Requirements Specification

Team name:
LangLens

Faculty Mentor
Italo Santos

Sponsor
Dr. Okim Kang

Team Members
Stefan Mihailovic, Daniel Navarette Martin
Brian Ruiz, Sami Tanquary
Kyle Young

Overview
Description of our project's requirements, both functional and non-functional, and forms the contractual basis for the expectations to be fulfilled by the development team.

Accepted as baseline requirements for the project:
Team lead signature: Client signature:

_____ _____

Version 2.0

December 5, 2022

Contents

1

Introduction

Language is arguably one of the most crucial aspects of human connection. While there are many different methods for communicating with one another, written and verbal communication allows for people of differing cultures and identities to connect and understand each other; however, this is only possible through the studying and mastering of foreign languages and/or translators to help bridge what is considered a language barrier. When learning a new language, it is common practice for learners to strengthen their understanding via reading textbooks, writing, practicing conversation, and simply memorization of endless lists of vocabulary. By following these outdated methods for language learning, it can take some learners years to master just one new language. Now, this is not to say that traditional language learning methods are useless, but recent advances in technology have progressed the world into an era of unlimited access to information and endless potential from a device that a majority of people carry in their pockets. Why not enhance the ability to learn a new language by ditching the textbooks and simply scanning the environment around you to see the world change into the language you're trying to learn right before your eyes? Or even better, interact with this augmentation of reality by saving what you see into digital flashcards and mastering languages all from the touch of a screen?

Our client, Dr. Okim Kang, is a professor in the Department of English at Northern Arizona University with special emphasis in Applied Linguistics and TESL (Teaching English as a Second Language). She has published numerous academic books and journals covering a wide range of research topics surrounding language learning and language teaching, and she has also been awarded with several honors regarding her outstanding contributions to the research and innovations of TESL and language pedagogy.

Dr. Okim, and her team, have been working in Computer-Assisted Language Learning (CALL) for several years and have observed a significant lack of tools targeted towards foreign language learners that utilize the latest advances in AR technology. A majority of language learning tools that are currently on the market only include a few aspects of speech and language pedagogy and fail to modernize their tools by rarely providing both object detection and optical character recognition (OCR) functionality. To fill this market gap, LangLens has partnered with Dr. Okim to create an innovative, web-based tool, designed for all mobile devices, that provides a free-to-use, immersive, language learning experience employing both object detection and OCR capabilities with the latest AR technology available.

2

Problem Statement

Our sponsor, Dr. Okim Kang, is a linguistics professor at NAU within the department of English. Her and her team have worked on and developed multiple systems in the past to provide computer assisted language learning tools to language learners. Over time they have noticed a distinct lack of language learning tools that utilize augmented reality.

Our client came up with an idea to fill this void with a web based application that utilizes augmented reality in terms of object detection, and character recognition to scan objects and text in the user's surroundings, and to assist them in language learning.

This project would serve as an independent tool, not necessarily a part of an already existing production flow, as LangLens would be creating a unique computer assisted language learning tool. While there are many tools that dabble with augmented reality, they all have certain deficiencies, missing capabilities, or limitations such as:

- Having either object detection or text recognition, but not both.
- If they do have either of the above, they don't specifically target language learning.
- Not being easily accessible, either due to being tied to a single platform, or hidden behind a paywall.
- Not enough focus on all key elements on language learning, such as meaning, use, and form.

LangLens' goal is to develop a tool that will tackle and resolve all of the issues raised above, and deploy an application that is free and widely available, regardless of the user's platform. That targets anyone wishing to learn a new language, while focusing on all the key elements of language learning, in an augmented reality environment providing an ability to detect and scan both objects and text in the user's surroundings.

3

Solution Vision

While there are many widely available language learning tools on the market, such as Duolingo or Mondly, these tools rarely incorporate the use of both object and text detection within their applications, and if they do, these tools do not center their focus around the key elements of word learning, which as described by our client, Dr. Okim Kang, are meaning, usage, and form. LangLens envisions to solve this lack of computer-assisted language learning tools by fully utilizing the potential of augmented reality in a web-app, optimized for mobile devices, that incorporates all forms of word learning elements, in a free and simple to understand interface that is accessible to the widest audience of language learners possible (i.e. students, teachers, researchers).

The following features of the LangLens solution aim to employ recent advances in object detection and optical character recognition (OCR) via detection modes that the user can toggle between upon opening the application. By accessing the user's mobile camera in a live viewing mode, recent AR technology will be utilized to allow language learners to quickly and accurately scan objects and text in their environment in order to learn the definitions and proper use of the scanned object / word in textual, visual and audible forms via a clickable link to an external learning page in the user's chosen target language.

