Software Requirements Specification for SFWRENG 4G06 Capstone Design Project: MotionMingle

Team #18, InfiniView-AI
Anhao Jiao
Kehao Huang
Qianlin Chen
Qi Shu
Xunzhou Ye

March 26, 2024

Contents

Table 1: Revision History

Date	Developer(s)	Change
25 September 2023	AJ, KH, QC, QS, XY	Initial draft
02 February 2024	AJ, KH, QC, QS, XY	rev 0.1
22 March 2024	AJ, KH, QC, QS, XY	rev 1

1 Introduction

This document serves as a comprehensive requirements specification for a ~~WebRTC-based video conferencing~~ Real-Time video streaming application tailored specifically for Tai Chi instruction. The purpose of this document is to provide a detailed guide for the development and implementation of the system, ensuring that all stakeholders and developers have a clear and unified understanding of the project's scope, objectives, and functional and non-functional requirements. This document follows a structured format of the Meyer's template, which has four major dimensions or PEGS: Project, Environment, Goals, System. The four dimensions correspond to the four major sections in this document. Slight modifications are made on the document structure to better fit our project:

- The Functionality section under the System dimension is divided into the functional requirements section and the non-functional requirements section for better clarity.

- Added the Use Case Diagram section to illustrate use cases.

- Added the Finite State Machine section to provide formal specification.

1.1 Glossary

Tai Chi A classical Chinese martial art system practiced for health promotion and rehabilitation.

Instructor A person who teaches a Tai Chi class through an online conference system.

Practitioner A person who learns Tai Chi through an online conference system.

Skeleton Annotation A type of annotation that the system can generate. This type of annotation looks like a human skeleton.

~~**Clock Annotation**~~ A type of annotation that the system can generate. This type of annotation looks like a physical clock surrounding the instructor.

Center of Mass Visualization A type of annotation that the application generates. This type of annotation shows the instructor's center of mass with the instructor's face and feet's direction.

Semantic Segmentation The process of linking each pixel in an image to a class label. It is a pixel-wise classification task that assigns a semantic label like "person", "car", "road", etc. To every pixel in the image.

Pose Estimation/Extraction The process of using machine learning models to estimate the pose of a person from an image or a video by estimating the spatial locations of the person's key body joints.

2D Pose Estimation The process of estimating the 2D position and orientation of a person or object in an image or video. This involves identifying the key points on the body or object, such as the head, shoulders, knees, etc. then estimating their position in 2D space.

3D Pose Estimation Similar to 2D pose estimation position estimation is in 3D space. One advantage of 3D pose estimation is it can show the rotation of a body part(i.e. Wrist rotation), whereas 2D pose estimation can't.

Machine Learning Model A mathematical model designed to find patterns and make predictions or decisions based on data.

Server-Side Rendering The technique of rendering streaming video on the server instead of on clients' devices.

Annotation Pipeline A sequence of processing elements connected in series, which are responsible for generating annotations.

Annotation Configuration Options for practitioners to configure the annotation preference.

UI User Interface

UX User Experience

AI Artificial Intelligence

OS Operating System

2 Project

This section provides a comprehensive overview of the project's key aspects. It outlines the project's structure, responsibilities of team members, technical decisions, and the plan for project execution.

2.1 Roles and Personnel

A team of five members is designated to carry out the planning, managing, documenting, and execution phases of the project specified in this requirements document (referred to as "the project"). The project is initiated and will be executed in the context of an accumulative academic course on software designs and life cycles. The situation presents time constraints and resource management challenges for all personnel involved. Members of the team are expected to identify potential problems that might hinder the progress of the project, voice their concerns or suggestions, acknowledge their shares of work and fulfill their own responsibilities with a high standard of quality. For each team member, the expected workload is 10 hours per week, totalling 50 hours per week. This is the typical workload of a full-time employee in the technology industry. Also constrained by the academic context, there is no preserved capacity for additional human resources. Induction of further personnel is not possible throughout the execution of the project. The project leader is expected to efficiently utilize and manage the limited resources and group members are expected to be held accountable for their duties.

2.1.1 Individual Responsibilities

- Xunzhou Ye specializes in the machine learning portion of the project. Subsequent responsibilities include learning and investigating preliminary studies completed before the initiation of the project, and developing from and optimizing existing machine learning-powered processing pipelines.

- From the managerial perspective of the project, Ye takes on the leadership role within the team. Subsequent responsibilities include envisioning the direction in which the project progresses, sustaining and/or enhancing the project momentum, coordinating team members to main-

tain a high level of productivity, actively listening to peer feedback, and timely stepping in and helping with resolving conflicts of interest.

- Qi Shu and Kehao Huang specialize in connecting and integrating machine learning with the component of the system that interfaces with the user. Subsequent responsibilities include studying the preliminary works in-depth and designing interfaces internal to the system to connect and integrate major components.

- From the managerial perspective, Qi is the co-project leader. Subsequent responsibilities include overseeing and documenting the progress of the project and communicating with managerial stakeholders such as academic and project supervisors.

- Qianlin Chen and Anhao Jiao specialize in articulating and implementing the means of interaction between the system and the end user. Subsequent responsibilities include designing the interfaces used for the communication between the system and the user and architecting both the software and hardware infrastructures to support the interfaces.

2.2 Imposed Technical Choices

N/A ~~The majority of the technical choices and decisions originate from the academic context in which the project is planned and carried out. Most of the technical resources shall be chosen from available open-source software, development tools, and intellectual assets. Despite the supervisor allotting a small budget for the completion of the project, team members shall minimize financial costs when making design choices and budgeting decisions. On the other hand, to accommodate the significant amount of computing resources required by machine learning-powered applications, our academic institution would offer computing hardware to support the operation of the specified system in the "System" book. The decision to utilize a self-hosted server, rather than opting for commercial cloud-hosted services, is clear and would help us adhere to our financial budget constraints. Other technical choices include the choice of programming language, Python, for the machine learning component of the system, and choosing JavaScript/TypeScript for the web application component of the system. Both of these choices are rooted in the fact that the language used to describe each component aligns with industry~~

standards and conventional practices. The technology ecosystems around the two languages are mature and well-tested.

