

# Software Requirements Specification for Software Engineering: subtitle describing software

Team #11, OKKM Insights

Mathew Petronilho

Oleg Glotov

Kyle McMaster

Kartik Chaudhari

October 7, 2024

# Contents

<b>1</b>	<b>Purpose of the Project</b>	<b>vii</b>
1.1	Goals of the Project . . . . .	vii
1.1.1	High Data Accuracy . . . . .	vii
1.1.2	Ease of use . . . . .	vii
1.1.3	Minimizing Cost to Analyze Images . . . . .	vii
1.1.4	Results Returned Within Appropriate Timeframe . . . . .	viii
1.1.5	High System Reliability and Accessibility . . . . .	viii
<b>2</b>	<b>Stakeholders</b>	<b>viii</b>
2.1	Client . . . . .	viii
2.2	Customer . . . . .	viii
2.3	Other Stakeholders . . . . .	viii
2.4	Hands-On Users of the Project . . . . .	viii
2.5	Personas . . . . .	ix
2.6	Priorities Assigned to Users . . . . .	ix
2.7	User Participation . . . . .	ix
2.8	Maintenance Users and Service Technicians . . . . .	ix
<b>3</b>	<b>Mandated Constraints</b>	<b>ix</b>
3.1	Implementation Environment of the Current System . . . . .	ix
3.2	Partner or Collaborative Applications . . . . .	ix
3.3	Off-the-Shelf Software . . . . .	x
3.4	Anticipated Workplace Environment . . . . .	x
3.5	Schedule Constraints . . . . .	x
3.6	Budget Constraints . . . . .	x
<b>4</b>	<b>Naming Conventions and Terminology</b>	<b>xi</b>
4.1	Glossary of All Terms, Including Acronyms, Used by Stakeholders involved in the Project . . . . .	xi
<b>5</b>	<b>Relevant Facts And Assumptions</b>	<b>xi</b>
5.1	Relevant Facts . . . . .	xi
5.2	Business Rules . . . . .	xi
5.3	Assumptions . . . . .	xi

<b>6</b>	<b>The Scope of the Work</b>	<b>xi</b>
6.1	The Current Situation . . . . .	xi
6.2	The Context of the Work . . . . .	xi
6.3	Work Partitioning . . . . .	xi
6.4	Specifying a Business Use Case (BUC) . . . . .	xi
<b>7</b>	<b>Business Data Model and Data Dictionary</b>	<b>xii</b>
7.1	Business Data Model . . . . .	xii
7.2	Data Dictionary . . . . .	xii
<b>8</b>	<b>The Scope of the Product</b>	<b>xii</b>
8.1	Product Boundary . . . . .	xii
8.2	Product Use Case Table . . . . .	xii
8.3	Individual Product Use Cases (PUC's) . . . . .	xii
<b>9</b>	<b>Functional Requirements</b>	<b>xii</b>
9.1	Functional Requirements . . . . .	xii
<b>10</b>	<b>Look and Feel Requirements</b>	<b>xii</b>
10.1	Appearance Requirements . . . . .	xii
10.2	Style Requirements . . . . .	xiii
<b>11</b>	<b>Usability and Humanity Requirements</b>	<b>xiii</b>
11.1	Ease of Use Requirements . . . . .	xiii
11.2	Personalization and Internationalization Requirements . . . . .	xiv
11.3	Learning Requirements . . . . .	xiv
11.4	Understandability and Politeness Requirements . . . . .	xiv
11.5	Accessibility Requirements . . . . .	xiv
<b>12</b>	<b>Performance Requirements</b>	<b>xiv</b>
12.1	Speed and Latency Requirements . . . . .	xiv
12.2	Safety-Critical Requirements . . . . .	xvi
12.3	Precision or Accuracy Requirements . . . . .	xvi
12.4	Robustness or Fault-Tolerance Requirements . . . . .	xvii
12.5	Capacity Requirements . . . . .	xvii
12.6	Scalability or Extensibility Requirements . . . . .	xviii
12.7	Longevity Requirements . . . . .	xviii