3.1 | Specific Solution Features

- Accessing the user's camera in a live viewing mode to fully utilize AR technology.
- A target language selector for translation of the scanned objects / words (English,

Spanish, French, and Korean are the preliminary languages available for initial development).

- Two detection modes: object detection and text recognition that will identify the object / word with a bounding box and label presented to the user in their chosen target language quickly and accurately.
- Upon each scan, a clickable link to an external word learning page with an open-source dictionary definition, a sentence usage example, a graphical image representation, and pronunciation example appears to fully incorporate all elements of word learning.
- Ability to return to the scanning screen and start the process again for repeated use.

3.2 | Stretch Goal Features

- Private user accounts with the ability to save the objects / words scanned into "PLAYlists" and review / interact with their lists in a practice mode for further reinforcement.
- A much larger selection of native / target languages for further audience reach (i.e. German, Russian, Arabic, Hindi, Portuguese etc.).

3.3 | Coarse System Architecture for LangLens

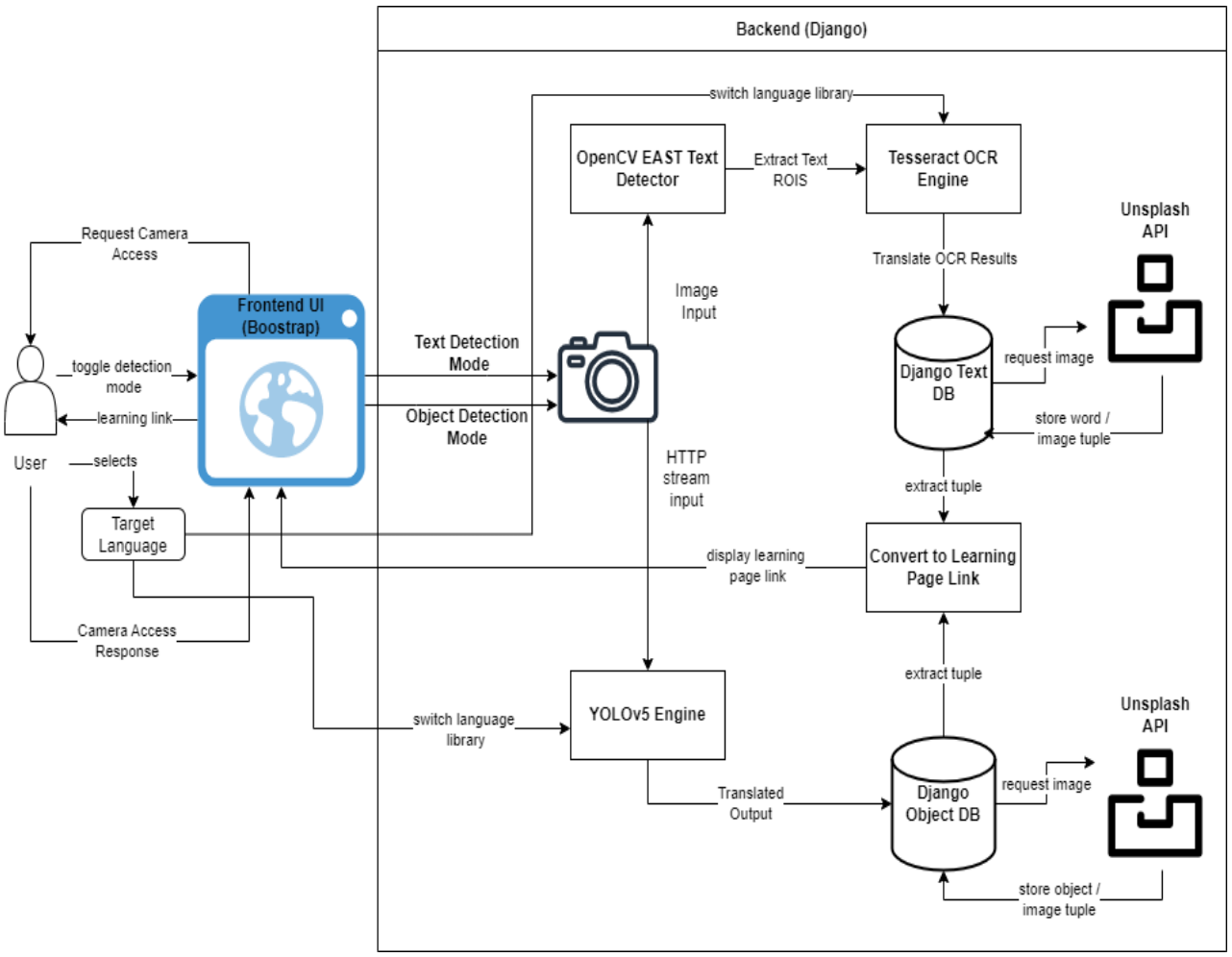

Figure 3.1: Coarse system architecture currently envisioned for LangLens. Subject to design revision.