Table 2: Milestones and Scheduled Due Dates

Milestone	Scheduled Due Date
Team Formed, Project Selected	18 September 2023
Problem Statement, Proof of Concept Plan, Development Plan	25 September 2023
Requirements Documents Revision 0	06 October 2023
Hazard Analysis Revision 0	20 October 2023
Verification and Validation Plan Revision 0	03 November 2023
Proof of Concept Demonstration	Week of 17 November 2023
Design Document Revision 0	17 January 2024
Project Demonstration Revision 0	Two weeks starting 05 February 2024
Verification and Validation Report Revision 0	06 March 2024
Project Demonstration Revision 1 (Final)	Two weeks starting 18 March 2024
Full Project Documentation Revision 1 (Final)	04 April 2024

2.3 Schedule and Milestones

Please refer to table 2

2.4 Tasks and Deliverable

For the documentation portion for each milestone, the work shall be broken down evenly for each team member. The distribution of work shall be carefully considered. Related or connected topics shall be distributed to an individual team member. Doing so would minimize the need for frequent sync-ups with other team members. Specific division of tasks shall be brought up and discussed in the deliverables planning meetings around 2 weeks ahead of the deliverable due. The description of the divided tasks and the expected completion date shall be posted as checkbox items on the Github meeting issue.

For the implementation aspect of each milestone, there are two stages

in total. The first one being to create a prototype for the proof of concept demonstration in mid-November. In this stage, the team shall be focusing on implementing the real-time video conferencing web application, which includes the user clients, UI designs, and the servicing infrastructure. At the same time, all team members shall be familiar with all technologies relevant to implementations. The second stage is to build on top of the prototype, carrying out the full implementation for the Revision 0 demonstration in February 2024. At this stage, the web application shall seamlessly integrate with the machine learning models and provide all specified functionalities and features to users.

2.5   Required Technology Elements

Programming Language: Python JavaScript/TypeScript

Linter Tool: Flake8, ESlint

Testing Framework: ~~pytest, Jest~~ aiohttp, React Testing Library

Code coverage: ~~pytest-cov, jest~~ React Testing Library

Version Contro:l git

CI/CD: GitHub Actions

Libraries: ~~WebRTC, PyTorch, Bun.js~~ aioRTC, MediaPipe, React.js

ML Models: ~~OpenPose, SMPL, ShapeMask~~ MediaPipe

UI Design framework: ~~Ant Design~~ Material UI

All of the listed tools are readily available to the public. The availability of these technology elements is crucial to the success of the project.

2.6   Risk and Mitigation Analysis

One potential risk for this project is insufficient background knowledge in some of the required technologies: ~~WebRTC~~ Real-Time streaming protocol, socket programming, and computer vision. To mitigate this, time will be allocated for team members to learn these technologies before utilization in the project.

Furthermore, the current technology selections and goals may evolve over the course of development. For example, it is uncertain whether the machine learning model can generate real-time annotations given factors like network traffic and chosen server-side rendering techniques. If significant changes occur, documentation will be updated accordingly to reflect implemented adjustments.

Another potential risk affecting our project's deliverables pertains to human factors. Given that team members are concurrently engaged in multiple other courses and possess diverse life priorities, the project's progress heavily relies on effective time management among team members. It is conceivable that our pace may decelerate during peak examination periods. To mitigate this, the team collectively agreed that each member should allocate a minimum of 10 hours per week to ensure consistent progress.

Ineffective communication could also pose a risk. To minimize miscommunications, the team will hold two scheduled meetings per week, along with preparation sessions before external stakeholder meetings to ensure every team member is on the same page. Platforms such as Discord and Teams will be used for daily communication and to clarify any misunderstandings.

Finally, this project presents not just technical challenges but also demands meticulous documentation of all development processes. Meeting both objectives within the given timeframe may prove infeasible. As discussed with the course instructor, documentation will take priority if such resource conflicts arise.

2.7   Requirements Process and Report

In the requirement elicitation process, project requirements will be acquired during our bi-weekly meetings with the supervisor. To facilitate efficient communication and effective elicitation, it is imperative to conduct brief question preparation sessions beforehand. Additionally, as the project progresses, a clearer distinction between our project implementation and objectives will emerge. This differentiation is likely to engender new project requirements, which will be systematically identified, summarized, and incorporated into our project documentation.

3 Environment

3.1 Glossary

Please refer to 1.1 Glossary

3.2 Components

3.2.1 Hardware Components

Webcam A Webcam is required for the instructor's client.

~~**Microphone** Microphone is optional for both the instructors and the practitioners.~~

Monitor A monitor is required for practitioners.

Host Server A host server is required for routing, processing video streams and annotation generation.

Network Connection Having access to the internet is required for both instructors and practitioners.

3.2.2 Software Components

OS The OS that the application runs on might affect the usability of the application. (i.e. the application doesn't run on some OS)

Docker Required for standardized application deployment from the developer's machine to the host server.

Machine Learning Models Machine Learning Models are required for instructor's pose extraction.

3.3 Constraints

Bandwidth The Application must not occupy more than 20MB/s download speed[6] and 5 MB/s upload speed[7], as these two numbers are the most common home internet speeds for Canadian families.

Compatibility The product must be able to operate on all Windows, Mac OS and Linux platforms because over 95% of the computers run on these three systems[5].

Security Data leak in any video conference system is not permitted, according to the Personal Information Protection and Electronic Documents Act[1].

Ease of Learning One of the target audiences of our system is senior people. This user group usually has some difficulties in using software as they have lower technical device adoption[3]. The UI interaction must be clear and intuitive enough for senior people to learn all the features without further instructions.

Compliance The application is required to comply with all applicable laws[1][2][4] and regulations.

3.4   Assumptions

Internet Access Users of the system are assumed to have access to a ~~reliable~~ internet connection. Packet loss and interruptions resulting from poor network connections are ~~generally not~~ accounted for.

Device Capability and Compatibility The system is assumed to run on modern portable/non-portable computing devices with adequate computational power and basic networking and real-time data streaming capabilities. The software runtime environment on top of the capable hardware is also assumed to be up-to-date. Extensive backward compatibility with outdated, deprecated hardware or software is not required.

