Bilkent University

Department of Computer Engineering

Senior Design Project

AnnoHub

Detailed Design Report

T2331

Emirkan Derköken, Bora Yılmaz, Dağhan Ünal, İlker Özgen, Onurcan Ataç

Supervisor: Shervin R. Arashloo

Table of Contents

1. Introduction

1.1 Purpose of the System

In the field of machine learning and artificial intelligence, high-quality labeled data is crucial for building quality models. However, large-scale data labeling poses challenges particularly in ensuring accuracy and consistency of the models. Additionally; handling sensitive data adds complexity, requiring a balance between having the dataset annotated and safeguarding privacy and corporate secrets.

Beyond the technical intricacies, sourcing a skilled workforce for data annotation poses a fundamental challenge, requiring careful consideration from recruitment to ongoing training. In contrast to the challenges of acquiring skilled labor for such tasks, finding skilled workers for simple data annotation is also difficult. These spikes in demand occur sporadically throughout the life-cycle of a business. As a result; companies are reluctant to establish in-house teams, given the irregular demand, making it economically impractical. Simultaneously, potential workers are deterred by the lack of stability in roles, resulting in a scarcity of individuals willing to perform straightforward data annotation tasks.

AnnoHub is designed to address the main challenges of data annotation explained above while offering a high-quality user experience through easy to use and user-friendly interfaces. Packed with many features that enhance the quality of life for both annotators and companies, ensuring an efficient and effective annotation process.

1.2 Design Goals

1.2.1 Scalability and Performance

- The application should be able to handle and optimize increasing volumes of data without compromising performance.
- AnnoHub needs scalable workforce management to scale labeling capacity up or down based on demand.
- The application should be able to integrate optimization tools to further enhance performance when needed.

1.2.2 Data Privacy and Security

- The project will include security facilities such as encryption of sensitive data at rest and in transit, data masking and anonymization tools to protect privacy.
- AnnoHub should also take into consideration the industry standards, state laws and regulations into account regarding data privacy and security.

1.2.3 System Availability and Reliability

- Reliable data storage and backup solutions should be integrated to AnnoHub.
- The application will need error handling and recovery mechanisms to ensure system stability and reliability.

1.2.4 Extensibility and Modularity

- The application should be built using a modular system design to allow for the addition of new features and components more easily.
- AnnoHub should be extensible in order to easily integrate with other systems and tools.
- API documentation for developers to extend and integrate the system will be needed.

1.2.5 Interoperability

- Standards-compliant data formats should be present to ensure interoperability with other systems.
- SDKs will also be needed for integration with external systems and data sources.
- The application should be compatible with common data labeling and machine learning tools.

1.2.6 Testability

- The application must be testable through various testing tools and frameworks to validate system functionality, stability, reliability and performance.

1.3 Overview

AnnoHub is a platform for thorough data annotation that supports two different modes of operation. Users have the option to openly share unlabeled datasets and hire independent annotators to work together to annotate them. Compared to engaging in-house annotators, this operational mode relieves publishers of the effort of finding the required manpower and annotates their data at a more economical rate. Furthermore, this alternative offers freelance annotators a new stream of side income that practically everyone may tap into. Alternatively, companies have the option to create private data annotation projects with their datasets, utilizing the platform as a sophisticated labeling tool to enlist annotators of their choosing.

AnnoHub integrates certain features to improve the consistency and accuracy of annotations, especially those made by external annotators on publicly available datasets, in order to look for potential contradictions both between and among annotators. These include methods for verifying the accuracy of the annotators' work,

like integrating known-annotation trick data into the dataset, self-validation, which requires annotators to label the same data multiple times, and cross-validation, which involves the same data being annotated by multiple annotators. Simultaneously; by the implementation of fraud detection algorithms, the negative effects of malicious use are aimed to be minimized. These measures and the information they are going to present to the companies collectively enhance the reliability and quality assurance of the annotation process on our platform.

Finally, owing to AnnoHub's system architecture prioritizing enhanced security, in cases where companies choose to create a private data annotation project, the data is going to bypass our servers entirely. Instead, it is directly shared with the annotators, ensuring the safeguarding of company secrets and adherence to data protection laws not through bonds, but through the system design.

2. Current Software Architecture

The utilization of AI and machine learning methodologies is expanding across diverse sectors, leading to an increased demand for high-quality labeled datasets. As of 2022, the global machine learning market attained a valuation of USD 36.73 billion, with an anticipated compound annual growth rate (CAGR) of 34.8% from 2023 to 2030. Consequently, it is plausible that existing tools in the market exhibit similar functionalities to our application. To address this, we conducted a comprehensive market analysis, identifying key competitors for AnnoHub. Following the research, three prominent tools -Labelbox, Label Studio, and Amazon Sagemaker Ground Truth- emerged as primary contenders due to their widespread usage and analogous functionalities.

