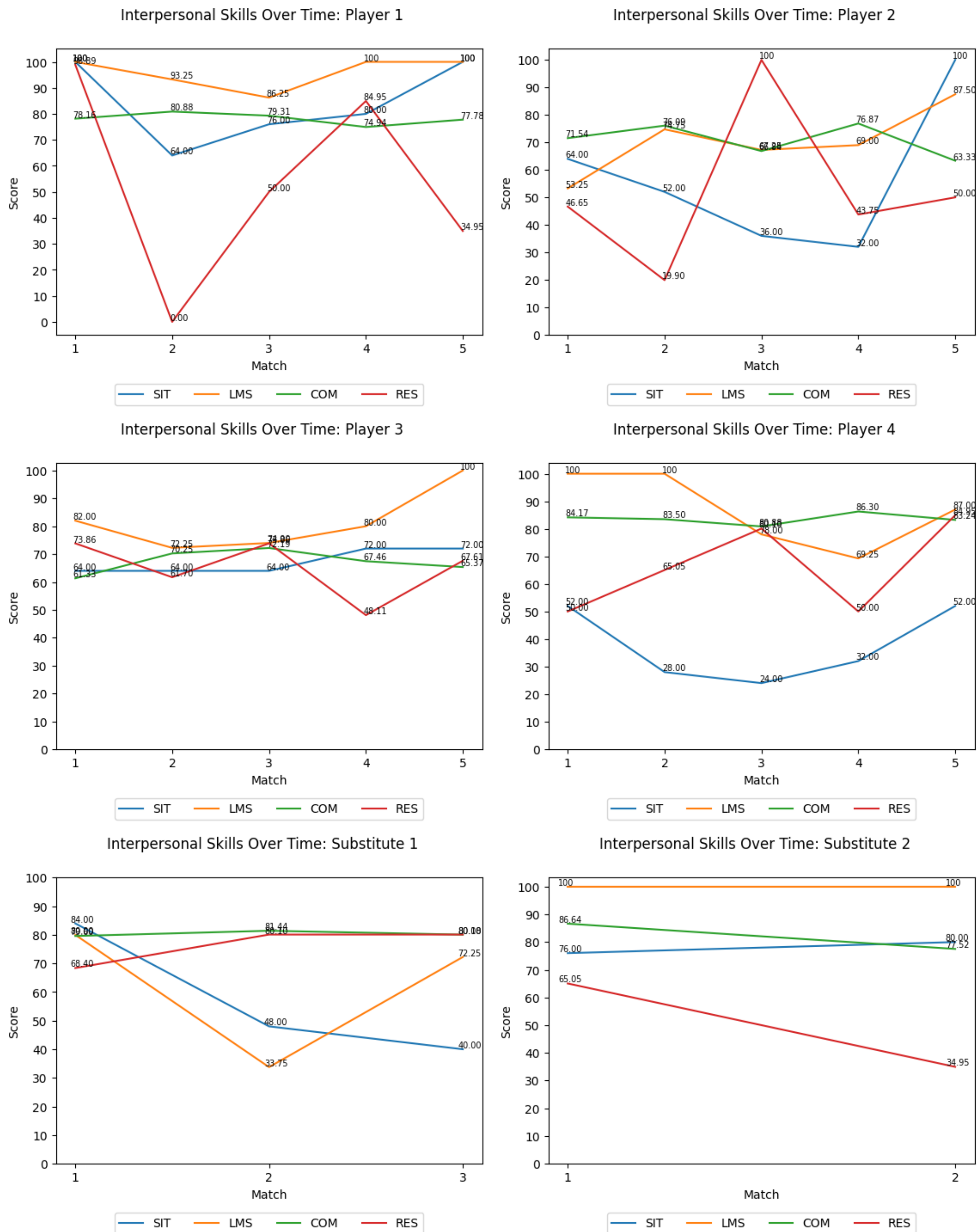

Figure C.2: Line charts showing the interpersonal skill profile scores of the players and substitutes of team 2 on a per-match basis.

Bibliography

[1] A. H. Olsen, *The evolution of esports: An analysis of its origin and a look at its prospective future growth as enhanced by information technology management tools*, 2015. arXiv: 1509.08795 [cs.CY].

[2] K. Lehnert, A. Walz, and R. Christianson, "The booming esports market: A field day for fans," *Journal of Business Strategy*, vol. 43, no. 2, pp. 122–128, Dec. 2020, ISSN: 0275-6668. DOI: 10.1108/jbs-07-2020-0159. [Online]. Available: http://dx.doi.org/10.1108/JBS-07-2020-0159.