Understanding of the Annotation After adding annotations to the instructor's streaming video, practitioners are assumed to understand these annotations.

Cognitive Ability Assume users are able to understand and use the interface of the system, presuppose that users can navigate, process information, and make informed decisions within the interface without struggling.

Supported Video Resolutions Most users' built-in webcams in laptops and phones are assumed to support video resolution from 480p ~~up~~ to 1080p, which is commonly used for streaming and daily uses.

Scalability of the System The system will be designed with scalability in mind, assuming a growing user base.

3.5   Effects

3.5.1   Improved Online Tai Chi Instruction Quality

By providing visual annotations, the system improves the clarity and effectiveness of online Tai Chi instruction. Instructors can convey movements more clearly, leading to better learning outcomes for practitioners.

3.5.2   Increased Accessibility to Tai Chi

By minimizing hardware requirements and providing an intuitive user interface, the system enables individuals who may not have access to traditional Tai Chi classes or specialized equipment to participate in Tai Chi practice from the comfort of their homes.

3.5.3   Increased Tai Chi Awareness

By making Tai Chi accessible to a wider audience, the platform can raise awareness about the benefits of Tai Chi, potentially leading to increased interest in and participation in Tai Chi practice in society.

3.5.4   Reduced Carbon Footprint

By offering Tai Chi lessons online, practitioners can practice from their own homes, reducing the need for travel to physical locations. This can result in fewer carbon emissions associated with transportation.

3.6   Invariants

Camera The camera used by the instructor during Tai Chi sessions must remain active and operational for the entire duration of the session.

Internet Bandwidth Sufficient internet bandwidth for both the instructor and participants must be consistently available to maintain the quality and reliability of Tai Chi sessions.

Data Encryption Throughout the streaming lesson, it is imperative to consistently uphold and ensure the secure transmission of user data and video call content to safeguard user privacy.

Video Transmission The video information transmitted to the client remains invariant with respect to the source video, excluding any annotations. In other words, annotations only supplement the original content with additional information and should not alter the amount of information or the quality provided by the original videos.

4 Goals

This section outlines the goals and objectives of the project, providing context for its development. It also presents the expected benefits of the project and offers a high-level overview of the system's functionality, limitations, and stakeholders involved in the project.

4.1 Context and Overall Objective

The prevailing situation in Tai Chi instruction involves a transition from traditional in-person classes to online training sessions induced by the enforcement of social distancing during the COVID-19 pandemic. The transition has imposed numerous challenges for both instructors and practitioners. This project aims to address the pressing need for an enhanced online learning platform for Tai Chi. By providing real-time annotations, minimizing hardware requirements, ensuring data security, and offering user-friendly interfaces, our project is expected to elevate online Tai Chi instruction to a smooth and enriching experience. The project is subsequently decomposed into two major components: the video conferencing platform that supports video and audio communications, and the machine learning pipelines that process the video stream from instructors and generate real-time annotations.

4.2 Current Situation

During the COVID-19 pandemic, group exercise classes like Tai Chi transitioned to the online format due to the enforcement of social distancing. Currently, virtual Tai Chi training is usually conducted by either offering

live sessions with webcams or watching pre-recorded videos. Both ways of virtual Tai Chi training fall short compared to traditional in-person instruction as widely perceived among Tai Chi practitioners and instructors for several reasons. One significant drawback of learning Tai Chi Online is that practitioners may not be able to see all the angles of the instructor's body needed to understand and mimic the movements. Also, the lack of real-time feedback from instructors and interaction among participants is also limiting the learning outcomes and engagement for practitioners. Therefore, the virtual Tai Chi training model is considered inferior to the traditional in-person teaching model.

4.3   Expected Benefits

The successful execution of the project promises several business benefits. First, it will significantly elevate the learning and teaching experience for online Tai Chi training sessions. By providing a video conferencing platform that supports real-time audio and video communication, along with offering various user-configurable annotations that enhance the visibility of the instructor's movements, our solution aims to bridge the gap between traditional in-person instruction and the online learning model. This enhancement is expected to result in higher engagement levels and improved learning outcomes of online Tai Chi training.

Secondly, the project emphasizes the safety of user data, ensuring trust and privacy among users. By safeguarding user data and preventing any case of data leakage, the application aims to provide a secure and reliable platform for Tai Chi instructors and practitioners.

Third, the application aims to minimize the hardware requirements for end users. Ideally, the application should be accessible to users as long as they have a device that supports video conferences and has access to the internet. Therefore, the application is accessible to audiences who have an interest in learning Tai Chi online but only have digital devices with limited computing power.

Lastly, the system's ease of use will make Tai Chi accessible to individuals of all age groups. This application should be intuitive enough to use for all users that no further instruction is required. The inclusiveness of this application not only promotes the practice of Tai Chi but also contributes to the overall well-being of individuals, aligning with the broader mission of health promotion and rehabilitation associated with Tai Chi.

To conclude, the successful execution of the proposed project is foreseen to revolutionize the landscape of virtual Tai Chi training, making it a more efficient, accessible and enriching experience for both practitioners and instructors.

4.4 Functionality Overview

The system is designed with multiple features tailored for practitioners. First, it ensures seamless video streaming, allowing users to watch and follow videos in real-time. Additionally, practitioners have the flexibility to select and view specific annotations on the video, which provide supplementary information or guidance. A standout capability of this system is its automated motion detection feature. This means that when the instructor's movements are detected, relevant annotations are automatically overlaid onto the video, aiding practitioners in better understanding and mirroring the demonstrated actions.

4.5 High-Level Usage Scenarios

For high-level usage scenarios, the system can be utilized for various purposes. The primary intended usage scenario is supporting Tai Chi teaching and learning sessions in a synchronous setting. Body movements and forms demonstrated by a Tai Chi instructor are captured in an ordinary media stream and then live-streamed to all interested practitioners. Features of the movements and forms are extracted from the media stream and used to achieve or aid in achieving the functionalities described above.