Labelbox, LabelStudio, and Amazon Sagemaker Ground Truth are all comprehensive data labeling tools, they come with their advantages and disadvantages. They are advanced platforms employed for annotating datasets to facilitate the training of machine learning models. Labelbox offers robust capabilities for managing extensive datasets while providing customizable annotation workflows. LabelStudio, being open-source, offers flexibility in annotation tasks, supporting iterative model refinement processes. In contrast, Amazon SageMaker Ground Truth, as a managed service, ensures scalability and accuracy through automated labeling mechanisms complemented by human review functionalities. Labelbox and Amazon Sagemaker are known for their prices being on the higher end after passing a limit, and it is difficult to pay for some publishers. Moreover, Labelbox and Amazon Sagemaker partially support outsourcing labeling, neither provides a comprehensive public workforce system. Although Amazon employs a similar system known as Amazon Mechanical Turk, our project's innovation expert, Ahmet Kocamaz, expressed dissatisfaction with the service due to inadequate information about labeled data metrics. Mechanical Turk is also notorious for not returning to the applications of the public.

AnnoHub aims to solve the main issues that the existing software products present. To solve these issues, AnnoHub aims to establish a public workforce, allowing customers the option to label their data through this workforce. Customers can select qualified public labelers via proficiency tests tailored to the labeling task, and publishers also can require specific documents from the labelers. AnnoHub strives to offer both cost-effective labeling and job opportunities through this system, a distinctive feature absent in competitors. By this system,

AnnoHub will have a massive advantage over the use in academia and startup companies, because of the cheaper pricing enabled by the public workforce.
Another primary challenge addressed by our application is the facilitation of data labeling without the necessity of sharing sensitive information with external servers. Labelbox mandates customers to upload data on Google Cloud Services, storing it compulsorily in the US. Catering to clients unwilling or unable to share data beyond their national borders (e.g., due to KVKK laws in Türkiye), AnnoHub offers a solution by maintaining only references to sensitive data on its servers and sharing the actual data exclusively with designated labeler employees. Consequently, AnnoHub ensures privacy by design, alleviating the need to entrust external entities with sensitive information.

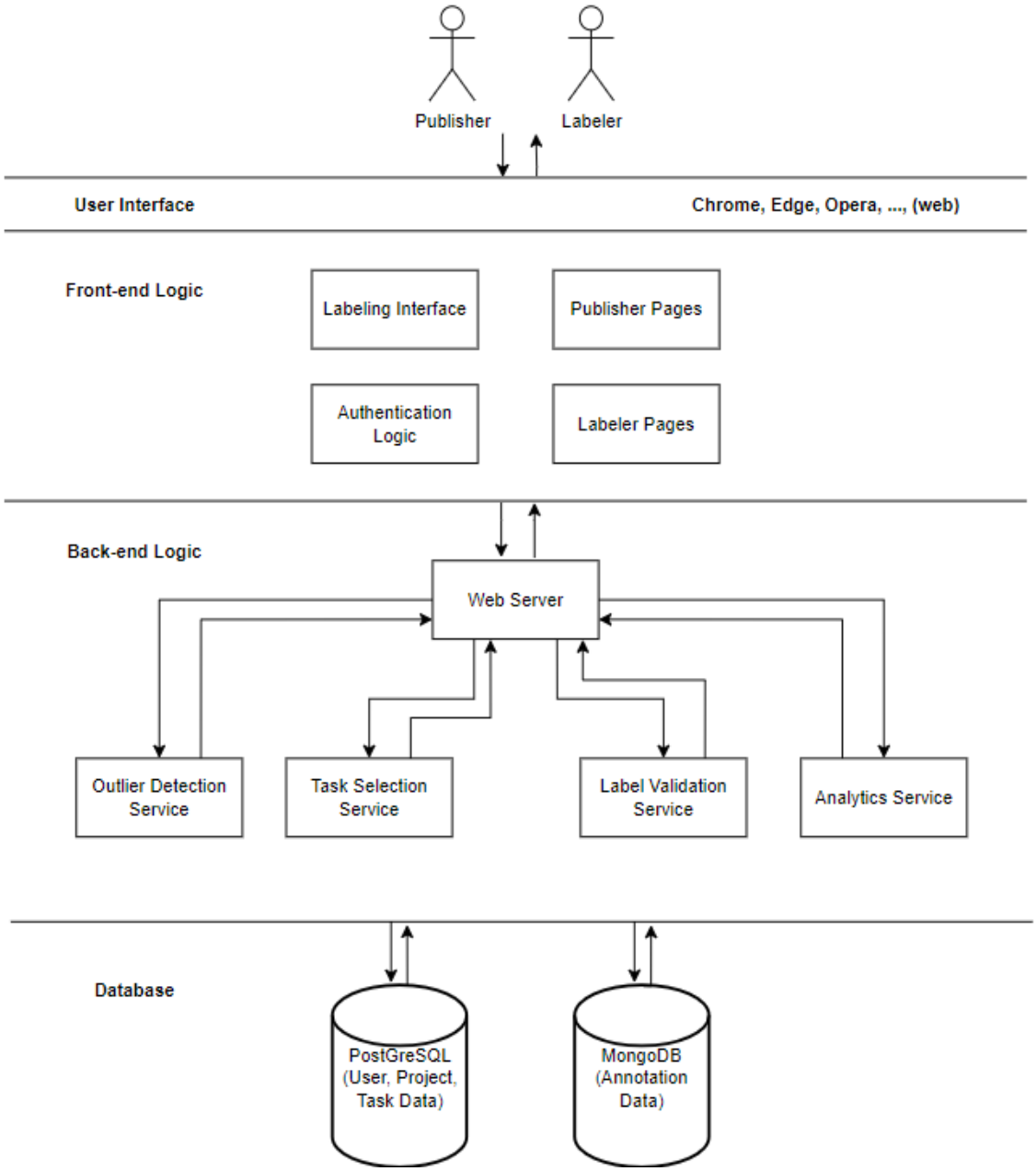
AnnoHub will not support model training like Amazon Sagemaker Ground Truth, but it will match the labeling functionalities by providing interfaces for various data types. Moreover, AnnoHub employs machine learning algorithms, including cross-validation, self-validation, and fraud detection algorithms. With the help of these algorithms, AnnoHub will also provide extensive analytics and metrics to the publishers. These mechanisms, coupled with the public workforce system, underscore AnnoHub's commitment to ensuring efficient and effective labeling, ultimately contributing to heightened customer satisfaction.