<b>13 Operational and Environmental Requirements</b>	<b>xviii</b>
13.1 Expected Physical Environment . . . . .	xviii
13.2 Wider Environment Requirements . . . . .	xviii
13.3 Requirements for Interfacing with Adjacent Systems . . . . .	xviii
13.4 Productization Requirements . . . . .	xviii
13.5 Release Requirements . . . . .	xviii
<b>14 Maintainability and Support Requirements</b>	<b>xix</b>
14.1 Maintenance Requirements . . . . .	xix
14.2 Supportability Requirements . . . . .	xix
14.3 Adaptability Requirements . . . . .	xix
<b>15 Security Requirements</b>	<b>xix</b>
15.1 Access Requirements . . . . .	xix
15.2 Integrity Requirements . . . . .	xxi
15.3 Privacy Requirements . . . . .	xxi
15.4 Audit Requirements . . . . .	xxii
15.5 Immunity Requirements . . . . .	xxii
<b>16 Cultural Requirements</b>	<b>xxii</b>
16.1 Cultural Requirements . . . . .	xxii
<b>17 Compliance Requirements</b>	<b>xxiii</b>
17.1 Legal Requirements . . . . .	xxiii
17.2 Standards Compliance Requirements . . . . .	xxiv
<b>18 Open Issues</b>	<b>xxiv</b>
<b>19 Off-the-Shelf Solutions</b>	<b>xxiv</b>
19.1 Ready-Made Products . . . . .	xxiv
19.2 Reusable Components . . . . .	xxiv
19.3 Products That Can Be Copied . . . . .	xxiv
<b>20 New Problems</b>	<b>xxv</b>
20.1 Effects on the Current Environment . . . . .	xxv
20.1.1 Data Privacy and Security Concerns . . . . .	xxv
20.1.2 Environmental Footprint . . . . .	xxv
20.2 Effects on the Installed Systems . . . . .	xxv
20.2.1 Integration Challenges . . . . .	xxv

20.2.2	Legacy Systems Compatibility . . . . .	xxvi
20.3	Potential User Problems . . . . .	xxvi
20.3.1	Usability Issues . . . . .	xxvi
20.3.2	Training and Onboarding . . . . .	xxvi
20.3.3	Compensation and Incentives . . . . .	xxvi
20.3.4	Technical Support . . . . .	xxvii
20.4	Limitations in the Anticipated Implementation Environment That May Inhibit the New Product . . . . .	xxvii
20.4.1	Regulatory Constraints . . . . .	xxvii
20.4.2	Technological Dependencies . . . . .	xxvii
20.5	Follow-Up Problems . . . . .	xxvii
20.5.1	Data Quality Management . . . . .	xxviii
20.5.2	User Engagement and Retention . . . . .	xxviii
20.5.3	Ethical Considerations . . . . .	xxviii
20.5.4	Dependency on User Participation . . . . .	xxviii
20.5.5	Technical Debt Accumulation . . . . .	xxviii
<b>21</b>	<b>Tasks . . . . .</b>	<b>xxix</b>
21.1	Project Planning . . . . .	xxix
21.2	Planning of the Development Phases . . . . .	xxix
<b>22</b>	<b>Costs . . . . .</b>	<b>xxix</b>
22.1	Development Costs . . . . .	xxix
22.1.1	Software Development . . . . .	xxix
22.1.2	UI/UX Design . . . . .	xxix
22.1.3	Testing and Quality Assurance . . . . .	xxx
22.2	Infrastructure Costs . . . . .	xxx
22.2.1	Cloud Services . . . . .	xxx
22.2.2	Data Acquisition . . . . .	xxx
22.3	Operational Costs . . . . .	xxx
22.3.1	User Compensation . . . . .	xxx
22.3.2	Maintenance and Support . . . . .	xxxi
22.3.3	Security and Compliance . . . . .	xxxi
22.4	Marketing and User Acquisition Costs . . . . .	xxxi
22.4.1	Promotional Activities . . . . .	xxxi
22.4.2	Onboarding and Training . . . . .	xxxi
22.5	Contingency and Miscellaneous Costs . . . . .	xxxi
22.5.1	Unexpected Expenses . . . . .	xxxi

22.6 Budget Forecast . . . . .	xxxix
<b>23 User Documentation and Training</b>	<b>xxxii</b>
23.1 User Documentation Requirements . . . . .	xxxii
23.2 Training Requirements . . . . .	xxxii
<b>24 Waiting Room</b>	<b>xxxii</b>
<b>25 Ideas for Solution</b>	<b>xxxii</b>

## Revision History

Date	Version	Notes
10/1/2024	Mathew Petronilho	Added Purpose Of Project
Date 2	1.1	Notes

# **1 Purpose of the Project**

There is currently a lack of high-quality, labeled satellite imagery datasets tailored for specific use cases. Many industries require specialized data for tasks like disaster response, environmental monitoring, urban planning, or defense, but building these datasets manually is time-consuming, costly, inefficient and may require expert data analysis. This hinders the development and deployment of accurate computer vision models for critical use cases across these various industries.

The purpose of this project is to create an online platform that accelerates this process and brings simplicity to satellite imagery data analysis.

## **1.1 Goals of the Project**

### **1.1.1 High Data Accuracy**

The system should have high classification accuracy for objects reported in the images. The core problem this system must solve is extracting useful information from the provided images. One key metric to determine the utility of the information found, is the classification accuracy of objects identified in the images. If the system is not able to determine what is contained in an image, it will not be useful to stakeholders.