Furthermore, such a video could be extended to more generic usage scenarios. Virtually any online media conference or offline instructional record which involves showcasing and placing emphasis on human limb movements, and is created for demonstrative and educational purposes, could benefit from the additional information provided by the generated annotations. Specific examples of the use cases include but are not limited to group or personal gym workout sessions, breakdowns of movements and techniques in various sports, and demonstrations of physiotherapy or rehabilitation exercises. Finally, it has the potential to perform motion tracking, pose estimation, and body auto-rigging solely with video data input.

4.6 Limitations and Exclusions

The system is tailored with specific guidelines in mind. It's worth noting that the product is specially crafted to analyze and provide insights on slow movements, much like those found in Tai Chi practices. As such, it might not deliver optimal results for fast-paced or dynamic actions. There's no user login mechanism because it's deemed unnecessary for the system's core functionality. Furthermore, the feedback and instruction are designed to be general, encompassing various Tai Chi techniques without the need to segregate them into different classes. Lastly, the system focuses on providing real-time and interactive guidance, it is not compatible with video playback features.

4.7 Stakeholders and Requirements Sources

4.7.1 Stakeholders

Tai Chi Practitioners Individuals who practice Tai Chi and can benefit from real-time instruction video after model rendering.

Tai Chi Instructors Professionals who teach Tai Chi and can utilize the system for synchronous teaching sessions.

Tai Chi Potential Practitioners Individuals who are learning Tai Chi and can use the system to improve their practice.

Capstone Project Supervisor A supervisor who supervises the progress of the project.

4.7.2 People Other Than Stakeholders

Capstone Course Instructor The instructor who oversees the capstone project and provides guidance and evaluation.

Course Teaching Assistants Teaching assistants who support the capstone course and may provide additional guidance or feedback.

Capstone Project Peer Reviewers Individuals who review and provide feedback on the capstone project.

4.7.3 Predating Documentations

There are two primary documents that predate the current research. The first encompasses feedback from users who faced challenges with plain video calls during the COVID-19 user preferences and opinions regarding annotation methods and features. The second compiles feedback from demo studies and focus groups, shedding light on user preferences and opinions regarding annotation methods and features. Together, these feedback sources offer a comprehensive view of user needs and project requirements.

5   System

This section provides a detailed breakdown of the system's desired functionalities and behavior with priorities, proposes a high-level verification plan with acceptance criteria, and states all the likely and unlikely changes of the system.

5.1   Components

5.1.1   Front-End (Client-Side)

- User Interface (UI)

- ~~WebRTC~~ Real-Time Communication Client

- ~~Real-Time Communication Engine~~

5.1.2   Back-End (Server-Side)

- Signaling Server

- Selective Forwarding Unit(SFU)

- STUN/TURN Servers:

    - STUN: Session Traversal Utilities for NAT
    - TURN: Traversal Using Relays around NAT

- API Server

- Machine Learning Models (Annotations)

5.1.3 Infrastructure

- Networking

- Cloud

For the ~~WebRTC-based video conferencing~~ Real-Time video streaming application using an SFU architecture, participants initiate sessions through a user interface. The ~~WebRTC~~ Real-Time Communication Client, aided by STUN and TURN servers, manages connection establishment and multimedia stream transmissions. Real-time Communication Engine establishes, maintains, and terminates the real-time sessions. Session setup details, communicated via the Signaling Server using the Session Description Protocol (SDP), determine the following media rendering. During the video conference, the Selective Forwarding Unit (SFU) routes and optimizes media stream distribution to participants. The video streams will be extracted from the SFU server, and transferred to the API server by the socket. Then the API server will assign its integrated Machine Learning Models to render the video and send it back to the client side. The underlying infrastructure ensures efficient data transmission, maintaining overall application performance.

5.2 Finite State Machine

21
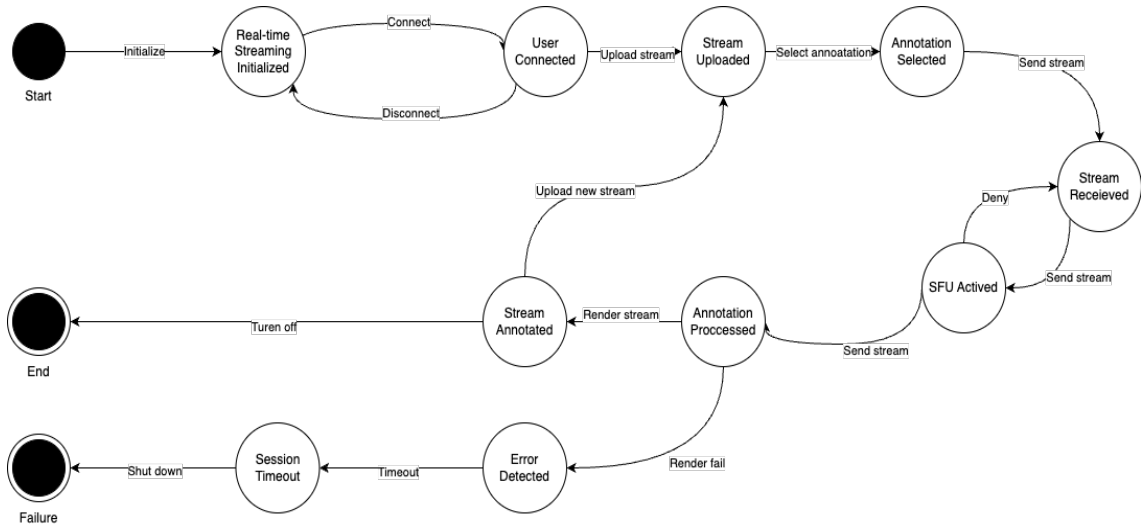
5.3 Functional Requirements

FR1 The system shall provide a real-time video streaming interface for Tai Chi instructors to conduct live sessions.

- Rationale: This functionality allows instructors to deliver live Tai Chi classes to practitioners, facilitating real-time interaction and instruction.
- Priority: Critical

FR2 The system shall provide a real-time video streaming interface for Tai Chi practitioners to watch live sessions.

