

Software Requirements Specification

for

AI/ML QMS Sentiment Analysis

Version 1.0 approved

Prepared by Joshua Soteras
Brandon Lopez
Jorge Arias
Javier Gonzalez
Kenneth Castro
Walter Najera
Andres Quezada
Johny Vu
Harshaun Khehra
Darrin Du

Leidos QTC Health Services

November 31, 2024

Table of Contents

Table of Contents.....	pg #
Revision History.....	pg #
1. Introduction.....	pg #
1.1. Purpose.....	pg #
1.2. Intended Audience and Reading Suggestions.....	pg #
1.3. Product Scope.....	pg #
1.4. Definitions, Acronyms, and Abbreviations	pg #
1.5. References.....	pg #
2. Overall Description.....	pg #
2.1. System Analysis.....	pg #
2.2. Product Perspective.....	pg #
2.3. Product Functions.....	pg #
2.4. User Classes and Characteristics.....	pg #
2.5. Operating Environment.....	pg #
2.6. Design and Implementation Constraints.....	pg #
2.7. User Documentation.....	pg #
2.8. Assumptions and Dependencies.....	pg #
2.9. Apportioning of Requirements.....	pg #
3. External Interface Requirements.....	pg #
3.1. User Interfaces.....	pg #
3.2. Hardware Interfaces.....	pg #
3.3. Software Interfaces.....	pg #
3.4. Communications Interfaces.....	pg #
4. Requirements Specification.....	pg #
4.1. Functional Requirements.....	pg #
4.2. External Interface Requirements.....	pg #
4.3. Logical Database Requirements.....	pg #
4.4. Design Constraints.....	pg #
5. Other Nonfunctional Requirements.....	pg #
5.1. Performance Requirements.....	pg #
5.2. Safety Requirements.....	pg #
5.3. Security Requirements.....	pg #
5.4. Software Quality Attributes.....	pg #
5.5. Business Rules.....	pg #
6. Legal and Ethical Considerations.....	pg #
Appendix A: Glossary.....	pg #
Appendix B: Analysis Models.....	pg #
Appendix C: To Be Determined List.....	pg #

Revision History

Name	Date	Reason For Changes	Version
Joshua Sotersa	November 31st	Added Information	1.00

1. Introduction

1.1 Purpose

This is the Software Requirements Specification (SRS) (aka Software Requirements Document (SRD)) for the initial release of the AI/ML Sentiment Analysis Project. This document will cover the application's standalone capabilities and their respective requirements, as well as the requirements for the database and web API used for the application's synchronization feature.

1.2 Intended Audience and Reading Suggestions

This SRS contains user requirements (high-level abstract requirements) and system requirements (detailed description of the software's functions). Section 2 describes the need for the system with brief and high-level descriptions of functions. Section 3 provides high-level descriptions of the system's interactions with external interfaces. Section 4 contains functional system requirements. Section 5 contains non-functional requirements. Section 6 contains legal and ethical considerations.

Managers interested in using this document to plan a bid for the system or for planning the development process should read sections 2, 3, 5, and 6.

System customers involved in requirements validation should read sections 2, 3, and 5.

System engineers who need to understand what system is to be developed should read all sections.

System test engineers should use the requirements in sections 4 and 5 for verification testing. If involved in validation testing, they should also read sections 2 and 3.

System maintenance engineers who need to understand the system's parts and how they work together should read sections 2, 3, 4, and 5.

All readers should be aware of section 6.

1.3 Product Scope

The project aims to develop an AI/ML-powered Quality Management System (QMS) that analyzes survey responses related to examinee and staff appointments. The system will determine positive or negative ratings, generate key performance indicator (KPI) data, and display data on a dashboard. Additionally, it will create incident tickets for all negative responses with ratings less than 3.