### **1.1.2 Ease of use**

The system should be very easy for stakeholders to use. There should be very low friction for users to classify images and objects found within images, with minimal training. It should also be simple for users to upload images to be analyzed. To maximize the information gained from users who are contributing to classification efforts, the system must ensure it is simple for users to get started with, and continue using the system. This is necessary to build a large enough user base, which will make it more likely to get insights in an acceptable amount of time.

### **1.1.3 Minimizing Cost to Analyze Images**

The system should minimize the cost for users request insights from images. This could be implemented through intelligent algorithms for task delegation. Users of the system who upload images are interested in getting an



appropriate return for their investment. If the cost to analyze is too high, the platform will not retain a sufficiently large user base of purchasers.

#### **1.1.4 Results Returned Within Appropriate Timeframe**

The system should ensure the time it takes to obtain information from images is within a specified limit, as determined by users who upload images. Purchasers will have some time limit they require the system to process images within. To ensure timing needs are met, the system should provide realistic timelines and stick to them.

#### **1.1.5 High System Reliability and Accessibility**

The system should be useable remotely for purchasers and labellers, and have minimal downtime. The system should allow purchasers to upload images without being physically located where the system is hosted to ensure flexibility of use. The same should also be true for labellers, as they should be able to perform their tasks remotely. In both cases, the system should have low down time as to not introduce additional friction into the completion of tasks.

## **2 Stakeholders**

### **2.1 Client**

*Insert your content here.*

### **2.2 Customer**

*Insert your content here.*

### **2.3 Other Stakeholders**

*Insert your content here.*

### **2.4 Hands-On Users of the Project**

*Insert your content here.*

## 2.5 Personas

*Insert your content here.*

## 2.6 Priorities Assigned to Users

*Insert your content here.*

## 2.7 User Participation

*Insert your content here.*

## 2.8 Maintenance Users and Service Technicians

*Insert your content here.*

# 3 Mandated Constraints

- The solution must be fully compatible with the latest stable releases of Google Chrome, Firefox, Microsoft Edge, and Safari browsers.

**Rationale:** Users will interact with the web application through various modern web browsers, so ensuring cross-browser compatibility is essential for providing a consistent user experience.

**Fit Criteria:** The web application must display consistently, maintain full functionality, and support core features across all specified browsers without major visual or functional discrepancies. Testing should be conducted on each browser to validate compatibility.

## 3.1 Implementation Environment of the Current System

There is no current environment in which our application must be implemented.

## 3.2 Partner or Collaborative Applications

There are no constraints regarding external applications that must be used alongside our product.

### 3.3 Off-the-Shelf Software

There is no required off-the-shelf software that must be used for our application.

### 3.4 Anticipated Workplace Environment

There is no particular location where users are required to work and use the product. As a web application, it can be accessed from most computers with an internet connection. We do not anticipate that the users' environment will physically constrain their ability to use the app in any way.

### 3.5 Schedule Constraints

- The proof of concept for this project must be ready to demonstrate by **November 11, 2024**. Not meeting this deadline will result in uncertainty about overcoming major risks associated with the project.
- The first project demonstration must be ready by **February 3, 2025**. Missing this deadline will reduce the time available to make refinements based on feedback and findings.
- The final demonstration must be ready by **March 24, 2025**. Missing this milestone would prevent the project from being presented and result in a significant loss of marks.

To see other documentation deadlines related to this project, refer to our [Development Plan](#).

### 3.6 Budget Constraints

- The project budget must not exceed \$750. All funds will be sourced from the team itself.

## **4 Naming Conventions and Terminology**

### **4.1 Glossary of All Terms, Including Acronyms, Used by Stakeholders involved in the Project**

*Insert your content here.*

## **5 Relevant Facts And Assumptions**

### **5.1 Relevant Facts**

*Insert your content here.*

### **5.2 Business Rules**

*Insert your content here.*

### **5.3 Assumptions**

*Insert your content here.*

## **6 The Scope of the Work**

### **6.1 The Current Situation**

*Insert your content here.*

### **6.2 The Context of the Work**

*Insert your content here.*

### **6.3 Work Partitioning**

*Insert your content here.*

### **6.4 Specifying a Business Use Case (BUC)**

*Insert your content here.*

## 7 Business Data Model and Data Dictionary

### 7.1 Business Data Model

*Insert your content here.*

### 7.2 Data Dictionary

*Insert your content here.*

## 8 The Scope of the Product

### 8.1 Product Boundary

*Insert your content here.*

### 8.2 Product Use Case Table

*Insert your content here.*

### 8.3 Individual Product Use Cases (PUC's)

*Insert your content here.*

## 9 Functional Requirements

### 9.1 Functional Requirements

*Insert your content here.*

## 10 Look and Feel Requirements

### 10.1 Appearance Requirements

**Requirement LF1:**

- **Description:** The application shall adapt to various screen sizes, ensuring legibility and an uncluttered layout.