- Rationale: This functionality allows practitioners to receive live Tai Chi sessions.
- Priority: Critical

FR3 The system shall turn on the webcam on the instructor's device.

- Rationale: The system needs the video stream from the instructor's device as input.
- Priority: Critical

FR4 The system shall establish a connection between the instructor and the server.

- Rationale: The system needs instructor-server connections to transmit the video and/or audio streams.
- Priority: Critical

FR5 The system shall establish the connection between each practitioner and the server for downstream data transmission.

- Rationale: The system needs practitioner-server connections to transmit video and/or audio streams.
- Priority: Critical

FR6 The system shall allow practitioners to configure the type of annotations.

– Rationale: The system needs user configuration details from practitioners as parameters to arrange the corresponding machine-learning pipeline.

– Priority: Critical

FR7 The system shall listen to any update on annotation configurations from practitioners.

– Rationale: The server needs to update and rearrange machine learning pipelines according to annotation configurations.

– Priority: Critical

FR8 The system shall send the annotation configuration info from the practitioner-client to the server.

– Rationale: The server needs to receive annotation configurations from clients.

– Priority: Critical

FR9 The system shall process the annotation configuration received from the practitioner-client.

– Rationale: The system needs user configuration details from practitioners as parameters to create the corresponding machine-learning pipeline.

– Priority: Critical

FR10 The system shall arrange ~~machine learning~~ annotation pipeline(s) based on annotation configuration.

– Rationale: The system needs to generate correct annotations with the pipeline according to the practitioner's configuration.

– Priority: High

FR11 The system shall render the instructor's video stream with correct annotation ~~through machine learning pipelines~~.

– Rationale: This is one of the outputs of the system.

– Priority: Critical

FR12 The system shall transmit the annotated video stream to each practitioner-client through an established practitioner-server connection.

  – Rationale: This is one of the outputs of the system.
  – Priority: Critical

FR13 The signaling server shall always be available to establish ~~WebRTC~~ Real-Time streaming connections.

  – Rationale: Signaling is essential for establishing real-time communications.
  – Priority: Critical

FR14 The system shall provide 2D Pose Estimation skeleton annotation.

  – Rationale: Increase Tai Chi learning experience by providing the instructor's 2D Pose Estimation skeleton annotation.
  – Priority: Critical

FR15 The system shall provide 3D Pose Estimation skeleton annotation.

  – Rationale: Increase Tai Chi learning experience by providing the instructor's 3D Pose Estimation skeleton annotation.
  – Priority: High

FR16 The system shall provide semantic segmentation annotation.

  – Rationale: To distinguish the instructor from the background and make the annotation more accurate.
  – Priority: Critical

FR17 The system shall provide footwork annotation.

  – Rationale: To identify the instructor's footwork annotation and increase Tai Chi learning experience.
  – Priority: Critical

5.4   Non-Functional Requirements

5.4.1   Look and Reel Requirement

LF1 The system shall have a user interface design that prioritizes clear and logically organized menus

- Rationale: To ensure that users can comfortably use the system.
- Fit Criteria: 95% of users are satisfied with the appearance of the UI interface.
- Priority: Medium

5.4.2   Usability and Humanity Requirement

Ease of Use Requirement:

UH1 ~~The system shall have a straightforward and intuitive user interface design.~~ At least 90% of the users can complete all operations and nevigation without assistance

- Rationale: To minimize user frustration and ensure a positive user experience.
- Fit Criteria: Users can complete common tasks within the system without extensive assistance.
- Priority: Medium

Ease of Learning Requirement

UH2 The system shall be easy to learn for users of all age groups and experience levels.

- Rationale: To accommodate a diverse user base, including older individuals who may be less familiar with technology.
- Fit Criteria: Elder users can perform basic tasks and navigate the system effectively ~~after a short learning period~~ within 10 minutes.
- Priority: Medium

5.4.3 Performance Requirements

Speed Requirement

PR1 The system shall respond to user interactions (e.g. button clicks, menu selections) within 1 second.

  – Rationale: To provide a responsive and smooth user experience.
  – Fit Criteria: User interactions result in near-instantaneous system responses under typical conditions.
  – Priority: Medium

Reliability and Availability Requirements

PR2 The signaling server, SFU, and STUN/TURN servers shall operate with high reliability, minimizing service interruptions during live sessions.

  – Rationale: To ensure a consistent and uninterrupted learning experience for users.
  – Fit Criteria: Real-time communication services are always available.
  – Priority: High

Capacity Requirement

PR3 The system shall support a minimum of 5 users using the system during peak usage periods.

  – Rationale: To accommodate a potentially large user base.
  – Fit Criteria: The system can handle 5 simultaneous users without performance degradation.
  – Priority: Medium

Robustness or Fault-Tolerance Requirement

PR4 The system shall implement zero fault tolerance with respect to user errors .

  – Rationale: It is essential that users receive appropriate feedback when errors occur to support system understandability.

- Fit Criteria: Warning messages indicating errors shall be displayed to the user to confirm lack of fault tolerance per the zero fault tolerance requirement.
- Priority: Medium

Scalability of Extensibility Requirement

PR5 ~~The system shall be able to add new extensions.~~

- ~~Rationale: Further improvements and extensions of the system are needed.~~
- ~~Fit Criteria: Not applicable.~~
- ~~Priority: Medium~~

PR6 The Selective Forwarding Unit (SFU) shall be scalable to accommodate an increasing number of simultaneous video streams as the user base grows.

- Rationale: To support a growing user community without performance degradation.
- Fit Criteria: The SFU can handle a minimum of 10 simultaneous video streams during peak usage.
- Priority: Medium

5.4.4 Operational and Environmental Requirements

OE1 The system shall be compatible with the latest versions of Windows, Linux, and macOS operating systems.

- Rationale: To ensure that users can access the platform regardless of their choice of operating system.
- Fit Criteria: The platform functions seamlessly on Windows, Linux, and macOS without critical issues.
- Priority: High

OE2 The system shall be compatible with commonly used web browsers, including but not limited to Chrome, Firefox, Safari, and Edge.

– Rationale: To make the platform accessible to users who prefer different web browsers.