3. Proposed System Architecture

3.1 Overview

The software architecture for our project is designed to satisfy our functional requirements, ensuring the system's effectiveness and scalability. At its core, the architecture is structured around clear layers, enabling modular development and facilitating future enhancements. These layers are delineated through diagrams and class representations, offering a visual roadmap. Persistent data management is a critical aspect, with our architecture supporting various database systems tailored to handle diverse data types and volumes. Functional requirements dictate the need for efficient data ingestion and partitioning capabilities, handling diverse data types such as text, images, audio, video, and 3D point clouds. Automatic and manual partitioning mechanisms ensure privacy and manageability while also providing clients with flexibility in data storage options. Furthermore, robust data labeling and annotation functionalities are integrated, featuring a user-friendly interface and support for tasks such as classification, object detection, and semantic segmentation. Quality assurance measures, including automated checks and real-time monitoring, are embedded within the architecture to maintain label accuracy and ensure high-quality data outputs. Supervisors are empowered with tools to review and correct labels, rate labelers based on performance, and conduct qualification tests to certify labelers' competency. Data aggregation and retrieval mechanisms enable the efficient organization and access of labeled datasets, while role-based access control and security protocols safeguard sensitive information. These functionalities are seamlessly integrated into the architecture, providing a solid foundation for the development and deployment of our project.

3.2 Subsystem Services



3.3 Persistent Data Management

The user data will be stored in AnnoHub's cloud servers. In these servers, severely sensitive information such as user passwords are going to be encrypted using the one-way SHA-256 password hashing algorithm, in order to prevent user passwords from being stolen by both internal and external cyber attacks. The remaining user data will be stored after being encrypted using the symmetric-key encryption algorithm AES-128 in order to protect it from third-party attackers.

The publishers may choose to upload their datasets to AnnoHub's cloud servers, or they may store and serve their own datasets in their own servers. If the

publishers choose to upload their datasets to AnnoHub's servers, the datasets will be encrypted using the AES-128 algorithm before being physically stored and will be decrypted before being served to the annotators. In the case that the publishers choose to store and serve their datasets using their own servers, the annotators will iteratively request the data from the publisher's servers directly without AnnoHub acting as an intermediary. However, since the front-end logic is served by AnnoHub, annotators are still going to be expected to contact AnnoHub's servers and additionally, since the back-end logic is also served by AnnoHub's servers, certain features such as the "outlier detection" will not be operational for publishers that choose to store their datasets in their own servers.

Finally, all the communication between the publishers, the annotators and the AnnoHub's servers will be carried out using the HTTPS protocol, which guarantees the safety and the privacy of the data throughout the transportation.

3.4 Access Control and Security

In terms of access control, AnnoHub is complicated and has lots of cases in order to give access to only the authorized users the abilities to work with the data and projects. Currently, AnnoHub architecture has two main types of users: Labeler and Publisher. According to access control, publishers should only be able to have authorization over managing the projects that they have created. In a similar fashion, the labelers should only be able to access the projects that they are a part of and should only be able to label the data instances that have been assigned to them.If a project is initialized as private, that project should not be seen by the labelers when they try to search for that project through the dashboard, even after exact filtering. Furthermore, as labeler and publisher roles are not designed to be crossed (in other words, a user will not be a labeler and a publisher at the same time), further authorization revisions will be needed. Although these are among the majority of the access control cases, more authorization scenarios are present in AnnoHub. To sum up, in terms of access control, the application has a lot of cases and possible scenarios despite having only two major roles in practice; so, in order to achieve data privacy and authorization security, AnnoHub has many authorization rules.

In terms of the application's security, Annohub has a multi-layer security approach. For data security, encryption serves an important part, safeguarding sensitive data both stored and processed actively. Utilizing strong encryption throughout the application is therefore one of the security features to be benefited from. As the application is initially planned to be running on the cloud services, it will also use the cloud security facilities as much as possible, alongside with the cloud provider's firewall utilities. These utilities include access controls and regular audits to monitor for any vulnerabilities, or potential privacy and security attacks. Additionally, the application has been adjudicated to have end-to-end encryption, which will further enhance confidentiality as well as user trust, guaranteeing that data remains unreadable to any intermediaries. Furthermore, continuous security updates and patches will ensure that our application will keep on being resilient against emerging threats, which will be another trust factor for the users.