The diagram 3.1 depicts two very similar, but computationally different, approaches to the flow of data within LangLens depending on the detection mode selected. The results of either detection mode are sent to the OpenCV EAST Text Detector or YOLOv5 Engine respectively via Bootstrap and are received by the Django backend API for direct communication with the detection engines for translation and the creation of the external learning pages, via a Bootstrap template. LangLens aspires to be a free-to-use software, but because of a lack of paid services within our system, there are a few tradeoffs in terms of speed and performance; however, LangLens is confident that even

if competitors are faster, there are very few of them that offer both object and text detection in one application, as well as really benefit the language learner with the centralized aspect of a learning page that includes all of the key elements of word learning. LangLens may not be the first of its kind in terms of computer-assisted language learning, but it will definitely create a strong foundation for the innovation of future language learning tools.

4

Project Requirements

For team LangLens to create an immersive augmented reality language learning tool, we would need to incorporate the high-level requirements outlined below. These domain level requirements are based on the key requirements from the user's perspective. These are broken into two sections: Key Requirements and Stretch Goal Requirements.

Key Requirements:

- The web-app will be accessible from any mobile device with internet access and a camera.
- The web-app will be packaged in a simple, easy to understand, user interface capable of quickly and securely scanning objects or retrieving text from an image.
- The web-app will offer the ability to choose a target language for translation from English to Spanish, French and/or Korean.
- The web-app will be capable of toggling between two detection modes: Detect Object and Detect Word.
- The web-app will offer the ability to restart the scanning process.
- The web-app will scan objects and texts in a live viewing mode for real time detection.

- The web-app will display a clickable link to an external learning page that provides a definition, word-in-use guide, image representation, and pronunciation example.

Stretch Goal Requirements:

- The web-app will be capable of storing objects and text within private user accounts to be used within a practice mode.
- The web-app will be capable of translating words and objects from any detected base language (not just English).
- The web-app will be capable of translating words and objects to a larger selection of target languages suggestable by the user.
- The web-app will offer secure and private user accounts.

These requirements will be able to be defined and proven in a way that will be able to be validated. These requirements will be only as good as its correlated test. Team LangLens will create an application that services these requirements by Spring of next semester. The requirements above have been divided into the following categories:

- Functional Requirements.
- Non-Functional Requirements.
- Environmental Requirements.

4.1 | Functional Requirements

4.1.1 | User Actions

The User Actions consist of all the specific user functions associated with the UI. Our target users for this system include anyone who would like to learn a new language. Users will have multiple utilities accessible to them when interacting with the UI that serve as the core functions for the application.Therefore, users must be able to execute the following actions:

- The user will be able to allow the application to have permission or denial to access the camera in order for the application to be used or not.
- The user will be able to select a target language for translation offered in Spanish, French, or Korean via a target language selection prompt and drop down options of the currently available languages.
  - **Stretch Requirements:** The user will be able to change the native base language of the application to another language other than English.
    - Additionally, the user will be able to choose from a larger selection of target languages.
- The user will be able to toggle between two detection modes: one for detecting objects and another for detecting words. A possible **stretch goal** would allow users to toggle to a third mode: learning mode.
  - When in the object detection mode, the user will be able to scan objects in their environment and see the translation in their chosen target language.
  - When in text recognition mode, the user will be able to scan a word in their environment and see the translation in their chosen target language.
  - **Stretch Requirement:** When in the learning mode, the user will be able to practice the saved words by playing a card guessing game. The user will be able to practice all saved objects or words in a customizable list.
- When an object or word is scanned, the user will be able to click on an external learning page popup link.
  - On the learning page, the user will be able to read a definition of the translation, see an example of a proper sentence usage, see a graphical representation, and click on an audible pronunciation of the detected object / word.
- The user will be able to restart the scanning process as many times as they desire by clicking a button that starts the scanning process over again.
- **Stretch Requirement:** The user will be able to create customizable "PLAYlists" related to a category that they would like to save their scanned objects / words.
  - The user will be able to save the desired scanned words within their account and add them to their previously created or newly created "PLAYlists".

- **Stretch Requirement:** The user will be able to register and create an account:
  time the user accesses the application, a menu will appear requesting login or account creation information.accounts will be based on a secure, unique “Username” and “Password”.
  - The user will be able to log out of the account and continue to use the application without logging in.
  - The application will not require an account to be made but will require an account if the user desires to utilize the learning mode and save their scanned objects and words.