– Fit Criteria: The platform functions correctly on major web browsers without significant deviations in performance or appearance.

– Priority: High

OE3 The system shall run smoothly on standard personal computers or laptops equipped with a camera ~~and optionally a microphone~~.

– Rationale: To minimize hardware-related barriers to using the platform.

– Fit Criteria: Users with standard personal computers or laptops can use the platform ~~without significant performance issues~~ with a video streaming latency of less than 500 ms while maintaining a minimum frame rate of 10 fps.

– Priority: High

5.4.5 Maintainability and Support Requirements

MS1 The system shall be designed for ease of maintenance and updates.

– Rationale: To ensure that the system can be efficiently maintained and improved over time.

– Fit Criteria: Updates and maintenance tasks can be performed with minimal disruption to users.

– Priority: Medium

MS2 ~~The system shall take no longer than 4 hours to be maintained/updated.~~

– ~~Rationale: To ensure the availability of the system.~~

– ~~Fit Criteria: Updates can be done within 4 hours.~~

– ~~Priority: Low~~

28

5.4.6 Security Requirements

SR1 ~~The system shall securely transmit and store user data, including personal information.~~

- ~~Rationale: To protect user privacy and data integrity.~~
- ~~Fit Criteria: User data is encrypted both in transit and at rest, and security protocols follow industry best practices.~~
- ~~Priority: Critical~~

SR2 The system shall not collect any non-essential information from users.

- Rationale: This prevents unauthorized disclosure of users' information.
- Fit Criteria: Any information gathering requires users' consent.
- Priority: Critical

SR3 The system shall not be able to be modified by unauthorized users.

- Rationale: This prevents the system from malfunctioning.
- Fit Criteria: Any updates/modifications to the system shall only be made and reviewed by the development team.
- Priority: Critical

5.4.7 Cultural Requirements

CR1 The system shall not display any offensive cultural symbol/language.

- Rationale: The intended users of the system are considered to have various cultural and religious backgrounds.
- Fit Criteria: Not applicable.
- Priority: Critical

5.4.8 Legal Requirements

LR1 The system shall meet the ISO/IEC 12207.

- Rationale: ISO/IEC 12207 is an international standard for software lifecycle processes.
- Fit Criteria: The system meets the standard of ISO/IEC 12207.
- Priority: Critical

5.4.9 Health and Safety Requirements

HS1 The system shall not cause the computers to overload.

- Rationale: The system should not overload the users' computers.
- Fit Criteria: The hardware running the system is under normal temperature 65°C.
- Priority: Medium

HS2 The system shall not affect users' physical and mental health.

- Rationale: The system must not harm users' health and safety.
- Fit Criteria: 99% of the users feel comfortable using the system in various situations.
- Priority: Critical

5.5 Interfaces

5.5.1 Interface Provided By The System

User Interface This is the primary medium most users will interact with and its design is based on the popular UI design pattern (like the Ant Design). Features include video display windows, control buttons (e.g. mute, camera toggle), participant lists, settings panels, etc.

5.5.2 Hardware Interfaces

Camera ~~and Microphone~~ To capture video ~~and audio~~ from users' sides, the system interfaces with hardware drivers and OS-level APIs.

Network For managing connections, the system might interact with the underlying operating system's networking stack.

5.6 Detailed Usage Scenarios

The real-time video conference application leverages its underlying infrastructure, which includes components like the ~~WebRTC~~ Real-Time Communication Client, SFU, and API server, to offer a seamless and enriched teaching and learning experience. The instructor's video rendered by machine learning annotations acts as a real-time guide, enhancing the practitioners' understanding and mastery of Tai Chi.

The detailed usage for the two main users, Tai Chi practitioners and instructors, is listed as follows:

- Practitioners initiate their sessions through the application's user interface. Upon joining, the ~~WebRTC~~ Real-Time Communication Client, with the assistance of STUN and TURN servers, ensures a stable connection for the participant. As they enter the session, they receive the instructor's video feed that's routed through the Selective Forwarding Unit (SFU). This media stream, having been extracted from the SFU server, is then relayed to the API server via a socket connection. The API server, having integrated annotation capabilities, renders the instructor's video by overlaying instructional annotations. This enhanced media stream, filled with real-time annotations detailing specific Tai Chi movements, is then returned to the practitioner's side. Throughout this process, practitioners have a passive role, in that they do not upload videos but solely benefit from the real-time instruction and the added value of the annotations on the instructor's video.

- Instructors initiate their teaching sessions via the same user interface. Their ~~WebRTC~~ Real-Time Communication Client, in tandem with the STUN and TURN servers, establishes a connection. As instructors demonstrate Tai Chi forms and movements, their video stream is managed by the Real-time Communication Engine, passing session setup details via the Signaling Server using the Session Description Protocol (SDP). Their live video then gets channelled through the SFU, ensuring optimal distribution to all attending practitioners. Post this, the instructor's video stream undergoes a transformation at the API server where it's augmented with instructional annotations. These annotations might highlight specific postures, movement trajectories, or key focus areas, thereby enhancing the instructional value of the video. This annotated video is then disseminated in real-time to all the prac-

titioners, ensuring they have a clear and enhanced understanding of the Tai Chi instructions being provided.