- **Rationale:** Users will have computers with varying screen sizes, so a consistent experience across all these sizes is ideal.
- **Fit Criterion:** Visual elements must not exceed the boundaries of a screen with a size between the range 1024×768 pixels to 1920×1080 pixels.

#### **Requirement LF2:**

- **Description:** Interactive elements such as buttons shall provide visual feedback to the user.
- **Rationale:** This will allow users a better understanding of when their actions have been processed by the application.
- **Fit Criterion:** Every interactive element changes colour or displays additional visual cues, such as animations or shadows, to indicate interaction.

## **10.2 Style Requirements**

#### **Requirement LF3:**

- **Description:** The application should maintain a unified visual design across all components.
- **Rationale:** A consistent appearance enhances the application's cohesiveness and conveys a professional aesthetic.
- **Fit Criterion:** Font type, sizing, and colour, along with background tones are all consistent throughout the application.

## **11 Usability and Humanity Requirements**

### **11.1 Ease of Use Requirements**

*Insert your content here.*

## 11.2 Personalization and Internationalization Requirements

*Insert your content here.*

## 11.3 Learning Requirements

*Insert your content here.*

## 11.4 Understandability and Politeness Requirements

*Insert your content here.*

## 11.5 Accessibility Requirements

*Insert your content here.*

# 12 Performance Requirements

## 12.1 Speed and Latency Requirements

### NFR-PR0

- **Description:** The system shall process new user requests within 15 minutes of a Customer completing the account creation process 90% of the time, and within 48 hours in all cases.
- **Rationale:** The user likely has urgent needs if they are signing up to access the system. They must have quick access to the services offered by the system, and authentication is the first step of this process.
- **Fit Criterion:**  
Let  $t_{\text{newCustomerAccount}}$  be the time it takes to process a new Customer account request, in hours.  
 $\mathbb{P}(t_{\text{newCustomerAccount}} < 0.25) \geq .90 \wedge \mathbb{P}(t_{\text{newCustomerAccount}} < 48) = 1$

### NFR-PR1

- **Description:** The system shall process new user requests within 15 minutes of a Labeler completing the account creation process 90% of the time, and within 48 hours in all cases.
- **Rationale:** To improve engagement from Labelers, there should be minimal delay in getting started.
- **Fit Criterion:**  
Let  $t_{\text{newLabelAccount}}$  be the time it takes to process a new Labeler account request, in hours.  
 $\mathbb{P}(t_{\text{newLabelAccount}} < 0.25) \geq 0.9 \wedge \mathbb{P}(t_{\text{newLabelAccount}} < 48) = 1$

### NFR-PR2

- **Description:** The system shall return a complete report of results to the Customer within the negotiated amount of time, 90% of the time, and within 48 additional hours in all cases.
- **Rationale:** The user is expecting to be able to act on the insights the system provides. To build trust and loyalty, the system must ensure that it is meeting the agreed upon timelines.
- **Fit Criterion:**  
Let  $t_{\text{serviceRequestTime}}$  be the time it takes to complete a service request, in hours.  
Let  $t_{\text{serviceRequestTimeLimit}}$  be the negotiated time limit for completing a service request, in hours.  
 $\mathbb{P}(t_{\text{serviceRequestTime}} < t_{\text{serviceRequestTimeLimit}}) \geq 0.9$   
 $\wedge \mathbb{P}(t_{\text{serviceRequestTime}} < t_{\text{serviceRequestTimeLimit}} + 48) = 1$

### NFR-PR3

- **Description:** The system shall take no longer than 10 seconds to display the next image to be labeled to a labeler, if there is one available.
- **Rationale:** To reduce friction for labelers, the system must ensure there is no unnecessary delay in preparing the next job. For users exposed to modern media, attention spans can be expected to be less than 10 seconds [src](#).



- **Fit Criterion:**

Let  $t_{\text{imageServing}}$  be the time it takes to serve the next image, in seconds.  
 Let  $x_{\text{nextImage}}$  equal True if there is a next image available, and False otherwise.

$$x_{\text{nextImage}} \Rightarrow t_{\text{imageServing}} \leq 10$$

#### NFR-PR4

- **Description:** The system shall deliver earned payouts to labelers within 7 business days of a request being made through the system.
- **Rationale:** The labelers are entitled to their earned income, and there must not be unnecessary delay. 7 days accounts for delays in the platform used to distribute payments.
- **Fit Criterion:**

Let  $t_{\text{payoutDelay}}$  be the time it takes for a user to receive their payments after a request is received by the system, in days.

$$t_{\text{payoutDelay}} < 7$$

## 12.2 Safety-Critical Requirements

N/A

## 12.3 Precision or Accuracy Requirements

#### NFR-PR5

- **Description:** The system shall report accurate labels 75% of the time.
- **Rationale:** Average data label accuracy for competitors is greater than 75% ([link](#)). The system should at a minimum provide the same label accuracy.
- **Fit Criterion:**

Let  $O$  be the set of objects to label.