These requirements will be able to be validated by the user by giving a true/false result for switching to a target language for translation, as well as being able to have a visual representation of changing between the object detection mode and text recognition mode, and eventually, a learning mode. Validation of the camera being used will be permitted via a permission prompt requesting camera access, as well as the visual display of the camera working in real time. The ability to click on the external learning page pop-up link will be tested simply by fact checking that the appropriate components are displayed properly on the template page and through various beta tests run by our team before deployment. If the stretch goal is met, creating an account within the Langlens application will be able to be tested by viewing the created account, both within the backend and within the user’s experience, after the user creates it. The user specific functions on the UI will be validated through beta tests that will check for accuracy and performance of the core features presented to the users upon accessing the application.

4.1.2 | Object Detection

Object Detection is a crucial portion of the LangLens application. It is what makes LangLens stand out from other language learning applications, as well as being a cutting edge technology that the members of team LangLens are going to and are excited to explore. Team LangLens is using a Python based program called YOLOv5 to detect objects. Object detection must contain the following requirements.

- Must scan in real time through the user movement of the camera: The object detection mode works in real time allowing the user to keep scanning objects while moving through their environment.

- Must show the user's desired scanned object in a visual bounding box with a high accuracy rate.
- Upon scanning, the translated label presented to the user on the UI must match the target language selected from the target language selection prompt.

Object Detection will be testable in a way of being able to scan an object and if the visual bounding box is correctly in place and the identity of the detected objects can be confirmed in a true or false manner, meaning we have a good requirement for consistent testing.

4.1.3 | Text Recognition

Text Recognition is an important portion of what makes LangLens stick out from other augmented reality language applications, as no application does both text and object detection within the same application. Text recognition will be powered by OpenCV and Tesseract OCR engine. Text recognition will contain the following requirements.

- Must scan text from a picture taken by the user of their current environment.
- OCR must recognise the text in the image for translation.
- Must show the user's desired scanned word in a visual bounding box.

Our text recognition system will be able to be tested by the ability for the system to be able to take a taken picture and recognize the text within a visual bounding box, giving a clear true or false response of the detection of a word. Testing that the text is being correctly taken from the picture into a string to be sent to the translator will be more difficult, but will be able to be tested by correctly setting flags within the software.

4.1.4 | Learning Mode - Stretch Goal Requirement

Learning Mode is an anticipated stretch goal of LangLens where users can practice all that they have saved in both object and text detection modes, and use it as a language learning page. The learning mode provides a card guessing game as a learning method. The learning mode, if implemented, will contain the following requirements:

- Must allow you to create playlists and add objects to them.
- Must show a menu with the created playlists to practice.
- Must retrieve the objects in a playlist along with their images.
- Must mix the objects so they do not appear in the same order every time a playlist is practiced.
- Must allow to pass to the next card and to flip the current card.
- Must display the object translation once the card is flipped.

Learning mode will be able to be tested by a demonstration of the requirement, and confirming the features of being able to create and view playlists as well as retrieve the specified objects within the playlists. Randomizing the order of the objects can be easily tested within a true or false manner, as well as confirming the pass or flip of the current card. Checking the translation of the object when flipped can be done within the code, and will be obvious within the demonstration.

4.1.5 | Translation

Translation will occur after the text or object has been detected. The word or object will be translated from the users base language to the users selected target language. YOLOv5, our object detection program, has built in object translation, while text translation will be handled by the free Google Translate API for Python that is compatible with the image processing handled by OpenCV and the Google Tesseract OCR engine.

- Translate output of object detection to the chosen target language
- Translate output of text detection to the chosen target language
- **Stretch Requirement:**
  - Translate output of object or text detection from any native language to the chosen target language.
  - Translate output of object and text detection from a larger selection of chosen target languages.

Testing translation could be difficult on the software side because words and their translations are not inherently equal, unlike values which are easily comparable. The testing of the translation of the output of the object detection would be a manual process where we check the word that was the output of the object detection, as well as the translated output being the correct translation. Creating a dictionary of words that are translated from one target language to another would help with the testing of translation, but this process would not cover every case. The testing of the translation of the text detection would generally fall under the same process of the testing for the object detection.

4.1.6 | Learning Page

A learning page will be shown to the user after a successful object scan or text detection. This learning page will have its own link that is shown to the user. The contents of the learning page will be the retrieved image, and the word in target language. The point of the learning page is for the user to be able to store words and objects so that they may later use them for learning possibilities in a way that is user friendly. The learning page must follow the following specification:

- External learning page must be clickable via a link to the user after each scan
- Learning page must include an open-source dictionary definition, a sentence usage example, a graphical image representation, and pronunciation example.