[3] Block, Sebastian and Haack, Florian, "Esports: A new industry," *SHS Web of Conf.*, vol. 92, p. 04002, 2021. DOI: 10.1051/shsconf/20219204002. [Online]. Available: https://doi.org/10.1051/shsconf/20219204002.

[4] E. Mangeloja, "Economics of esports," *Electronic Journal of Business Ethics and Organization Studies*, vol. 24, no. 2, 2019.

[5] S. King, "How the FaZe CSGO Super-Team is Being Saved By Rookies," *Dexerto*, Jul. 8, 2020. [Online]. Available: https://www.dexerto.com/csgo/how-the-faze-csgo-super-team-is-being-saved-by-rookies-1389739/ (visited on 03/01/2024).

[6] A. Mujacic, "Team Elephant disbanding holds a lesson for all superteam builders," *Esports.net*, Nov. 14, 2021. [Online]. Available: https://www.esports.net/news/team-elephant-disbands-lessons-learned/ (visited on 03/01/2024).

[7] R. Mukherjee, "The failure of Team Liquid and Vitality might be the definitive end to super-teams in western League of Legends," *Sportskeeda*, Sep. 5, 2022. [Online]. Available: https://www.sportskeeda.com/esports/the-failure-team-liquid-vitality-might-definitive-end-super-teams-western-league-legends (visited on 03/01/2024).

[8] Seulgi, "Cloud9 part with yay," *VLR.gg*, Mar. 1, 2023. [Online]. Available: https://www.vlr.gg/180598/cloud9-part-with-yay (visited on 03/01/2024).

[9] teja, "Cloud9 release vanity," *VLR.gg*, Mar. 10, 2023. [Online]. Available: https://www.vlr.gg/184666/cloud9-release-vanity (visited on 03/01/2024).

[10] M. Švejda, "Cloud9 no match to ENCE at Gamers8," *HLTV*, Aug. 18, 2023. [Online]. Available: https://www.hltv.org/news/36900/cloud9-no-match-to-ence-at-gamers8 (visited on 03/01/2024).

[11] GSTK, "Cloud9 crash out of EPL after 0-2 loss to Eternal Fire," *HLTV*, Sep. 23, 2023. [Online]. Available: https://www.hltv.org/

news/37099/cloud9-crash-out-of-epl-after-0-2-loss-to-eternal-fire (visited on 03/01/2024).

[12] H. Richards, "Complexity send Cloud9 packing," *HLTV*, Oct. 18, 2023. [Online]. Available: https://www.hltv.org/news/37275/complexity-send-cloud9-packing (visited on 03/01/2024).

[13] D. Morris, "Falcons fail to qualify for PGL Major Copenhagen 2024," *Esports.net*, Feb. 16, 2024. [Online]. Available: https://www.esports.net/news/counter-strike/falcons-fail-to-qualify-for-pgl-major-copenhagen-2024/ (visited on 03/01/2024).

[14] S. Popkes, "Community reacts: Astralis miss out on PGL Major Copenhagen," *HLTV*, Feb. 23, 2024. [Online]. Available: https://www.hltv.org/news/38331/community-reacts-astralis-miss-out-on-pgl-major-copenhagen (visited on 03/01/2024).

[15] R. Flin, K.-M. Goeters, H. Hoermann, R. Amalberti, C. Valot, and H. Nijhuis, "Development of the notechs (non-technical skills) system for assessing pilots' crm skills," in Jan. 2003, pp. 221–243, ISBN: 978-0-7546-2829-3.

[16] J. van Avermaete and Nationaal Lucht- en Ruimtevaartlaboratorium, *NOTECHS: Non-technical Skill Evaluation in JAR-FCL* (NLR TP.: Nationaal Lucht- en Ruimtevaartlaboratorium). Nationaal Lucht- en Ruimtevaartlaboratorium, 1998. [Online]. Available: https://books.google.nl/books?id=qZNSHAAACAAJ.

[17] E. T. S. Tan, K. Rogers, L. E. Nacke, A. Drachen, and A. Wade, "Communication sequences indicate team cohesion: A mixed-methods study of ad hoc league of legends teams," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CHI PLAY, pp. 1–27, Oct. 2022, ISSN: 2573-0142. DOI: 10.1145/3549488. [Online]. Available: http://dx.doi.org/10.1145/3549488.