As a bonus feature, the system will incorporate Generative AI chat capabilities to provide aggregate data trends and insights while ensuring compliance with HIPAA and protecting Personally Identifiable Information (PII)

1.4 Definitions, Acronyms, and Abbreviations

NLP - Natural Language Processing

HIPAA - Health Insurance Portability and Accountability Act of 1996

SQL - Structured Query Language

roBERTa - Robustly Optimized BERT Approach

BERT - Bidirectional Encoder Representations from Transformers

1.5 References

[1] Section 508 Guide - Section 508 guide for compliance that outlines the tools needed for accessible web content

[2] **Database Transaction Unit (DTU)** - metric used by Azure SQL Database to measure the overall performance resources available to the sentiment analysis application, which include CPU processing, memory, and read/write input/output operations needed to handle data

[3] Service Level Agreements (SLA) for Online Services

This document outlines Microsoft's service level commitments for online services, including uptime guarantees, support response times, and compensation for service disruptions. Uptime for free tiers is not included.

[4] **Service Level Agreement (SLA)**: A document defining the expected service level between a service provider and a client. It outlines the metrics by which the service is measured, such as availability, performance, and the responsibilities of both parties.

[5] Azure Pricing Calculator

A Microsoft-provided tool that estimates the costs associated with using Azure services. The calculator allows users to calculate pricing based on selected Azure products, usage estimates, and other variables.

[6] **Azure SQL Database Recovery / Azure App Services Recovery**

The first document provides an overview of Active Geo-Replication for Azure SQL Database, which allows for the creation of secondary databases in different regions for recovery. The second document details Azure App Service's reliability, including automatic scaling, load balancing, and recovery options.

2. Overall Description

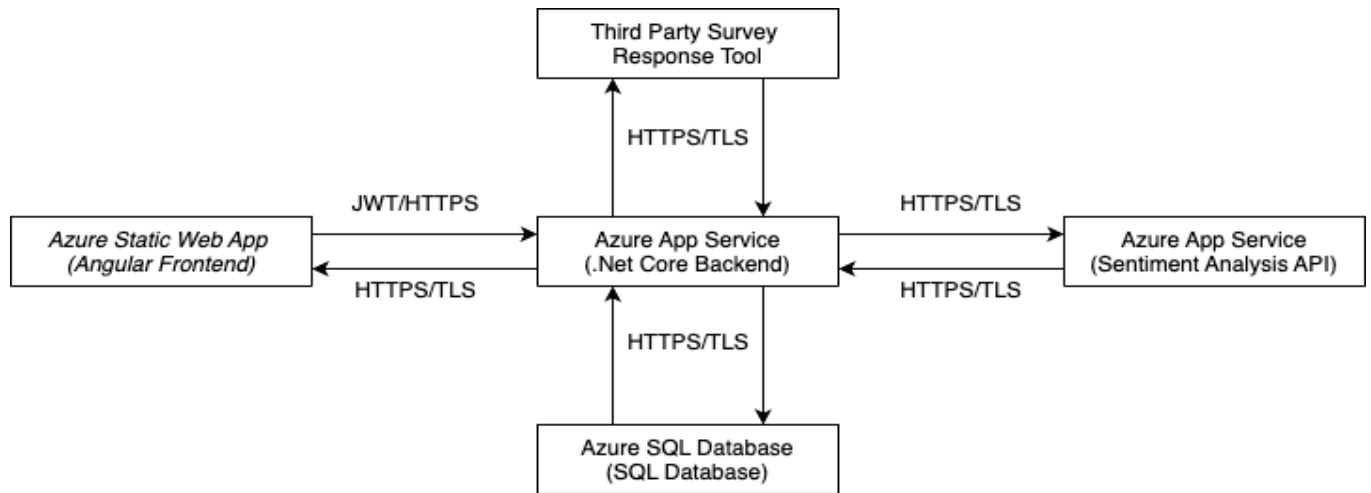
2.1 System Analysis

In order to help Leidos QTC Health Services to improve services, surveys are import to gather patients/examinees' experiences. With machine learning and AI, surveys will expand and be more in detail; instead of a star rating system, user's will be able to write responses which will be analyzed as positive, negative, neutral. This process of analyzing text is called Sentiment Analysis in which companies use to better their services.

The project will incorporate a user interface for both patients and Leidos employees. For employees, they will perceive a KPI (key performance indicator) interface in which they will be able to see in what areas they can improve on and the consensus of the reviews.