5.7   Use Case Diagram
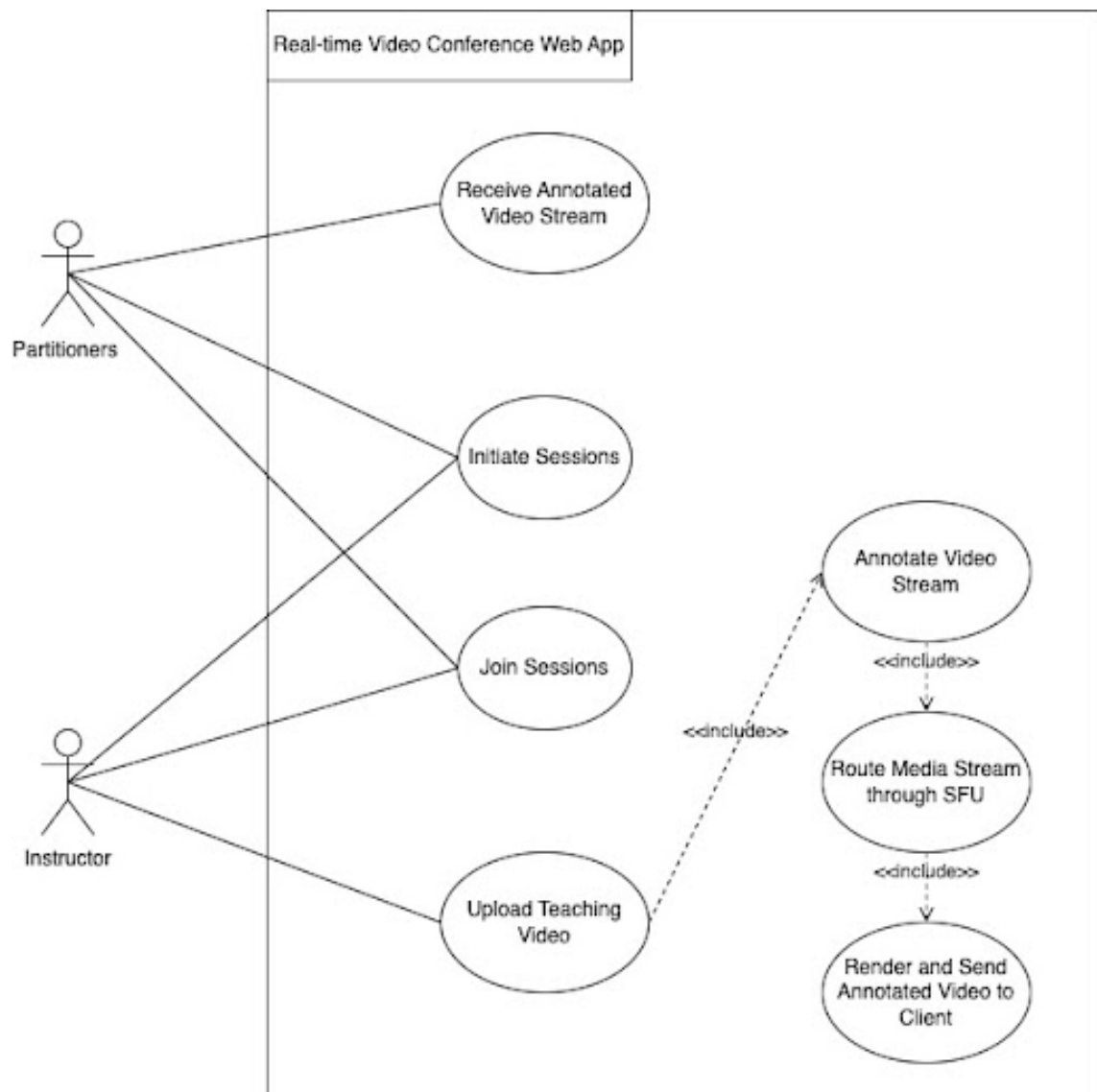
5.8 Prioritization

The priorities of functionalities and behaviours are classified into four groups: Critical, high, medium and low.

- Critical:

    FRs: FR1 5.3, FR2 5.3, FR3 5.3, FR4 5.3, FR5 5.3, FR6 5.3, FR7 5.3, FR8 5.3, FR9 5.3, FR11 5.3, FR12 5.3, FR13 5.3

    NFRs: SR1 5.4.6, SR2 5.4.6, SR3 5.4.6, CR 5.4.7, LR1 5.4.8, HS2 5.4.9

- High:

    FRs: FR10 5.3,

    NFRs: PR2 5.4.3, OE1 5.4.4, OE2 5.4.4, OE3 5.4.4

- Medium:

    FRs: Not applicable.

    NFRs: LF1 5.4.1, UH1 5.4.2, UH2 5.4.2, PR1 5.4.3, PR3 5.4.3, PR4 5.4.3, PR5 5.4.3, PR6 5.4.3, MS1 5.4.5, HS1 5.4.9

- Low:

    FRs: Not applicable.

    NFRs: MS2 5.4.5

5.9 Verification and Acceptance Criteria

The high-level verification and acceptance criteria outlined below are intended to ensure that the system meets the specified requirements and presents desired behaviours.

5.9.1 Integration of Data Streams

- Verification:

    - Conduct module testing to ensure that video and audio data streams are transmitted and received without significant loss of quality.

- Perform system testing to verify that data streams maintain integrity during real-time communication sessions.

- Conduct static analysis to identify and mitigate potential issues related to data stream integrity.

- Acceptance Criteria:

  - Video streams should maintain a minimum resolution of 720p with minimal artifacts.

  - Audio streams should have clear and intelligible sound quality.

  - During system testing, no critical issues related to data stream integrity should be identified.

5.9.2 Accurate Response to Client Requests and Annotation Configurations

- Verification:

  - Perform integration testing to ensure that the system accurately responds to client requests and annotation configurations.

  - Conduct user acceptance testing with practitioners to validate the system's responsiveness to annotation preferences.

- Acceptance Criteria:

  - The system should accurately apply annotation configurations specified by practitioners.

  - Practitioners should confirm that their requested annotation preferences are correctly implemented during user acceptance testing.

5.9.3 Real-Time Processing

- Verification:

  - Conduct system testing to assess the system's ability to process video streams in real-time.

  - Evaluate the system's performance under various load conditions, ensuring real-time processing is maintained.

- Acceptance Criteria:

  - The system should consistently process video streams with minimal delay, aiming for a latency of less than 500 ms while maintaining a minimum frame rate of 10 fps.
  - Under typical load conditions, the system should sustain real-time processing without significantly degrading the performance.

5.9.4 Starting/Joining Call Sessions

- Verification:

  - Perform integration testing to ensure that both instructors and practitioners can start/join call sessions without encountering critical errors.
  - Conduct user acceptance testing to assess the ease of starting/joining call sessions.

- Acceptance Criteria:

  - Users should be able to start/join call sessions without experiencing crashes or disruptions.
  - User acceptance testing should confirm that the process of starting/joining a call session is straightforward and user-friendly.