[18] ESM.ONE, Inc. "Popular esports games in 2023 by viewership." (2023), [Online]. Available: https://escharts.com/top-games?order=peak&year=2023 (visited on 03/01/2024).

[19] Riot Games, Inc. "VALORANT: Riot Games' competitive 5v5 character-based tactical shooter." (2024), [Online]. Available: https://playvalorant.com/en-us/ (visited on 03/01/2024).

[20] Riot Games, Inc. "Beginner's guide." (2020), [Online]. Available: https://playvalorant.com/en-us/news/announcements/beginners-guide/ (visited on 03/01/2024).

[21] D. Murko, "Riot Games to launch Valorant partner league in China — what's new in VCT 2024?" *Esports Charts*, Dec. 12, 2023. [Online]. Available: https://escharts.com/news/whats-new-vct-2024 (visited on 03/01/2024).

[22] D. Himmelstein, Y. Liu, and J. L. Shapiro, "An exploration of mental skills among competitive league of legend players," *International Journal of Gaming and Computer-Mediated Simulations*, vol. 9, no. 2, pp. 1–21, Apr. 2017, ISSN: 1942-3896. DOI: 10.4018/ijgcms.201704

0101. [Online]. Available: http://dx.doi.org/10.4018/IJGCMS.2017040101.

[23] T. Zandt, “Cs:go as a serious game for the navy,” Bachelor’s Thesis, University of Twente, Jul. 2021.

[24] Riot Games, Inc., *Riot Developer Portal: APIs*, 2021. [Online]. Available: https://developer.riotgames.com/apis.

[25] F. F. d. Nascimento Junior, A. S. d. C. Melo, I. B. da Costa, and L. B. Marinho, “Profiling successful team behaviors in league of legends,” in *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, ser. Webmedia ’17, ACM, Oct. 2017. DOI: 10.1145/3126858.3126886. [Online]. Available: http://dx.doi.org/10.1145/3126858.3126886.

[26] E. Tan, A. Wade, A. Kokkinakis, G. Heyes, S. Demediuk, and A. Drachen, “Less is more: Analysing communication in teams of strangers,” in *Proceedings of the 54th Hawaii International Conference on System Sciences*, ser. HICSS, Hawaii International Conference on System Sciences, 2021. DOI: 10.24251/hicss.2021.348. [Online]. Available: http://dx.doi.org/10.24251/HICSS.2021.348.

[27] T. Manninen, “Virtual team interactions in networked multimedia games case: "counter-strike" - multiplayer 3d action game,” *Teleoperators and Virtual Environments - Presence*, Nov. 2001.

[28] V. Lipovaya, Y. Lima, P. Grillo, C. E. Barbosa, J. M. de Souza, and F. J. d. C. M. Duarte, “Coordination, communication, and competition in esports: A comparative analysis of teams in two action games,” en, 2018. DOI: 10.18420/ECSCW2018_11. [Online]. Available: https://dl.eusset.eu/handle/20.500.12015/3122.

[29] E. Reid, R. L. Mandryk, N. A. Beres, M. Klarkowski, and J. Frommel, ““bad vibrations”: Sensing toxicity from in-game audio features,” *IEEE Transactions on Games*, vol. 14, no. 4, pp. 558–568, Dec. 2022, ISSN: 2475-1510. DOI: 10.1109/tg.2022.3176849. [Online]. Available: http://dx.doi.org/10.1109/TG.2022.3176849.

[30] J. Frommel, V. Sagl, A. E. Depping, C. Johanson, M. K. Miller, and R. L. Mandryk, “Recognizing affiliation: Using behavioural traces to predict the quality of social interactions in online games,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20, ACM, Apr. 2020. DOI: 10.1145/3313831.3376446. [Online]. Available: http://dx.doi.org/10.1145/3313831.3376446.

[31] T. Wanyi, “Understanding esports from the perspective of team dynamics.,” *Sport Journal*, p. 1, 2018, ISSN: 1543-9518. [Online]. Available: https://thesportjournal.org/article/understanding-esports-from-the-perspective-of-team-dynamics/.

[32] A. V. Carron and M. A. Eys, *Group Dynamics in Sport, 4th Edition*. Morgantown, WV: Fitness Information Technology, Jan. 2012, ISBN: 978-1-935412-35-9.

[33] B. A. De Jong, K. T. Dirks, and N. Gillespie, "Trust and team performance: A meta-analysis of main effects, moderators, and covariates.," *Journal of Applied Psychology*, vol. 101, no. 8, pp. 1134–1150, Aug. 2016, ISSN: 0021-9010. DOI: 10.1037/apl0000110. [Online]. Available: http://dx.doi.org/10.1037/apl0000110.