Let  $C$  be the set of classes an object in  $O$  can be.

Let  $L_{\text{True}} : O \rightarrow C$  be a function which maps objects in  $O$  to their true classes in  $C$ . Let  $L_{\text{Guess}} : O \rightarrow C$  be the function derived from

the system which maps objects in  $O$  to their assumed classes in  $C$ .  
 $(\forall o \in O : \mathbb{P}(L_{\text{True}}(o) = L_{\text{Guess}}(o)) \geq 0.75)$

## 12.4 Robustness or Fault-Tolerance Requirements

### NFR-PR6

- **Description:** The system shall have 97% uptime.
- **Rationale:** It is crucial that in emergency response use cases, the system is able to accept and process requests with minimal delay.
- **Fit Criterion:**  
 Let  $t_{\text{uptime}}$  be the uptime of the system.  
 Let  $t_{\text{downtime}}$  be the downtime of the system.  

$$\frac{t_{\text{uptime}}}{t_{\text{uptime}} + t_{\text{downtime}}} > 0.97$$

## 12.5 Capacity Requirements

### NFR-PR7

- **Description:** The system shall have the capacity to support enough labelers to meet all service request deadlines.
- **Rationale:** This requirement is critical to satisfy NFR-PR2.
- **Fit Criterion:** See NFR-PR2.

### NFR-PR8

- **Description:** The system shall have the capacity to store and process large image files.
- **Rationale:** This requirement is necessary to obtain information from satellite images, which can be several gigabites in size.
- **Fit Criterion:** The system will not crash or fail to store when given images files 50 Gb in size.

## 12.6 Scalability or Extensibility Requirements

### NFR-PR9

- **Description:** The system shall be able to scale to meet the capacity specified in NFR-PR7.
- **Rationale:** This requirement is critical to satisfy NFR-PR7.
- **Fit Criterion:** See NFR-PR7.

## 12.7 Longevity Requirements

N/A

## 13 Operational and Environmental Requirements

### 13.1 Expected Physical Environment

*Insert your content here.*

### 13.2 Wider Environment Requirements

*Insert your content here.*

### 13.3 Requirements for Interfacing with Adjacent Systems

*Insert your content here.*

### 13.4 Productization Requirements

*Insert your content here.*

### 13.5 Release Requirements

*Insert your content here.*

## 14 Maintainability and Support Requirements

### 14.1 Maintenance Requirements

#### NFR-MR1

- **Description:** All maintainance required for the system shall be possible to complete by a competent software developer after reading all of the documentation provided in the source repository.
- **Rationale:** The system should be well documented, and therefore maintainable after reading said documents.
- **Fit Criterion:** A competent software developer, as determined by the original developers or their agents, can solve problems related to the core of the system, as determined by the features outlined in this document.

### 14.2 Supportability Requirements

N/A

### 14.3 Adaptability Requirements

N/A

## 15 Security Requirements

### 15.1 Access Requirements

#### Requirement SR1:

- **Description:** The application shall only allow users with labeling access, including labellers, customers, and admins, to view active projects and label images.
- **Rationale:** We do not want random users with no stake in the process to effect the results.

- **Fit Criterion:** Users who have not logged in to the application have no way of viewing projects or labeling images. Users logged in as labellers, customers, or admins have access to these features.

#### Requirement SR2:

- **Description:** The application shall only allow users with customer access and above to create new image analysis projects.
- **Rationale:** Unidentified users creating projects would be impossible to facilitate. Also, labellers have no need to access project creation.
- **Fit Criterion:** Users who have not logged in to the application have no way of creating an image analysis project. Users logged in as customers or admins have access to these features.