This will be implemented with a machine learning model, backend service, frontend frame world and deployment strategy.

2.2 Product Perspective



2.3 Product Functions

Sentiment Analysis - Analyze survey responses (scores and written) to determine positive or negative ratings. Generate incident tickets for responses with ratings less than 3.

Dashboard and Reporting - Display KPI data on a user-friendly dashboard. Provide insights on common low and high scores, top negative and positive survey trends, issue categories, and resolution categories.

Incident Management - Automatically create alert and track incident tickets for negative responses. Bonus: Provide resolution suggestions for similar negative scores.

Generative AI Chat - Enable users to inquire about aggregate data trends and insights. Ensure compliance with HIPAA and protection of Personally Identifiable Information (PII)

2.4 User Classes and Characteristics

Patients

Patients and examinees that are associated with Leidos QTC Health Services will only be able to access a survey from their perspective. Necessary to navigate through a webpage, read, is a necessary expertise. Patients will have their own login page and will be accessed through email and password information. By default they can send an anonymous survey through a specific part of the web interface if they do not want to login to their account.

Employees

Employees that are under Leidos QTC Health Services will see a different interface which will display the key performance indicators (KPI) and survey responses and the following responses. Necessary employee information is needed to login within the application to view the KPI interface. This login page will ask for information such as employee ID, email, and password.

2.5 Operating Environment

The software will be hosted by Azure cloud services. Azure is a cloud computing platform from Microsoft that will allow the Leidos QTC to build, run and manage our project application. These include the SQL database needed to store the data, hosting the machine learning model for sentiment analysis and Angular as a framework. All resources will be managed by Azure.

2.6 Design and Implementation Constraints

Safety and Security: Test Data

Because of HIPAA regulations and compliance, using and accessing patient data is out of our scope. Thus the project will be utilizing pseudo data in .csv format. This might impact the accuracy and effectiveness of our sentiment analysis due to limited and accurate data provided.

2.7 User Documentation

As explained in section 2.4, the application should be simple enough to use that training documentation or training videos are generally not needed to use the application's features.

However, the "customize UI" feature may require documentation on how color schemes work; whether or not this will be needed will be resolved later in development (it depends on whether the application can explain how the colors work in its own menu and how the colors for components are even decided). Such documentation would not be excessive, and would likely only require a section in the project's README file.

If time permits, documentation demonstrating the functions listed in section 2.3 may be developed. If such documentation will be created, it will likely either be in the project's README file or in a PDF file.

2.8 Assumptions and Dependencies

If Azure doesn't fit the standards needed by the company or the application, then there exist other cloud services such as AWS and Google Cloud which hold the necessary tools to host and deploy our application. The Continuous Integration / Development would then be switched to Jenkins, a third party software used for automation testing and development.

Furthermore, the application plans to use a deep learning model fine tuned by Cardiff NLP that specializes in Natural Language Processing to perform Sentiment Analysis. This model is provided by Hugging Face; an open source platform and community for machine learning (ML) and data science that provides pre-trained models and over 30,000 datasets for training models. In case the chosen model will not be available, the project will have to manually fine-tune (train) a model for sentiment analysis.

2.9 Apportioning of Requirements

Voice surveys in which users (patients) will be answer survey questions through their mobile device will be delayed until all necessary project milestones (essential requirements) are met (see sections 4 and 5)

3. External Interface Requirements

3.1 User Interfaces



User Interface Compliance with Section 508

The system's user interface will comply with Section 508 of the Rehabilitation Act, assuring accessibility for all users, including those with disabilities. All of the interactive elements, such as navigation buttons and data input fields, will be accessible via keyboard (both mobile and desktop advices)

User Interface will be similar and accessible in either mobile, tablet, laptop, or desktop devices, meaning that the interface will adapt to the users screen resolution or constraints.

Surveys

For patient surveys, the system collects data from customer surveys, which have ten questions. Some questions are multiple-choice with options ranging from 1-5, and one question requires a text response. Thus, a user interface will be displayed with a designated text box and selectable multiple choice answers and a submission button. Indicators will display to the user on which answers are selected and which questions have been answered.