[34] A. Leavitt, B. C. Keegan, and J. Clark, "Ping to win?: Non-verbal communication and team performance in competitive online multiplayer games," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI'16, ACM, May 2016. DOI: 10.1145/2858036.2858132. [Online]. Available: http://dx.doi.org/10.1145/2858036.2858132.

[35] J. Wuertz, S. Bateman, and A. Tang, "Why players use pings and annotations in dota 2," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17, ACM, May 2017. DOI: 10.1145/3025453.3025967. [Online]. Available: http://dx.doi.org/10.1145/3025453.3025967.

[36] Riot Games, Inc., *Ask VALORANT - May 19*, May 2022. [Online]. Available: https://playvalorant.com/en-us/news/dev/ask-valorant-may-19/.

[37] D. Williams, S. Caplan, and L. Xiong, "Can you hear me now? the impact of voice in an online gaming community," *Human Communication Research*, vol. 33, no. 4, pp. 427–449, Oct. 2007, ISSN: 1468-2958. DOI: 10.1111/j.1468-2958.2007.00306.x. [Online]. Available: http://dx.doi.org/10.1111/j.1468-2958.2007.00306.x.

[38] A. Tang, J. Massey, N. Wong, D. Reilly, and W. K. Edwards, "Verbal coordination in first person shooter games," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ser. CSCW '12, ACM, Feb. 2012. DOI: 10.1145/2145204.2145292. [Online]. Available: http://dx.doi.org/10.1145/2145204.2145292.

[39] E. Kiourti, "Layering literacies and metagaming in counter strike: Global offensive," *L1-Educational Studies in Language and Literature*, pp. 1–27, Jul. 2022, ISSN: 1573-1731. DOI: 10.21248/l1esll.2022.22.2.367. [Online]. Available: http://dx.doi.org/10.21248/l1esll.2022.22.2.367.

[40] H. Hoermann, K. Burger, and H. Neb, "Evaluation of pilot´s interpersonal skills. usability of an integrated proficiency reporting from within lufthansa," Apr. 2003.

[41] P. D. Birch, L. Greenlees, and B. T. Sharpe, "An exploratory investigation of personality in counter-strike: Global offensive," *Journal of Electronic Gaming and Esports*, vol. 1, no. 1, Jan. 2023, ISSN: 2836-3523. DOI: 10.1123/jege.2022-0027. [Online]. Available: http://dx.doi.org/10.1123/jege.2022-0027.

[42] A. Ceschi, A. Costantini, V. Zagarese, E. Avi, and R. Sartori, "The notechs+: A short scale designed for assessing the non-technical skills (and more) in the aviation and the emergency personnel,"

Frontiers in Psychology, vol. 10, May 2019, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.00902. [Online]. Available: http://dx.doi.org/10.3389/fpsyg.2019.00902.

[43] J. Falwell, "Exploring tilt in esports," M.S. thesis, Illinois State University, 2023.

[44] O. E. D. Hamlet, A. Irwin, R. Flin, and N. Sedlar, "The development of the helicopter non-technical skills (helinots) behavioural marker systems," *Ergonomics*, pp. 1–10, Apr. 2023, ISSN: 1366-5847. DOI: 10.1080/00140139.2023.2194591. [Online]. Available: http://dx.doi.org/10.1080/00140139.2023.2194591.

[45] S. Yule, R. Flin, S. Paterson-Brown, N. Maran, and D. Rowley, "Development of a rating system for surgeons' non-technical skills," *Medical Education*, vol. 40, no. 11, pp. 1098–1104, Nov. 2006, ISSN: 1365-2923. DOI: 10.1111/j.1365-2929.2006.02610.x. [Online]. Available: http://dx.doi.org/10.1111/j.1365-2929.2006.02610.x.

[46] A. Irwin, I.-R. Tone, and N. Sedlar, "Developing a prototype behavioural marker system for farmer non-technical skills (flints)," *Journal of Agromedicine*, vol. 28, no. 2, pp. 199–207, Jun. 2022, ISSN: 1545-0813. DOI: 10.1080/1059924x.2022.2089420. [Online]. Available: http://dx.doi.org/10.1080/1059924X.2022.2089420.