#### Requirement SR3:

- **Description:** The application shall validate the email format the user provides when creating an account.
- **Rationale:** We do not want users using invalid emails to sign up.
- **Fit Criterion:** Let  $E$  represent the set of all email addresses, and let  $V$  represent the set of all valid email addresses. A valid email address conforms to the general pattern:

$$V = (\forall \text{ email} \in E \mid \text{email matches the pattern } [a-zA-Z0-9+_{-}]+@[a-zA-Z0-9_{-}]+[a-zA-Z])$$

#### Requirement SR4:

- **Description:** The application shall validate the password format the user provides when creating an account.
- **Rationale:** We do not want users using weak passwords to sign up.
- **Fit Criterion:** Let  $P$  represent the set of all passwords, and let  $V$  represent the set of all valid passwords. A valid password has a at least one lowercase, uppercase, number and special character and is a

minimum of 8 characters in length:

$$V = (\forall \textit{password} \in P \mid \textit{password matches the pattern } (?=.*[a-z])(?=.*[A-Z])(?=.*[0-9])(?=.*[\#\$\%\&])[a-zA-Z0-9\#\$\%\&]\{8,\})$$

## 15.2 Integrity Requirements

### Requirement SR5:

- **Description:** The application shall prevent incorrect data from being introduced.
- **Rationale:** The database of information should always reflect correct and up to date information.
- **Fit Criterion:** The system must validate user inputs for data accuracy and format before they are saved. Any invalid data must trigger error messages, preventing it from being entered into the database. Users must be required to correct errors before proceeding.

## 15.3 Privacy Requirements

### Requirement SR6:

- **Description:** User data will be securely encrypted to protect user's privacy.
- **Rationale:** This will help to avoid user's being compromised if a data leak occurs.
- **Fit Criterion:** An encryption algorithm is used on sensitive user data such as passwords.

### Requirement SR7:

- **Description:** The application shall ensure that all payment transactions are processed securely using encryption and comply with relevant security standards, such as PCI-DSS, which helps to protect payment account data (PCI Security Standards Council, 2024).

- **Rationale:** Protecting users' financial information is critical to maintaining trust. Failing to secure payments can lead to data breaches, financial loss, and legal liabilities.
- **Fit Criterion:** All payment transactions must use industry-standard encryption to protect sensitive data. Payment information, such as credit card details, must not be stored locally on the application and must be processed via a secure, PCI-DSS-compliant third-party payment gateway.

## 15.4 Audit Requirements

These requirements are not applicable as we are not an organization that is currently subject to audits.

## 15.5 Immunity Requirements

### Requirement SR8:

- **Description:** The application shall use parameterized queries or prepared statements for all database interactions.
- **Rationale:** We want to prevent SQL injection attacks which can lead to unauthorized data access or manipulation.
- **Fit Criterion:** All database queries must be implemented using parameterized queries or prepared statements. Dynamic SQL strings that concatenate user input must not be used in the codebase.

# 16 Cultural Requirements

## 16.1 Cultural Requirements

### NFR-CUR1

- **Description:** The system shall present users with the option to select the most popular language in each country it is deployed in.
- **Rationale:** It is important that the users of the program can understand what is said in each step.

- **Fit Criterion:** A drop down will allow users to select from the list of languages. At a minimum, the most popular language by number of speakers will be available for each country.

## 17 Compliance Requirements

### 17.1 Legal Requirements

#### NFR-COR1

- **Description:** The system shall not be available in any country currently facing economic sanctions by the Government of Canada.
- **Rationale:** Legal requirement to operate.
- **Fit Criterion:** Website must not be reachable in sanctioned countries.

#### NFR-COR2

- **Description:** The system shall follow Canadian tax code when accepting and paying out earnings.
- **Rationale:** Legal requirement to operate.
- **Fit Criterion:** All relevant tax codes must be satisfied when accepting payment from Customers or paying out earnings to Labelers.

#### NFR-COR3

- **Description:** The system shall allow Customers to restrict Labelers from certain regions from labeling the images related to their service request.
- **Rationale:** Customers may be uploading sensitive images, which are inappropriate for international Labelers to view.
- **Fit Criterion:** No restricted image, as identified by the customer, shall be shown to a restricted group.



## 17.2 Standards Compliance Requirements

N/A

## 18 Open Issues

**Task Assignment Algorithm:** To ensure labelers are engaged as they complete tasks and to obtain the highest quality of information possible, the system must implement an intelligent task allocation system. This system has not yet been determined.

**Label Consensus Algorithm:** Similarly, the algorithm for combining multiple user labels into one accurate label has not yet been determined.

**Labeling Services Offered:** The system has determined several potential labeling services to be offered by the system, but has not confirmed with certainty what will be included. This will be determined after more research has been completed on the task assignment algorithm.

## 19 Off-the-Shelf Solutions

### 19.1 Ready-Made Products

**Amazon Mechanical Turk:** A web-based crowdsourcing platform. Instead of building a novel front end, the system could obtain labels through this platform instead.

### 19.2 Reusable Components

**Label Studio:** A React library which contains components for building a web-based data annotation platform.

### 19.3 Products That Can Be Copied

**Tolka AI:** A general purpose image label crowdsourcing site. Supports image segmentation, bounding box drawing, and more computer vision labeling tasks.