Testing within the learning pages will be a crucial part of the process of developing the application. The learning page is an integral part of the purpose of the application and testing to make sure that the users have a good learning experience is important. Testing the creation of the learning page will require tests within both the front and the back-end of the application within testing that the backend is sending the correct information to the learning page which can be tested in a true or false manner, and the front end creating the page which would be tested within bootstraps testing environment.

4.1.7 | Front-end Requirements

The front-end will have a user interface, designed and developed in Bootstrap. Users will interact with the UI to perform the following tasks:

- **User Interface**
  - UI will request camera access from the user displayed as a permission prompt.
  - A target language selection prompt for user's translation preference.
    - displays the preliminary options of Spanish, French and Korean with a note to the user that more languages are in development.
  - Two buttons for toggling between the detection modes: "Detect Word" and "Detect Object".
    - **Stretch Requirement:** A third button available for "Learning Mode" that will only be accessible if a user has created an account.
  - Display learning page link to user immediately following each scan.
  - A button at the bottom of the UI that if clicked will restart the scanning process for unlimited, repeated use.
  - **Stretch Requirements:**
    - The user will be able to create an account.
    - The user will be able to log-in.
    - The user will be able to log-out.
    - A button to save the object or word scanned into a playable list to utilize in the learning mode.
    - The user will be able to select from a larger variety of target languages when prompted with the target language selection screen.
      - The user will be able to suggest languages for translations if their target language is not currently offered via a submission box at the bottom of the target language selection screen.

The user interface and other general front-end requirements will be able to be tested and validated through a bootstrap testing software.

4.1.8 | Back-end Requirements

The back-end which will be powered by Django, a back-end API that can be constructed in Python, will be able to do the following requirements:

- Store the data of the word and its translation within its text database.

- Store the data of the object and its translation within its object database.
- Store the data of the image retrieval for both text and objects.
- Send the data of specified word or object to the learning page.
- Only accessible by admin and not available to users.
- **Stretch Requirement:**
  - Securely store private user account information for log-in and account creation.

The back-end requirements will be able to be tested through Django's extensive built-in testing process.

4.1.9 | Camera Actions

Requirements are needed from the camera to access our application to create a seamless language learning experience. The camera will do the following:

- Use the camera to detect object.
  - using a real-time HTTP livestream as input for the object detection engine (YOLOv5).
- Use the camera to detect words.
  - using a real time HTTP live stream as input for the text detection software (OpenCV) and Tesseract's OCR Engine for text recognition.

4.1.10 | Image Retrieval

Image retrieval will be handled by Unsplash API to search for an image corresponding to the translated word or object. The image retrieval will handle the following functional requirements.

- The system automatically searches a URL image if it has not been previously searched.

- The system automatically stores a URL image if it is searched for the first time.
- The system automatically retrieves a URL image from the database if it has been searched before.

The testing of the image retrieval system will check if the word and translated word match up to the retrieved image.

4.2 | Non-functional Performance Requirements

While in the section above we outlined the functions our application will have, in this section we will discuss how the listed functions will be expected to perform.

4.2.1 | Object Detection

In terms of performance Object Detection has two crucial benchmarks which will need to be met. These are speed, and accuracy. For the application to be usable, we will need it to be very fast as well as accurate when it comes to detecting objects. The application should not take more than 1 second to successfully detect and scan an object, while the accuracy should not fall below 80%. Both of these benchmarks are in the worst case scenario, and we are aiming for our application to be even faster and more accurate.

4.2.2 | Text Recognition

Very similarly to object detection, text recognition has the same two performance expectations in terms of speed and accuracy. Unlike object detection however, text recognition is given a slightly larger buffer in both. Accuracy is expected to stay above 65%, while the text recognition wait time should not be longer than 5 seconds. Similarly to how it was stated above, these are in the worst case scenario, and we are aiming for our application to be both faster and more accurate.

4.2.3 | Translation

When it comes to translation, YOLOv5 has built-in functionality to almost instantly translate scanned objects into a desired language. The Google Translate API for Python works just as seamlessly with the Tesseract OCR to provide translations for scanned

words into the user's desired target language. Giving both a slight buffer, translation should not take more than 1 second.

4.2.4 | Learning Page

The Learning Page will be generated with the data requested via a Bootstrap template, and as a result should take no longer than 5 seconds.