4. Test Cases

4.1 Functional Test Cases

4.1.1. Create Project Tests

Test ID	UI_CP_1	Category	Functional	Severity	High

Objective	To ensure creating project functionality works properly.
Date	13.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Main Page.	Display the page content.	
2	Click on the ‘’New Project” button.	Display the create project pop up.	
3	In the “Project Settings” tab, fill in the “Project Name”, “Description”, “Cross-Validation Count”, “Self-Validation Frequency”, “Project Access”, “Price Per Assignment”, “Hiring Status”, “Due Date”, and “Thumbnail” informations properly.	Save the settings to local storage until finalizing the project creation.	
4	Navigate to the “Labeling Setup” tab, fill in the labeling setup information properly.	Save the settings to local storage until finalizing the project creation.	
5	Navigate to the “Labeling Examples” tab, fill in the example labeling information properly.	Save the settings to local storage until finalizing the project creation.	
6	Finalize creating the project by clicking the save project.	Display the new project in the “Active Projects” tab in the main page.	

4.1.2. Project Settings Tests

Test ID	UI_PS_1	Category	Functional	Severity	Medium

Objective	To ensure the ‘General’ tab in settings page and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Settings Page.	Display the page content.	
2	Click on the ‘General’ Tab.	Display the Tab Content.	
3	Update project name and description.	Project name and description fields will be set for updating.	
4	Under the ‘General Properties’ section, update the modifiable parts (such as deadline and cross validation count).	Other modifiable fields will be set for updating.	
5	Update the thumbnail by dragging and dropping an image file.	New image should be seen from the thumbnail part.	
6	Try to update the thumbnail by dragging and dropping a non-image file and get an error.	New image should be seen from the thumbnail part.	
7	Click on the ‘Save Changes’ button.	Update the project according to the changes made.	

Test ID	UI_PS_2	Category	Functional	Severity	Medium

Objective	To ensure the ‘Labels’ tab in settings page and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Settings Page.	Display the page content.	
2	Click on the ‘Labels’ Tab.	Display the Tab Content.	
3	Under the ‘Add Label’ section, enter a new label name to the text field.	Label name field will be filled.	
4	Under the ‘Add Label’ section, choose a color for the label from the color palette.	A custom color to the label is given.	
5	Click on the ‘Add Label’ button.	The new label should be added to the project.	
6	Under ‘Current Labels’ section, see the newly added label.	The new label should be seen at the tablet.	

Test ID	UI_PS_3	Category	Functional	Severity	Medium

Objective	To ensure the ‘Add Labelers’ subtab in ‘Labelers’ tab in settings page and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Settings Page.	Display the page content.	
2	Click on the ‘Labelers’ Tab.	Display the Tab Content.	
3	Click on the ‘Add Labelers’ subtab.	Display the subtab Content.	
4	Under ‘Labeler Candidates’ section, see the labelers to be included to the project and check a candidate via the respecting checkbox, under ‘Accept’ column.	Respective checkbox components should be checked.	
5	Click on the ‘Add Labelers’ button.	The selected labelers will be added to the project.	

Test ID	UI_PS_4	Category	Functional	Severity	Medium

Objective	To ensure the ‘See Suspended Labelers’ subtab in ‘Labelers’ tab in settings page and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Settings Page.	Display the page content.	
2	Click on the ‘Labelers’ Tab.	Display the Tab Content.	
3	Click on the ‘See Suspended Labelers’ subtab.	Display the subtab Content.	
4	Under ‘Suspended Labelers’ section, see the suspended labelers to be excluded from the project and acquit a labeler by clicking on the button under ‘acquit’ column.	A popup for executing the action should be shown.	
5	See the necessary popup and press the ‘Accept’ button.	The selected labeler should be able to keep labeling data as usual.	
6	Under ‘Suspended Labelers’ section, see the suspended labelers to be excluded from the project and exclude a labeler by clicking on the button under ‘kick’ column.	A popup for executing the action should be shown.	
7	See the necessary popup and press the ‘Accept’ button.	The selected labeler should be excluded from the project.	

Test ID	UI_PS_5	Category	Functional	Severity	Medium

Objective	To ensure the ‘Workflow’ tab in settings page and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Settings Page.	Display the page content.	
2	Click on the ‘Workflow’ Tab.	Display the Tab Content.	
3	Under the ‘Remove Data from Workflow’ section, see the current data instances included to the project and check a candidate via the respecting checkbox, under ‘Remove’ column.	The checkboxes respective to the chosen data instances to be removed should be selected.	
4	Click on the ‘Remove Instances’ button.	The data and all related labeling to that data instance should be removed.	
5	Under the ‘Import Data to Workflow’ section, drag and drop files to be added to the project workflow.	The dragged and dropped files should be uploaded under the project’s data instances.	
6	Under the ‘Import Data to Workflow’ section, click on the field below and choose the files to be added through your OS’ file manager.	The chosen files should be uploaded under the project’s data instances.	

Test ID	UI_PS_6	Category	Functional	Severity	Medium

Objective	To ensure the ‘Instructions’ tab in settings page and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Settings Page.	Display the page content.	
2	Click on the ‘Instructions’ Tab.	Display the Tab Content.	
3	Under the ‘Instruction Details’ section, see the current instruction and update the instruction through the text field.	Modified instruction should be in the field.	
4	Click on the ‘Update Instructions’ button.	Project instructions should be updated.	
5	Under the ‘Example Label’ section, click on the ‘Update Label’ section.	Should be forwarded to the labeling interface for updating the example.	
6	Update the example by changing the example data instance and labeling it via using the labeling interface.	New labeling should be made on the updated example data instance.	
7	Update the description of the example from the text field.	The description related to the example should be modified.	
8	Click on ‘Update Example’ button.	Example label should be updated.	