## **20 New Problems**

### **20.1 Effects on the Current Environment**

The introduction of the OKKM Insights platform is expected to have several impacts on the existing technological and operational environment.

#### **20.1.1 Data Privacy and Security Concerns**

1. Impact: Handling sensitive satellite imagery and user-generated data raises significant privacy and security issues. Ensuring compliance with data protection regulations is paramount.
2. Potential Problem: Breaches or mishandling of data could lead to legal repercussions, loss of trust, and reputational damage.

#### **20.1.2 Environmental Footprint**

1. Impact: Increased computational requirements for AI processing and data storage may lead to higher energy consumption.
2. Potential Problem: This could conflict with sustainability goals and lead to higher operational costs.

### **20.2 Effects on the Installed Systems**

Introducing the OKKM Insights platform will interact with and potentially disrupt existing systems within the organization and for stakeholders.

#### **20.2.1 Integration Challenges**

1. Impact: The platform will need to integrate with existing data sources, cloud services, and possibly applicable third-party APIs.
2. Potential Problem: Incompatibilities or integration failures could result in data inconsistencies, system downtimes, or increased maintenance efforts.

### **20.2.2 Legacy Systems Compatibility**

1. Impact: Older systems may not support the latest technologies required by the new platform.
2. Potential Problem: Upgrading or adjusting to legacy systems can be costly and time-consuming, potentially delaying the platform’s deployment.

## **20.3 Potential User Problems**

Users are at the heart of the OKKM Insights platform, and several issues may arise that affect their experience and satisfaction.

### **20.3.1 Usability Issues**

1. Impact: If the platform is not intuitive or user-friendly, users may struggle to navigate and utilize its features effectively.
2. Potential Problem: Poor user experience can lead to reduced engagement, lower data labeling contributions, and higher dropout rates.

### **20.3.2 Training and Onboarding**

1. Impact: Users may require training to understand how to label data accurately and use the platform’s tools.
2. Potential Problem: Inadequate training resources can result in inconsistent labeling, decreasing the quality of the datasets and the reliability of the AI models.

### **20.3.3 Compensation and Incentives**

1. Impact: Users expect fair compensation for their contributions.
2. Potential Problem: Delays or inaccuracies in payment processing can lead to dissatisfaction, reducing user retention and the overall quality of data labeling efforts.

#### **20.3.4 Technical Support**

1. Impact: Users may encounter technical issues that require timely resolution.
2. Potential Problem: Insufficient support can frustrate users, leading to decreased platform usage and negative word-of-mouth.

### **20.4 Limitations in the Anticipated Implementation Environment That May Inhibit the New Product**

Several environmental and contextual limitations could hinder the effective implementation and operation of the OKKM Insights platform.

#### **20.4.1 Regulatory Constraints**

1. Limitation: Different countries have varying regulations regarding satellite data usage, privacy, and AI applications.
2. Impact: Navigating these regulatory landscapes can be complex and may restrict the platform's operations in certain regions, limiting market potential.

#### **20.4.2 Technological Dependencies**

1. Limitation: The platform relies on third-party services (e.g., cloud providers, satellite data suppliers) whose availability and reliability are beyond the project's control.
2. Impact: Downtime or changes in third-party services can disrupt platform functionality and user experience.

### **20.5 Follow-Up Problems**

Addressing the initial set of problems may give rise to additional challenges that need to be managed.

### **20.5.1 Data Quality Management**

1. Follow-Up Problem: Ensuring the ongoing accuracy and relevance of labeled datasets as new data is continuously added. This may require implementing robust quality assurance mechanisms and periodic reviews.

### **20.5.2 User Engagement and Retention**

1. Follow-Up Problem: Continuously engaging users to maintain an active labeling workforce may require ongoing incentives, gamification strategies, and community-building efforts to prevent user fatigue and attrition.

### **20.5.3 Ethical Considerations**

1. Follow-Up Problem: Addressing ethical concerns related to surveillance, data misuse, and the potential dual-use nature of satellite imagery data. Establishing ethical guidelines and oversight mechanisms will be necessary to prevent misuse.

### **20.5.4 Dependency on User Participation**

1. Follow-Up Problem: The platform's success heavily relies on active user participation for data labeling. Fluctuations in user engagement can directly impact data quality and availability, requiring strategies to stabilize user contributions.

### **20.5.5 Technical Debt Accumulation**

1. Follow-Up Problem: Rapid development to address emerging problems may lead to technical debt, where short-term solutions create long-term maintenance challenges. Proper code management and refactoring practices will be essential to mitigate this issue.