4.2.5 | Front-end Requirements

Front end will provide an intuitive and highly responsive user interface. User's should not need any training to use and navigate the application. In case of a more nuanced feature, there will be an included explanation on how it functions.

4.2.6 | Back-end Requirements

The most important metric of our application's back-end will be responsiveness. Since the back-end is responsible for storing and retrieving data, it should not take more than a second or two to do either.

4.2.7 | Image retrieval

Unsplash API will be responsible for handling image retrieval. It's speed and accuracy. While in this functionality we can sacrifice a bit of speed to guarantee the highest accuracy due to the way the application will handle image retrieval, we still want it to be fast and not take longer than 5 seconds to retrieve an image. Accuracy wise it will be accurate 90% of the time.

4.3 | Environmental Requirements

The following section will describe the non-functional requirements related to the constraints imposed upon the application by the client and the chosen solution software and hardware.

4.3.1 | Web-Based Mobile Application

As requested by our client, LangLens will be a web-based mobile application and will be accessible via any type of web browser (i.e. Google Chrome, Mozilla Firefox,

Safari, etc.) on any type of mobile operating system (i.e. Apple iOS, Google Android, Microsoft's Windows Phone OS, etc.). Due to the nature of the application, the mobile device that is accessing the LangLens web-app will need to have a stable internet connection in order to use the application.

4.3.2 | Free-to-Use Software

LangLens is required to utilize only freely available software and tools and will not utilize any paid services or platforms for development. The client has specifically requested that the LangLens tool be free-to-use by any user and that LangLens should not require the use of a developers license for distribution of the application via an app store.

4.3.3 | Mobile Device with Camera

A hardware requirement that was not explicitly stated by the client but is implied by the functionality of the application is that LangLens will require the use of a mobile device camera. This means that any user accessing LangLens must have a mobile device with access to a camera or if they are accessing the website from a computer or laptop, they must have a webcam.

4.3.4 | Tesseract "Traineddata" Files

The Tesseract OCR Engine chosen for the text recognition aspect of the application comes with the constraint of utilizing only Tesseract's prepackaged language models known as "tessdata" or "traineddata". These files are necessary for the text recognition feature of Tesseract and there are no other externally compatible models approved for use with Tesseract. This is not a limiting constraint however as Tesseract provides over 100 different trained language models specifically designed for their deep learning algorithm and will more than cover the necessary requirements of the text recognition aspect of the LangLens application.

5

Potential Risks

There are a number of different risks that can arise under certain circumstances. Here we analyzed some of the potential risks that could occur during our development process and certain problems that could happen once the product is deployed. The relevance of these potential risks are based on how severe the risk will be from a scale of 1 to 5, where 1 is the least severe and 5 is the most severe. We also go over solutions for each problem at the end of each explanation. There is also a table that describes the severity of each rating and how it can affect the user.

5.1 | The user does not have a camera

A core feature of LangLens revolves around utilizing augmented reality in order to scan objects. There may be a chance that if someone were to use our product that either their mobile or their at home device will not have a camera. With this being said, most devices such as laptops, computers and other mobile devices already have a camera. Though the chances of this happening is low, it has a severity rating of 5 out of 5 mainly because if someone does not have a camera the product cannot be used at all. The solution to this would be on the user's end, meaning they would need a device with a camera to use this application.

5.2 | A certain language is not supported

One of LangLens core functions is the ability to scan an object or text and then translate the detected label to its targeted language, however; there is a chance that there is a language that is not yet offered from our product. The severity of this would

be a 3.5 out of 5 because if someone wants to learn a certain language, but that language is not supported they cannot learn that language. On the other hand, the same can be said for someone who speaks a certain language who wants to learn French for example. The person cannot learn French because their own language that they speak is not supported. A resolution to this would be implementing new language libraries and allowing the user's to request new languages via suggestion box.

5.3 | The scanner does not scan accurately

Another major component is the actual scanning of the objects and text. Once the text or object is scanned, the product will bring up an image of the object, or a translation of the text. There is a chance that the product may bring up the incorrect image or translation because it was scanned incorrectly. This would get a rating of 2.5 out of 5 because it is expected that the scanner will not scan accurately 100% of the time and there may be instances where the scanner may scan something and interpret it as something else. For example, if you scan a bench and it brings up a sofa. A solution to this would be implementing new images of objects that do not exist in the database as well as creating new classes to scan from and offering a manual input option for users to correct any inaccuracy that they are aware of.

5.4 | The product does not translate accurately

Another big component of LangLens is translating. We need to be able to accurately translate objects and text so that the user is able to learn. With that being said, there may be instances where the product will not accurately translate everything either due to the product not being able to interpret slang or certain words and spellings are incorrect. This gets a rating of 3 out of 5 because this would mean that the user's will be misinformed about a certain translated word. A fix to this would be implementing a dictionary that contains the words that are known to cause issues with the translator of the application and having the user edit the response to correct any inaccuracies that they are aware of.