[47] S. Yule, R. Flin, N. Maran, D. Rowley, G. Youngson, and S. Paterson-Brown, "Surgeons' non-technical skills in the operating room: Reliability testing of the notss behavior rating system," *World Journal of Surgery*, vol. 32, no. 4, pp. 548–556, Feb. 2008, ISSN: 1432-2323. DOI: 10.1007/s00268-007-9320-z. [Online]. Available: http://dx.doi.org/10.1007/s00268-007-9320-z.

[48] S. Kozlowski and S. Kozlowski, "The oxford handbook of organizational psychology," Sep. 2012. DOI: 10.1093/oxfordhb/9780199928309.001.0001.

[49] R. Fernandez, S. Shah, E. D. Rosenman, S. W. J. Kozlowski, S. H. Parker, and J. A. Grand, "Developing team cognition: A role for simulation," *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, vol. 12, no. 2, pp. 96–103, Apr. 2017, ISSN: 1559-2332. DOI: 10.1097/sih.0000000000000200. [Online]. Available: http://dx.doi.org/10.1097/SIH.0000000000000200.

[50] N. J. Cooke, J. C. Gorman, J. L. Duran, and A. R. Taylor, "Team cognition in experienced command-and-control teams.," *Journal of Experimental Psychology: Applied*, vol. 13, no. 3, pp. 146–157, 2007, ISSN: 1076-898X. DOI: 10.1037/1076-898x.13.3.146. [Online]. Available: http://dx.doi.org/10.1037/1076-898X.13.3.146.

[51] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, *Robust speech recognition via large-scale weak supervision*, 2022. DOI: 10.48550/ARXIV.2212.04356. [Online]. Available: https://arxiv.org/abs/2212.04356.

[52] W. Rozelle. "VCT GAME CHANGERS." (2021), [Online]. Available: https://playvalorant.com/en-us/news/esports/vct-game-changers/ (visited on 06/15/2024).

[53] F. R. H. Zijlstra, M. J. Waller, and S. I. Phillips, "Setting the tone: Early interaction patterns in swift-starting teams as a predictor of effectiveness," *European Journal of Work and Organizational Psychology*, vol. 21, no. 5, pp. 749–777, Oct. 2012, ISSN: 1464-0643. DOI: 10.1080/1359432x.2012.690399. [Online]. Available: http://dx.doi.org/10.1080/1359432X.2012.690399.

[54] G. King and L. Zeng, "Logistic regression in rare events data," *Political Analysis*, vol. 9, 137–163, Feb. 2001.

[55] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[56] Y. Liu, M. Ott, N. Goyal, *et al.*, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019. arXiv: 1907.11692. [Online]. Available: http://arxiv.org/abs/1907.11692.

[57] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. [Online]. Available: https://aclanthology.org/N19-1423.

[58] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697.

[59] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[60] Martín Abadi, Ashish Agarwal, Paul Barham, *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/.

[61] T. Wolf, L. Debut, V. Sanh, *et al.*, "Huggingface's transformers: State-of-the-art natural language processing," *CoRR*, vol. abs/1910.03771, 2019. arXiv: 1910.03771. [Online]. Available: http://arxiv.org/abs/1910.03771.

[62] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1412.6980.

[63] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, S. van der Walt and J. Millman, Eds., 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

[64] T. pandas development team, *Pandas-dev/pandas: Pandas*, version latest, Feb. 2020. DOI: 10.5281/zenodo.3509134. [Online]. Available: https://doi.org/10.5281/zenodo.3509134.

[65] L. St»hle and S. Wold, "Analysis of variance (anova)," *Chemometrics and Intelligent Laboratory Systems*, vol. 6, no. 4, pp. 259–272, 1989, ISSN: 0169-7439. DOI: https://doi.org/10.1016/0169-7439(89)80095-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0169743989800954.

[66] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965, ISSN: 00063444, 14643510. [Online]. Available: http://www.jstor.org/stable/2333709.

[67] H. Levene, "Robust tests for equality of variances," *Contributions to probability and statistics*, pp. 278–292, 1960.

[68] B. L. WELCH, "ON THE COMPARISON OF SEVERAL MEAN VALUES: AN ALTERNATIVE APPROACH," *Biometrika*, vol. 38, no. 3-4, pp. 330–336, Dec. 1951, ISSN: 0006-3444. DOI: 10.1093/biomet/38.3-4.330. eprint: https://academic.oup.com/biomet/article-pdf/38/3-4/330/716814/38-3-4-330.pdf. [Online]. Available: https://doi.org/10.1093/biomet/38.3-4.330.

[69] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952. DOI: 10.1080/01621459.1952.10483441. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1952.10483441. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1952.10483441.

[70] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[71] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2.