4.1.3. Publisher Project Details Tests

Test ID	UI_PPD_1	Category	Functional	Severity	Medium

Objective	To ensure the Analytics Tab and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Details Page.	Display the page content.	
2	Click on the Analytics Tab.	Display the Analytics Tab Content.	
3	Click on the Name column in the Annotators Table 2 times.	Change the sorting order after each click.	
4	Click on the E-Mail column in the Annotatators Table 2 times.	Change the sorting order after each click.	
5	Click on Pagination Buttons in the Annotator Table.	Change the page of the table accordingly.	
6	Click on Pagination Buttons in the AI Report Table.	Change the page of the table accordingly.	
7	Click on Pagination Buttons in the Annotation History Table.	Change the page of the table accordingly.	

Test ID	UI_PPD_2	Category	Functional	Severity	Medium

Objective	To ensure the Workflow Tab and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Details Page.	Display the page content.	
2	Click on the Workflow Tab.	Display the Workflow Tab Content.	
3	Click on each column in the table 2 times.	Change the sorting order after each click.	
4	Click on Pagination Buttons in the table.	Change the page of the table accordingly.	
5	Provide a unique existing name in the search bar.	Filter the table content and show the row with that name.	
6	Provide a non-unique existing name in the search bar.	Filter the table content and show all the rows with that name.	
7	While the search filtering is active with a non-unique existing name, click on each column in the Annotators Table 2 times again.	Change the sorting order after each click while only showing all the rows with that name.	
8	While the search filtering is active with a non-unique existing name, click on Pagination Buttons in the table.	Change the page of the table accordingly while still showing the correct rows.	
9	Provide a non-existing name in the search bar.	Filter the table content and show an empty table.	
10	Empty the search bar.	Stop filtering the table content.	

Test ID	UI_PPD_3	Category	Functional	Severity	Medium

Objective	To ensure the Annotator Details Tab and its content work correctly and as expected, as well as all its functionality.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Details Page.	Display the page content.	
2	Click on the Annotator Details Tab.	Display the Annotator Details Tab Content.	
3	Click on each column in the table 2 times.	Change the sorting order after each click.	
4	Click on Pagination Buttons in the table.	Change the page of the table accordingly.	
5	Provide a unique existing name in the search bar.	Filter the table content and show the row with that name.	
6	Provide a non-unique existing name in the search bar.	Filter the table content and show all the rows with that name.	
7	While the search filtering is active with a non-unique existing name, click on each column in the Annotators Table 2 times again.	Change the sorting order after each click while only showing all the rows with that name.	
8	While the search filtering is active with a non-unique existing name, click on Pagination Buttons in the table.	Change the page of the table accordingly while still showing the correct rows.	
9	Provide a non-existing name in the search bar.	Filter the table content and show an empty table.	
10	Empty the search bar.	Stop filtering the table content.	

Test ID	UI_PPD_4	Category	Functional	Severity	Medium

Objective	To ensure the general components of the Publisher Project Details Page work correctly and as expected.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Publisher Project Details Page.	Display the page content.	
2	Click on the New Project Button.	Display the Create Project Page content.	
3	Click on the Settings Button.	Display the Settings Page content.	
4	Click on the Sidebar Drawer Button.	Collapse the Sidebar.	
5	Click on the Sidebar Button.	Open back the Sidebar.	
6	Navigate between different projects in the sidebar.	Display the page content dedicated to the clicked project.	

4.1.4. Annotator Project Details Tests

Test ID	UI_APDU_1	Category	Functional	Severity	Medium

Objective	To ensure that the data on the Annotator Unregistered Project Details Page is displayed correctly.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Annotator Dashboard Page.	The page should be loaded.	-
2	Click to the View button of the first suggested project shown on the Annotator Dashboard Page.	The user should be navigated to the Annotator Unregistered Project Details Page of the selected project.	-
3	Check if the project information presented on the page (image, title, description) matches the project clicked on the Annotator Dashboard Page.	The project information should match the project clicked on the Annotator Dashboard Page.	-
4	Check if the other information spots on the page about the project and the publisher (publisher name, project type, project access status, hiring status, due date, price per annotation, completion percentage, information about the other annotators, publisher statistics, publisher information) are shown on the page.	The information on the indicated spots should be shown on the page.	-
5	Check if the annotation information chart's data and distribution matches the shown information on the page.	The annotation information chart's data and distribution should match the shown information on the page.	-

Test ID	UI_APDU_2	Category	Functional	Severity	Medium

Objective	To ensure that the buttons on the Annotator Unregistered Project Details Page function as intended.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Annotator Dashboard Page.	The page should be loaded.	-
2	Navigate to the Annotator Unregistered Project Details Page by clicking to the View button of the first suggested project shown on the Annotator Dashboard Page.	The page should be loaded.	-
3	Click to the Apply Now button.	The user should be navigated to the Annotator Application Page.	-