Dashboard

The dashboard will continuously monitor and display the overall sentiment, which includes both scores and comments. The dashboard can also monitor and display the sentiment of the comments and tell us what percentage of the comments are positive, neutral, and negative. Thus a UI will be used to display these key performance indicators through the use of graphs, charts, color, and analysis timelines.

3.2 Hardware Interfaces

Hardware that the AI/ML Sentiment Analysis tool will interact with are mobile devices, laptops, tablets, desktops and or any device with a display that may access the web application. The application will interact with keyboards, touch screens (touch screen keyboards and selection), and mice to take user input.

Information will be transferred through HTTPS protocols, thus needing to be connected to the internet is essential,

3.3 Software Interfaces

ASP.NET Core v8.0.0 is used by Leidos QTC Health Services as their standard backend framework thus the team decided to use this backend framework to easily integrate the software within their ecosystem.

We will be using roBERTa transformer model to perform Sentiment Analysis which was fine-tuned by CardiffNLP which can be found within Hugging Face. This model will interact with the survey data (specifically written responses) to extract information and perform ratings of positive, neutral, and negative

3.4 Communications Interfaces

The project will primarily use HTTP to handle communication between the client and server, which is built on top of a cloud-based infrastructure (Azure). CSV files are utilized to store and transfer structured data, such as our client's responses, while JSON format is used as a lightweight method of transferring the data between Azure and client's devices. Azure provides tools in order to encrypt and decrypt the data when necessary. Additionally, robust authentication and authorization mechanisms shall be implemented in order to protect sensitive data. Monitoring and logging mechanisms shall be implemented to track performance, identify issues, and analyze usage patterns in order to notify staff on the system's conditions.

4. Requirements Specification

4.1 Functional Requirements

Sentiment Analysis - The application shall: analyze survey responses (scores and written) to determine positive or negative ratings. Generate incident tickets for responses with ratings less than 3.

Dashboard and Reporting - The application shall: display KPI data on a user-friendly dashboard. Provide insights on common low and high scores, top negative and positive survey trends, issue categories, and resolution categories.

Incident Management - The application shall: automatically create alerts and track incident tickets for negative responses. Bonus: Provide resolution suggestions for similar negative scores.

Generative AI Chat -The application shall: enable users to inquire about aggregate data trends and insights. Ensure compliance with HIPAA and protection of Personally Identifiable Information (PII)