5.5 | The user is not able to save their scanned item

LangLens will also allow the user to save any objects or words that they scan to later practice learning the word or object. One big risk that poses a threat is if the product

does not save the object or word the user scans. This would mean they would not find their scanned items in their saved items and would disencourage the user to learn the language. This is rated a 3 out of 5 because anything that the user can save can be later used in the learning page mode will not show up when they want to learn. To avoid this potential risk, we would implement a way for users to manually input the object or word that they tried to scan and save into their customizable deck within the learning mode. This could be executed via an option within the learning mode UI for adding custom practice words for the case that their desired scanned object / word did not save to their deck.

5.6 | The product does not translate everything

Our product is supposed to translate any scanned items, however, if the user already speaks another language they are able to change the language of the app to their preferred language. This means the entire application is translated to the language so that they are able to use it. One small problem would be that if the product did not translate everything on the application. Though the user will still be able to use the app, certain words or phrases may still be in the default language, which would be English during the development of our first version. This gets a rating of 2 out of 5, since it does not break the app and the user will still be able to use all of the features, but it poses the issue of comprehension on the user's end since they will not be able to understand certain words that were not translated. Here we would have to look through each section of the application in each language to see where the translator is malfunctioning. From there we could have the translator then translate whatever it was not translating before.

Severity	Risk Description	Mitigation
5	The user does not have a camera for the application to use. This means that the user will not be able to use the application at all whatsoever.	For this instance, the user will need to have a device with a camera in order to use the application and there is not much we can do to avoid this risk as it is fully related to the user's personal device.
3.5	Not all languages will be supported by our application. This means people will not be able to learn certain languages or certain people who speak certain languages won't know how to operate the application effectively.	Implementing new language libraries using the language libraries provided by YOLOv5 and Tesseract / Google Translate, as well as giving the users the option to request new languages via a suggestion box.
2.5	The scanner does not scan accurately. This could mean that the scanner scanned the object, but the scanner confuses the object to be a different object, therefore we get an incorrect output.	Adding images of objects that do not exist in the database and implementing new classes of objects to scan as well as offering a manual input option for users to correct any inaccuracy that they are aware of.
3	The application does not translate everything accurately. This could be because it is trying to translate some form of slang, but since slang is different in each country it may be hard to translate it. Therefore, this could potentially produce incorrect output.	Implementing a dictionary that contains words that cause issues with the translator to translate inaccurately, and having the user edit the response to correct any inaccuracy that they are aware of.
3	**Stretch Goal Risk:** The user is not able to save the object or word they scanned. This would mean that if something is scanned and not saved, then the user will not be able to utilize the learning mode effectively as their object or word would not appear in their customizable deck.	To avoid this potential risk, we would implement a way for users to manually input the object or word that they tried to scan and save into their customizable deck within the learning mode. This could be executed via an option within the learning mode UI for adding custom practice words for the case that their desired scanned object / word did not save to their deck.
2	The application does not convert the language of the application correctly. This means that if a user speaks Spanish and wants the application to be in Spanish, there may be a chance that when converting the language of the application not every word is translated.	Here we would have to look throughout the application to see where in the application it is not being converted to the appropriate language. Then, we would have the application print it so that it comes out translated.

6

Project Plan

The following section will go into our project execution plan in the order of what we plan to implement first. Below is a gantt chart that depicts a visual representation of our projected development plan for the 2023 Spring semester. Keep in mind that throughout the development process of our application we will be testing out each implementation of the features in the application to make sure everything is moving smoothly before progressing to the next requirement, as well as making sure that each section we implement works with the other sections as well.

At the beginning of the 2023 Spring Semester, we will create the foundation of our project which requires implementing the basic user interface as the entirety of the application is structured around these features. After creating the basic UI for the application, we will begin implementing object detection and text recognition as these contain many sub-level requirements branched off of the user actions in the UI previously described in the functional requirements section. We anticipate the remaining requirements of translation and the learning page to take longer than establishing the detection engines to ensure efficient and accurate results before polishing the UI and delivering the final product. Therefore, following the implementation of these detection modes is the implementation of the target language selector which is where we will start to develop the translator for the labels of the detected objects and words. This involves translating the scanned object and text to the appropriate language, as well as leaving room for modularity for the anticipated stretch goal of translating the entire application for non-english users. The last feature to implement will be the learning page as this is what ties in all of the previously implemented components corresponding to the object or text detection and is dependent on the preceding requirements being fully

implemented and tested. Once all of the requirements are implemented and working, we will be able to polish the user interface so that it is visually appealing and fully functional for the last step of final product deployment.