## **21 Tasks**

### **21.1 Project Planning**

*Insert your content here.*

### **21.2 Planning of the Development Phases**

*Insert your content here.*

## **22 Costs**

Developing and implementing the requirements for the OKKM Insights platform involves various financial and effort-related expenditures. Below is a detailed breakdown of the primary cost components associated with building the platform:

### **22.1 Development Costs**

#### **22.1.1 Software Development**

1. **Front-End Development:** Creating a user-friendly web interface requires experience working with front-end technologies such as HTML, CSS, JavaScript, and frameworks like React or Angular.
2. **Back-End Development:** Developing robust server-side infrastructure using technologies like Node.js, Flask, and Python. This includes setting up databases, APIs, and integrating with cloud services (AWS/Azure).
3. **AI and Machine Learning Integration:** Building and integrating AI-powered features for automatic data labeling and computer vision model training will require specialized expertise in machine learning, potentially increasing development costs.

#### **22.1.2 UI/UX Design**

1. Investing time into UI/UX design to ensure the platform is intuitive and easy to use for both data labelers and end clients. This includes

designing workflows, views, and ensuring responsive design across devices.

### **22.1.3 Testing and Quality Assurance**

1. Comprehensive testing to identify and fix bugs, ensure cross-platform compatibility, and maintain high system reliability. This involves both automated and manual testing processes.

## **22.2 Infrastructure Costs**

### **22.2.1 Cloud Services**

1. Hosting: Utilizing cloud platforms like AWS or Azure for hosting the backend services and databases. Costs will scale with usage, including storage for satellite images, computational resources for AI processing, and data transfer fees.
2. Scalability: Ensuring the infrastructure can scale to handle increasing numbers of users and data volume, which may involve additional costs for load balancing, auto-scaling, and enhanced security measures.

### **22.2.2 Data Acquisition**

1. Purchasing high-quality, commercially available satellite images from third-party providers. These costs can vary based on the resolution, coverage area, and frequency of image updates required for different use cases.

## **22.3 Operational Costs**

### **22.3.1 User Compensation**

1. Labeling Incentives: Allocating funds to compensate users who contribute to the image labeling process. This includes setting competitive rates to attract and retain a large and active user base.
2. Payment Processing Fees: Costs associated with handling financial transactions, including fees from payment gateways for distributing earnings to users and receiving payments from clients.

### **22.3.2 Maintenance and Support**

1. Ongoing maintenance of the platform to ensure uptime, implement updates, and address technical issues. This also includes providing customer support to both data labelers and end clients.

### **22.3.3 Security and Compliance**

1. Implementing robust security measures to protect sensitive satellite data and financial transactions. Costs may include encryption technologies, regular security audits, and compliance with data protection regulations.

## **22.4 Marketing and User Acquisition Costs**

### **22.4.1 Promotional Activities**

1. Marketing efforts to attract both data labelers and end clients to the platform. This includes digital marketing campaigns, partnerships with relevant organizations, and participation in industry events.

### **22.4.2 Onboarding and Training**

1. Creating tutorials, documentation, and training materials to facilitate easy onboarding of new users and ensure they can effectively contribute to the labeling process with minimal friction.

## **22.5 Contingency and Miscellaneous Costs**

### **22.5.1 Unexpected Expenses**

1. Allocating a budget for unforeseen challenges such as technical setbacks, additional feature requests, or changes in market conditions that may require pivoting the project strategy.

## **22.6 Budget Forecast**

A detailed budget forecast will be developed, encompassing all the aforementioned cost categories. This forecast will be periodically reviewed and



adjusted based on project milestones, market conditions, and actual expenditure patterns to ensure financial sustainability and efficient resource allocation.

*Insert your content here.*

## **23 User Documentation and Training**

### **23.1 User Documentation Requirements**

*Insert your content here.*

### **23.2 Training Requirements**

*Insert your content here.*

## **24 Waiting Room**

*Insert your content here.*

## **25 Ideas for Solution**

*Insert your content here.*

## References

1. PCI Security Standards Council. (2024, May 13). *PCI Security Standards Council – Protect Payment Data with Industry-driven Security Standards, Training, and Programs*. <https://www.pcisecuritystandards.org/standards/pci-dss/>

## Appendix — Reflection

The information in this section will be used to evaluate the team members on the graduate attribute of Lifelong Learning. Please answer the following questions:

1. What knowledge and skills will the team collectively need to acquire to successfully complete this capstone project? Examples of possible knowledge to acquire include domain specific knowledge from the domain of your application, or software engineering knowledge, mechatronics knowledge or computer science knowledge. Skills may be related to technology, or writing, or presentation, or team management, etc. You should look to identify at least one item for each team member.
2. For each of the knowledge areas and skills identified in the previous question, what are at least two approaches to acquiring the knowledge or mastering the skill? Of the identified approaches, which will each team member pursue, and why did they make this choice?