5.9.5 Start Annotation Pipeline Based on Need

- Verification:

  - Conduct module testing to verify that annotation pipelines are initiated based on practitioner preferences.
  - Evaluate the system's ability to dynamically start and stop annotation pipelines during live sessions.

- Acceptance Criteria:

  - Annotation pipelines should be initiated and terminated based on practitioner preferences.
  - Users should have a seamless experience with annotations starting, stopping or switching as needed during live sessions.

5.10 Traceability Matrix

Please refer to 3 and 4.

	FR1	FR2	FR3	FR4	FR5	FR6	FR7
LF1	X	X	X			X	
UH1	X	X	X			X	
UH2	X	X	X			X	
PR1	X	X				X	
PR2				X	X		
PR3	X	X		X	X		
PR4	X	X	X			X	
PR5							
PR6							
OE1	X	X	X				
OE2	X	X	X				
OE3	X	X	X				
MS1	X	X	X	X	X	X	X
MS2	X	X	X	X	X	X	X
SR1				X	X		
SR2	X	X					
SR3	X	X	X	X	X	X	X
CR1	X	X					
LR1	X	X	X	X	X	X	X
HS1	X	X					
HS2	X	X					

Table 3: Traceability Matrix Showing the Connections Between Functional Requirements and Non-functional Requirements

	FR8	FR9	FR10	FR11	FR12	FR13
LF1						
UH1						
UH2						
PR1						
PR2						X
PR3					X	X
PR4						
PR5			X		X	X
PR6			X		X	X
OE1						
OE2						
OE3						
MS1	X	X	X	X	X	X
MS2	X	X	X	X	X	X
SR1					X	X
SR2						
SR3	X	X	X	X	X	X
CR1						
LR1	X	X	X	X	X	X
HS1						
HS2						

Table 4: Traceability Matrix Showing the Connections Between Functional Requirements and Non-functional Requirements

5.11   Likely/Unlikely Changes

Likely Changes

- The application could be extended in the future to offer mobile compatibility, providing users accessibility on both desktop and mobile platforms.

- A login and access control system may need to be implemented to restrict video streaming capabilities to authorized users only, in order to adhere to security best practices.

- Additional features may need to be incorporated into future iterations

based on discoveries from real-world usage and direct end user feedback.

- Web application compatibility may be offered to provide increased reach and usability across different devices and operating systems.

- The initial system load capacity will be set conservatively due to unknown demand. As user adoption increases over time, load capacity can be scaled up to support larger volumes of traffic.

- The Real-Time Processing requirement might change as no metric has been provided from the stakeholders.

- The application may be extended in the future to offer functionality to allow practitioners to stream their videos and audios.

Unlikely Changes

- The application will be designed to comply with all applicable laws and regulations[1][2][4]. User personal information will only be collected or used with explicit consent.

- Given this is a student capstone project, the primary stakeholder and client is unlikely to change.

- As a student capstone project, schedule and budget constraints are relatively fixed and not expected to change significantly.

- The user input to the application and the outputs to users should remain unchanged, as the overall objective is determined.

- A key priority in the system design will be ease of use and learnability, given the target user base has less technical expertise.

- The application will implement access controls to ensure users only view data they are properly authorized to access.

- Annotation generation will occur in a dedicated backend system.

- The application will allow configurable annotation types displayed to practitioners.

- The application is intended to operate on Windows, Linux and MacOS platforms, but not limited to running on these systems natively.

- Secure data transmission and storage will be implemented to safeguard user data and mitigate data leakage risks.

6   Appdendix

6.1   Reflection

Writing the Software Requirements Specification was a crucial phase that laid the foundational blueprint for our entire development process. During this stage, we delved deeply into understanding and defining the core functionalities, non-functional requirements, and the overall scope of our WebRTC-based video streaming application tailored for Tai Chi instruction. We collaboratively identified the key features necessary for a successful implementation, such as real-time video streaming, user-centred design, and machine learning integration for annotations. In doing so, we faced the challenge of translating broad concepts and goals into clear, detailed specifications that could guide our development efforts effectively.

One of the most significant aspects of this phase was the risk and mitigation analysis. We identified potential risks, such as insufficient technical knowledge in specific areas like WebRTC and real-time streaming protocols, and the challenge of integrating machine learning models for real-time annotations. As a team, we not only pinpointed these risks but also developed strategies to mitigate them, which included allocating time for learning and experimentation. Moreover, the process of creating the SRS reinforced the importance of clear communication and collaboration amongst ourselves. We learned to articulate our thoughts and ideas precisely, ensuring that every team member had a unified understanding of the project's objectives and requirements.

References

[1]   L. S. Branch. Consolidated federal laws of Canada, Personal Information Protection and Elect URL: https://laws-lois.justice.gc.ca/ENG/ACTS/P-8.6/page-1.html#h-416888. (accessed: Oct.04.2023).

[2] Government of Canada. Canada's anti-spam legislation - Home," ised-isde.canada.ca, Mar. 19 URL: https://ised-isde.canada.ca/site/canada-anti-spam-legislation/en. (accessed: Oct.04.2023).

[3] Government of Canada. Criminal Code, Justice.gc.ca, 2019. URL: https://laws-lois.justice.gc.ca/eng/acts/C-46/. (accessed: Oct.04.2023).

[4] Government of Canada. Criminal Code, Justice.gc.ca, 2019. URL: https://laws-lois.justice.gc.ca/eng/acts/C-46/. (accessed: Oct.04.2023).

[5] statcounter. Desktop Operating System Market Share Canada Sept 2022 - Sept 2023. URL: https://gs.statcounter.com/os-market-share/desktop/canada. (accessed: Oct.04.2023).

[6] Petroc Taylor. Average internet download speed in Canada as of March 2018, by province. URL: https://www.statista.com/statistics/544943/average-data-download-speed-canada-province/. (accessed: Oct.04.2023).

[7] Petroc Taylor. Average mobile internet download and upload speeds in selected cities in Cana URL: https://www.statista.com/statistics/862704/canada-average-mobile-download-and-upload-speeds-by-city/. (accessed: Oct.04.2023).