LangLens Projected Development Plan - Gantt Chart

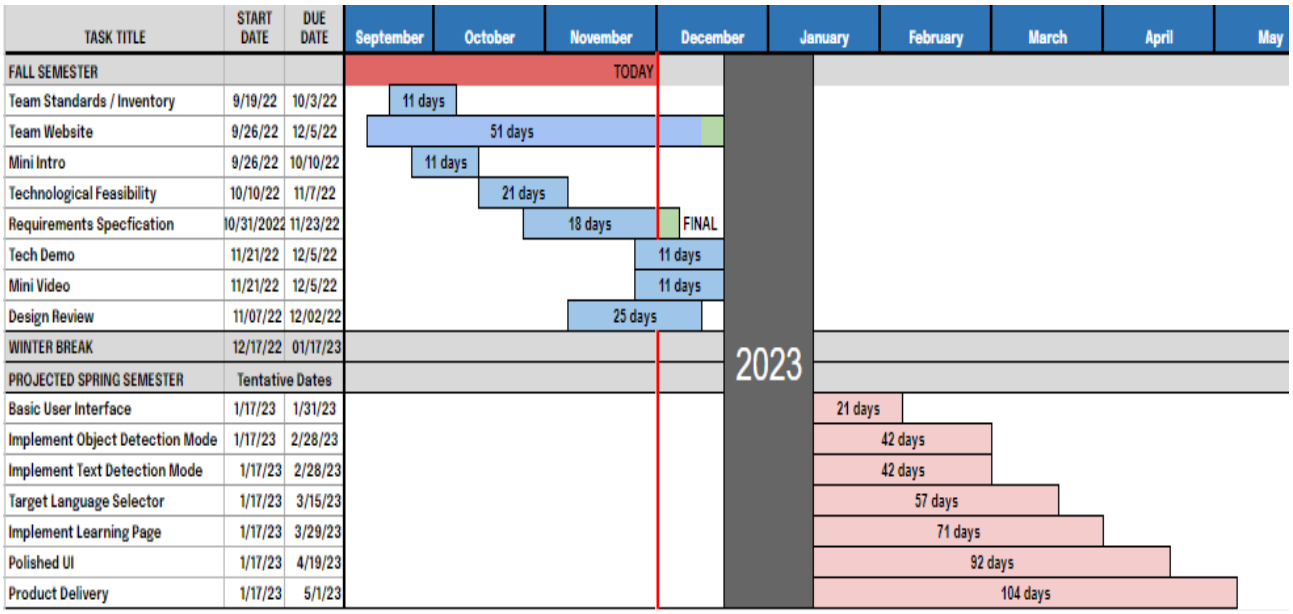

Figure 6.1: LangLens Projected Development Plan - Gantt Chart
*Subject to revision according to deadlines unknown at this time for the 2023 Spring Semester.
**Testing will occur after each requirement implementation

7

Conclusion

To fill the market gap of the observed lack of tools targeted towards foreign language learners that are modern, free, and emphasize the core of language learning pedagogy, LangLens has partnered with Dr. Okim Kang to create an innovative, web-based tool, designed for all mobile devices, that provides a free-to-use, immersive, language learning experience employing both object detection and OCR capabilities with the latest AR technology available.

LangLens intends to solve this problem by creating a website application that can serve as a reinforcement tool when learning a new language in a more interactive way than simply reading workbooks or through the existing learning applications on the market today. Users will be able to scan their environment in real time and detect objects and words, see their target language translation and be exposed to the key elements of word learning via a learning page which provides visual, textual, and auditory examples crucial to language learning. This turns the application not only into a translator, but into a language learning tool specifically designed for people of all backgrounds to simply pick up their mobile device and learn a new language whenever and wherever they are.

Throughout this document we have reviewed our extensive research to establish all of the functional requirements, non-functional (performance) requirements and environmental requirements to provide detailed descriptions of the expected functionality of the application which meet the expectations imposed by our client.

Now that we have already done a technological feasibility analysis, proposed a solution vision and established all the requirements for our application, we can start de-

veloping individual demonstrations for the core components of the application and test them to ensure that they fulfill their desired responsibilities. Once we have tested every individual demo, the following steps will be to integrate every component based on the architecture designed and deploy the final product. LangLens is confident in taking these next steps toward development and is positive that the team will be able to provide a fully functional, language learning tool that utilizes the latest AR advances and features described in this document by the end of the 2023 Spring Semester.