Test ID	UI_APDR_1	Category	Functional	Severity	Medium

Objective	To ensure that the data on the Annotator Registered Project Details Page is displayed correctly.
Date	07.03.2024

#	Steps	Expected	Result
1	Click on the button that opens the sidebar.	The sidebar along with the active projects should be displayed.	-
2	Click to the first active project shown on the sidebar.	The user should be navigated to the Annotator Registered Project Details Page of the selected project.	-
3	Check if the project title presented on the page matches the project clicked on the sidebar.	The project title should match the project clicked on the sidebar.	-
4	Check if the other information spots on the page about the project and the publisher (project image, description, publisher name, project type, due date, price per annotation, annotator contribution information) are shown on the page.	The information on the indicated spots should be shown on the page.	-
5	Check if the annotation information bar chart is visible on the page.	The annotation information bar chart should be visible on the page with 7 days of data.	-
6	Click to the dropdown of the annotation information bar chart and change the period from “Last 7 days” to “Last Month”.	The displayed annotation information bar chart should be visible with 30 days of data instead of 7 days.	-

Test ID	UI_APDR_2	Category	Functional	Severity	Medium

Objective	To ensure that the buttons on the Annotator Registered Project Details Page function as intended.
Date	07.03.2024

#	Steps	Expected	Result
1	Click on the button that opens the sidebar.	The sidebar along with the active projects should be displayed.	-
2	Click to the first active project shown on the sidebar.	The user should be navigated to the Annotator Registered Project Details Page of the selected project.	-
3	Click to the Continue Labeling button.	The user should be navigated to the Labeling Interface Page.	-

4.1.4. Annotator Dashboard Tests

Test ID	UI_ADB_1	Category	Functional	Severity	Medium

Objective	To ensure that the data of the annotator on the Annotator Dashboard Page is displayed correctly.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Annotator Dashboard Page.	Page should be loaded.	-
2	Check if the username on the navigation bar is correct.	The username that is displayed should match the username of the signed in annotator.	-
3	Check if the first chunk of suggested projects are loaded.	Suggested projects should be displayed properly with their images, titles and descriptions.	-
4	Click on the button that opens the sidebar.	Names of the active projects should match with the active projects that user has contributed to.	-

Test ID	UI_ADB_2	Category	Functional	Severity	Medium

Objective	To ensure that the buttons on the Annotator Dashboard work as intended.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Annotator Dashboard Page.	Page should be loaded.	-
2	Click to the username of the user on the navigation bar or the dropdown symbol beside it.	The dropdown that displays “My Account” and “Logout” options should be shown to the user.	-
3	Click to View button on the first suggested project.	The user should be navigated to Unregistered Project Details Page of the selected project.	-
4	Click on the button that opens the sidebar.	Names of the active projects should match with the active projects that user has contributed to.	-
5	Click to the first active project of the annotator if there is an active project.	The user should be navigated to the Registered Project Details Page of the selected project.	-

Test ID	UI_ADB_3	Category	Functional	Severity	High

Objective	To ensure that the search bar and the lazy loading feature on the Annotator Dashboard work as intended.
Date	07.03.2024

#	Steps	Expected	Result
1	Navigate to the Annotator Dashboard Page.	Page should be loaded.	-
2	Move downwards on the page until the projects shown are passed and the loading animation comes up.	New chunk of suggested projects should be loaded.	-
3	Click to the search bar on the navigation bar, type a query, click to the enter button or the search (magnifying glass) icon.	Existing projects on the dashboard should disappear, new chunk of projects should be loaded according to the query entered.	-
4	Move downwards on the page until the projects shown are passed and the loading animation comes up.	New chunk of projects that match with the entered query should be loaded.	-

4.1.5. Labeling Interface

Test ID	UI_LI_1	Category	Functional	Severity	Low

Objective	To test if the labeling interface hotkeys are working correctly
Date	12.03.2024

#	Steps	Expected	Result
1	Press 1-9	The class assigned to the pressed number is selected.	
2	Press ctrl + 1-9	The labeling assigned to the pressed number is selected.	
3	Press ctrl + enter	Submit the labeling.	
4	Press ctrl + space	The task is skipped.	
5	Press backspace	The selected labeling is deleted.	
6	Press ctrl + backspace	All of the labels are removed.	

Test ID	UI_LI_2	Category	Functional	Severity	Medium

Objective	To test if the labelings are properly saved within a session
Date	12.03.2024

#	Steps	Expected	Result
1	Submit a task	The task is submitted and the next task is loaded.	
2	Go back by pressing the “<” button in the application	The task that was just labeled is loaded with the correct labelings.	
3	Reload the page	The same task (at the end of step 2) is loaded once again with the correct labelings.	
4	Press the “>” button in the application	The new task that was loaded in at the end of step 1 is loaded once again.	
5	Press the left arrow button on the browser to navigate back	The previous task (at the end of step 2) is loaded with the correct labelings.	

Test ID	UI_LI_3	Category	Functional	Severity	Low

Objective	To test if the “Instructions” and “Report” buttons are working correctly
Date	12.03.2024

#	Steps	Expected	Result
1	Press the “instructions” button.	The instructions panel pops out and renders the HTML code for the instructions.	
2	Click the “close” button.	The instructions panel is closed.	
3	Press the “report” button.	The new panel pops out, asking for the category and the description of the problem.	
4	Choose a category and fill in the description text box and press the submit button.	The inputs are submitted to the back-end and the next task is loaded.	