4.2 External Interface Requirements

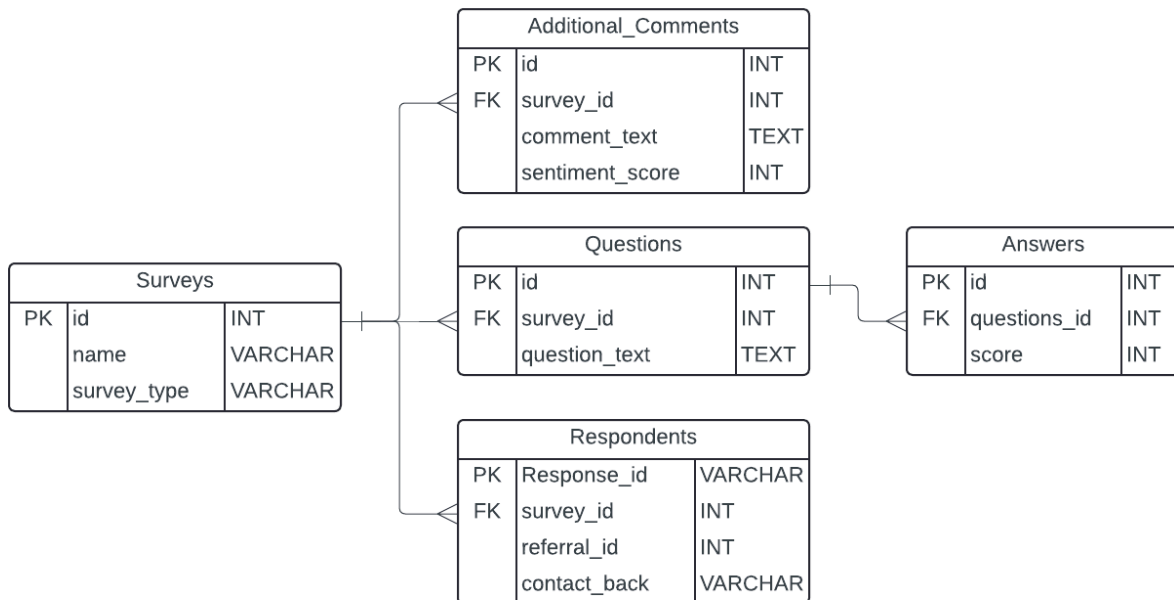
Machine Learning Model

A deep learning model called a transformer is necessary for sentiment analysis. To analyze human emotion within survey data and determine sentiment, the model needs to be trained by a vast amount of survey responses. Furthermore, to adjust or enhance the ratings, the model may be fine tuned with more or newer data. To train this model requires hardware which is offered by Azure Machine Learning.

Angular HTTPClient

Angular's HttpClient service is used to send data to the ASP.NET Core backend. HttpClientModule is needed within the Angular app module In ASP.NET Core, create an API controller that will handle the data sent from Angular. This Controller will receive data through the HTTP POST request and use Entity Framework Core (ASP.NET Core feature for grabbing and accessing data)

4.3 Logical Database Requirements



- **Surveys Table**
 - We will store the Surveys' ID, name, and survey type. This is essential for identifying which survey the data belongs to.
- **Questions Table**
 - Each question is linked to a survey via survey_id, and the actual question text is stored in question_text.
 - This design works well to associate questions with specific surveys, ensuring flexibility when adding new surveys or questions.
- **Respondents Table**
 - Tracks each respondent's Response_id along with their referral_id and is linked to a survey via survey_id.

- Stores if the respondent wants to be contacted back
- **Additional Comments Table**
 - Stores any extra comments (comment_text) and links to a survey_id and a sentiment_score
 - This provides a way to store open-ended responses (like comments or feedback) and sentiment analysis results.
- **Answers Table**
 - Stores the questions_id (foreign key linking to the Questions table) and score, representing each respondent's answer
 - This table stores the actual numeric responses for the questions asked in the survey.

POST	https://api.aiapi.com/analysis/request				
Request	Consumes	application/json			
		Name	Type	Required	Description
		comment_id	integer	Y	id for comment
		survey_id	integer	Y	id for survey
		comment_text	string	Y	comment for analysis
	Example	<pre>{ "comment_id" : 1, "survey_id" : 2, "comment_text" : "Great review" }</pre>			

SQL Database

- SQL database will be deployed to the Azure SQL Database server
- SQL database will send and receive requests from the backend
- The backend will manage read/write operations securely using Azure SQL Firewall, encrypted connections, and Azure Data Masking

4.4 Design Constraints

Anticipated Cost Growth

The anticipated cost growth over the next three years will be driven by two main factors: projected Azure pricing adjustments and the expected increase in user traffic. We can conservatively estimate annual Azure price increases at around 5%, reflecting industry trends. Additionally, as user traffic grows each year, the demand for resources will increase proportionally, leading to a rise in overall costs. Factoring in both Azure pricing adjustments and traffic growth will provide a more accurate forecast of future expenses.

Azure SQL Database Basic Tiers

The Basic Tier offered by Azure for their SQL Database services could pose an issue overtime, as while the Basic tier is ideal for handling small datasets for the development and testing phases of the project, in a working environment the Basic tier would reach max capacity overtime. As the project reaches more people, there's an increase in user traffic which correlates to increased demand for resources to be allocated to saving their responses, which is not what the Basic tier designed to handle.

5. Other Nonfunctional Requirements

5.1 Performance Requirements

Performance - The system should process and analyze survey responses in real-time. Dashboard and reports should load within 2 seconds.

Security -The system should ensure data encryption in transit and at rest. Implement role-based access control (RBAC) to protect sensitive information.