4.2 Non-functional Test Cases

4.2.1 Security Tests

Test ID	WS_S_1	Category	Non-Functional	Severity	High

Objective	To ensure that after a publisher authenticates, only the projects that the user has published should be accessible.
Date	07.03.2024

#	Steps	Expected	Result
1	Authenticate by entering email and password or via Google authentication	Authenticate to the application.	
2	On the left sidebar, see the projects that the authenticated user has published, click on any one of those projects.	See only the published projects.	
3	Interact with selected project's analytics and settings for Project administration.	Change the sorting order after each click.	

Test ID	WS_S_2	Category	Non-Functional	Severity	High

Objective	To ensure that when an authenticated but unauthorized user tries to navigate to a project's administration panels through URL, that user should not be able to reach to the project administration page.
Date	07.03.2024

#	Steps	Expected	Result
1	Authenticate by entering email and password or via Google authentication.	Authenticate to the application.	
2	Enter the URL of a project administration page.	Get 401-Unauthorized response. The user should not be able to access the page.	

Test ID	WS_S_3	Category	Non-Functional	Severity	High

Objective	To ensure that when a labeler tries to access a Publisher-related page through URL, that user should not be able to access that page.
Date	07.03.2024

#	Steps	Expected	Result
1	Authenticate by entering email and password or via Google authentication as a labeler.	Authenticate to the application.	
2	Enter the URL of a Publisher-related page.	Get 401-Unauthorized response. The user should not be able to access the page.	

Test ID	WS_S_4	Category	Non-Functional	Severity	High

Objective	To ensure that when a publisher tries to access a labeler-related page through URL, that user should not be able to access that page.
Date	07.03.2024

#	Steps	Expected	Result
1	Authenticate by entering email and password or via Google authentication as a publisher.	Authenticate to the application.	
2	Enter the URL of a labeler-related page.	Get 401-Unauthorized response. The user should not be able to access the page.	

Test ID	WS_S_5	Category	Non-Functional	Severity	High

Objective	To ensure that when an unauthenticated user tries to enter to any authentication-required page of the application (i.e., any page except login/register pages) through URL, user cannot access those pages.
Date	07.03.2024

#	Steps	Expected	Result
1	While being unauthenticated, enter the URL of an authentication-required page of the application.	The user should not be able to access the page. Get forwarded to the login page.	

Test ID	WS_S_6	Category	Non-Functional	Severity	High

Objective	To ensure that if an unauthorized labeler tries to access and label an instance (through the labeling interface) by entering the URL, the labeler cannot access that instance.
Date	07.03.2024

#	Steps	Expected	Result
1	Enter the URL of an unauthorized project's label instance.	The user should not be able to access the page. Get 401-Unauthorized response.	

Test ID	WS_S_7	Category	Non-Functional	Severity	Medium

Objective	To ensure that when a private project is not seen to the labelers who are not included in that project.
Date	07.03.2024

#	Steps	Expected	Result
1	Authenticate by entering email and password or via Google authentication as a labeler.	Authenticate to the application.	
2	Go to the projects page.	Access projects page.	
3	Scroll the page to check for the private project.	Should not be able to see that private project listed.	
4	Search that private project's information through the search bar at navbar.	Should not be able to see that private project listed.	

Test ID	WS_S_8	Category	Non-Functional	Severity	Medium

Objective	To ensure that if a user's bearer token is changed (damaged), user's authentication should be automatically disabled.
Date	07.03.2024

#	Steps	Expected	Result
1	Choose a page and in the source code of that page, manually damage the token (for any user).	When the page is loaded, it should damage the authenticated user's token.	
2	With an authorized user, enter that chosen page.	The user's token should be damaged. Should get redirected to the login page.	
3	Get rid of the source code update.	The token damaging code should be removed.	

Test ID	WS_S_9	Category	Non-Functional	Severity	High

Objective	To ensure that the data stored in the databases is properly encrypted
Date	10.03.2024

#	Steps	Expected	Result
1	Read the “Users” table from the database.	The email addresses and password of the users are obtained.	
2	Use the email addresses and password obtained in order to log in to the application.	No authentication was made with the credentials since the data read in step 1 are encrypted.	

Test ID	WS_S_10	Category	Non-Functional	Severity	High

Objective	To ensure that the private datasets stored in the client's servers does not pass through AnnoHub's servers
Date	10.03.2024

#	Steps	Expected	Result
1	Start a network sniffer application.	The network traffic is being monitored.	
2	Request the data from the publisher's server, annotate it and submit the annotation	The annotation is submitted	
3	Examine the network traffic to see if the data requested from the publisher's server has been sent to the AnnoHub's servers.	The data requested from the publisher's server were not sent to any other address.	

4.2.2 Reliability Tests

Test ID	WS_R_1	Category	Non-Functional	Severity	Medium

Objective	To ensure that while 50 users are authenticated, make sure 90% of user requests are handled in less than 5 seconds.
Date	07.03.2024

#	Steps	Expected	Result
1	Write a script to perform 2 GET requests and 1 PUT request within 7 seconds.	The script should be able to perform 2 GET requests and 1 PUT request within 7 seconds.	
2	Write a script to authenticate 50 users at the same time.	The script should be able to authenticate 50 users at the same time.	
3	Merge the previous scripts.	The new script should be working as expected.	
4	Perform the availability test while keeping the elapsed time for each request, for each user.	A list of response times to the requests should be obtained.	
5	Check all 150 elapsed time results.	90% of requests should be handled in less than 5 seconds.	