Scalability - The system should handle increasing volumes of survey data without performance degradation. Support for future enhancements and additional features. These include more than 10,000 examinees and more than 25 Quality Management Administrators.

Usability - The user interface should be intuitive and easy to navigate. Provide comprehensive user documentation and training materials.

Compliance -The system should ensure compliance with HIPAA regulations. Protect Personally Identifiable Information (PII) at all times

5.2 Safety Requirements

Data Protection - The application will have safeguards in place to prevent data corruption or loss. Some safeguards include regular backups, data validation, and error handling.

Regular backups - Implement automated backup procedures to create copies of critical data at regular intervals.

Data Validation- We will Implement data validation checks to ensure that incoming data is consistent with expected formats and ranges. This helps prevent corrupt or invalid data from entering the system.

Error Handling - We will develop error handling mechanisms to detect and handle errors such as data input errors. The errors will be logged for analysis and debugging.

5.3 Security Requirements

All data, especially sensitive information like personally identifiable or patient health information, should be encrypted at rest and in transit. This ensures that raw data is protected against unauthorized access via techniques such as advanced encryption standard(AES-256) encryption & secure sockets layer/transport layer security(SSL/TLS) encryption.

Although the data will use user IDs rather than personal information like names, it is still important to implement data anonymization techniques to protect individual responses that will be analyzed by sentiment analysis.

5.4 Software Quality Attributes

Ease of Use is prioritized over Ease of Learning, as the primary user group consists of experienced professionals. The user group are managers who will understand the survey data and the user interface is easy enough to use that there is little to no learning curve.

Furthermore, for increasing storage and resource consumption (such as training and fine tuning the transformer model) Azure is scalable depending on the applications needs. With SQL Database (Azure SQL Database) storage can be increased through a paywall of different tiers. More storage, the higher the cost.

Monitor

5.5 Business Rules

Licenses Compliance - for the Software and Ensuring that tools are safe to use. This includes compliance with open-source licenses, proprietary agreements, and enterprise-level contracts. Furthermore Software License must permit commercial use and integration without violating intellectual property laws or creating legal risk

Source Verification - tools and libraries must originate from reputable organizations with a proven track record of delivering secure and reliable software.

Use - Only Quality Management Administrators will be able to access the KPI (key performance indicator) page through their login credentials which will display survey and sentiment analysis. Patients or examinees will not be allowed to view this page.

6. Legal and Ethical Considerations

NIST (National Institute of Standards and Technology)

- NIST frameworks, especially NIST SP 800-53 and NIST Cybersecurity Framework, provide guidelines for securing federal information systems.

- Risk management and the implementation of cybersecurity controls across different domains (access control, incident response, etc.).
- Implement comprehensive security controls, conduct risk assessments, and apply continuous monitoring to ensure systems are secure.

ISO 27001 (International Organization for Standardization)

- ISO 27001 provides a framework for an Information Security Management System (ISMS).
- Protecting sensitive company and customer information through risk management and continuous improvement.
- We must define an ISMS, perform risk assessments, implement controls to mitigate risks and undergo audits to ensure compliance.

HIPAA (Health Insurance Portability and Accountability Act)

- HIPAA is a U.S. federal law that mandates the secure handling of PHI. HIPAA establishes guidelines for healthcare providers, insurers, and their business associates to ensure the confidentiality, integrity, and availability of PHI.
- Requirements
 - Implementation of safeguards to protect PHI (physical, administrative, and technical).
 - Compliance with privacy rules, including patient consent for sharing PHI.
 - Reporting data breaches involving PHI.

Section 508 Compliance (Rehabilitation Act of 1973)

- Section 508 requires federal agencies to make their electronic and information technology accessible to people with disabilities.
- Ensuring that websites, software, and other technology are accessible.
- Companies must implement accessibility features in software, websites, and electronic documents, ensuring compatibility with assistive technologies like screen readers.
- [1] [Accessible Web Guide](#)

Appendix A: Glossary.

TBA

Appendix B: Analysis Models

TBA

Appendix C: To Be Determined List

- Appendix A Glossary
- Appendix B Glossary