Test ID	WS_R_2	Category	Non-Functional	Severity	Medium

Objective	To ensure that while the application is running, deleting a data instance will not harm the user experience.
Date	07.03.2024

#	Steps	Expected	Result
1	Set the scenario: Labeler is at the labeling interface and labeling a selected instance.	A session is created with the related scenario.	
2	Set the scenario: Publisher project analytics page and seeing a selected instance.	A session is created with the related scenario.	
3	Delete the selected data instance from the database manually.	The deleted data instance and everything related to it should be removed.	
4	Check the databases.	All data related to that instance should be deleted.	
5	For the scenario in step 1, try to save the labeling task.	The task must not be saved to the database and a related error should be shown. The labeling experience should continue as normal.	
6	For the scenario in step 2, refresh the page and search for that data instance.	Any information related to that data instance should not be seen on the analytics page or the settings page.	

Test ID	WS_R_3	Category	Non-Functional	Severity	Low

Objective	To ensure that while the application is running, deleting a project instance will not harm the user experience.
Date	07.03.2024

#	Steps	Expected	Result
1	Set the scenario: Labeler is at the labeling interface and labeling an instance of the selected project.	A session is created with the related scenario.	
2	Set the scenario: Publisher is looking at the project analytics page.	A session is created with the related scenario.	
3	Set the scenario: A labeler is looking at the project details page of the selected project and is included in the project.	A session is created with the related scenario.	
4	Set the scenario: A labeler is searching for the selected project from the dashboard.	A session is created with the related scenario.	
5	Delete the selected project from the database manually.	The deleted project and everything related to it should be removed.	
6	Check the databases.	All data related to that project should be deleted.	
7	For the scenario in step 1, try to save the labeling task.	The task must not be saved to the database and a related error should be shown. The labeling experience should be interrupted The project should not	

		be seen in the sidebar (of current projects).	
8	For the scenario in step 2, refresh the page.	Publisher should get a 404 error. The project should not be seen in the sidebar (of current projects).	
9	For the scenario in step 3, refresh the page.	Labeler should get a 404 error. The project should not be seen in the sidebar (of current projects).	
10	For the scenario in step 3, refresh the page.	Labeler should not be able to find the project after searching.	

Test ID	WS_R_4	Category	Non-Functional	Severity	Low

Objective	To ensure that while the application is running, after a sudden and heavy traffic for 2 minutes, the application should restore optimal performance in at most 3 minutes.
Date	07.03.2024

#	Steps	Expected	Result
1	Before the test, send multiple GET, ADD, PUT, DELETE requests and calculate each request's elapsed time.	Find the optimal performance in terms of time for the request handling.	
2	While the application is running, use a CPU loading test for 2 minutes while the application is running.	CPU usage should be around 80% to 100%.	
3	During the test, send multiple GET, ADD, PUT, DELETE requests and calculate each request's elapsed time.	Elapsed times for the requests on average should be greater than the results before the CPU test.	
4	After the test, keep sending multiple GET, ADD, PUT, DELETE requests and calculate each request's elapsed time for 5 minutes.	Elapsed times should get lower over time compared to the results during the CPU test. Results should be roughly the same after 3 minutes compared to the results before the CPU test.	

5. Consideration of Various Factors in Engineering Design

5.1 Constraints

5.1.1 Operational & Implementation Constraints

- AnnoHub will be a web application.
- GitHub will be used for team collaboration, code implementation, and version control.
- Python Flask will be used for the backend implementation, and React will be used for the frontend implementation. This was found suitable for agile development.
- MongoDB will be used for data annotation because a NoSQL database allows flexibility in handling frequently changing schemas. PostgreSQL will be used to store user information since it performs better when dealing with bigger datasets.
- A cloud service (such as Amazon Web Services) is planned to be used for hosting.
- Electron JS will be used for the on-premise desktop application implementation because it is interpreted, and its HTML-CSS-JS implementation is beneficial for agile development.
- Customers are expected to have decent servers to ensure data privacy while keeping the application functional.
- Customers and labelers should exist for AnnoHub's operational purposes.
- The data to be labeled does not have to be hosted in AnnoHub. However; in the event of customers choosing to keep their data private, certain features -such as mislabel prediction- are either going to be disabled or the source code will be shared with the customer through the on-premise application since both cannot co-exist.

5.1.2 Economic Constraints

- The main economic constraint for AnnoHub is the expenses associated with server maintenance. This includes costs related to hosting, cloud services, and storage. Server resources must be efficiently managed to control these expenses and ensure that AnnoHub is functioning smoothly. The storage costs are expected to be higher than the others and should be given importance.
- Another economic constraint is the costs related to financial transactions, like money transfers to labelers. These transactions involve fees and currency conversion charges, which need to be considered in AnnoHub's financial planning.
- Initial investment is another significant economic constraint. While the project can operate with a moderate initial investment, future financial planning is essential. Adequate funding is needed not only for the initial setup costs but also for the ad expenses and other ongoing needs.

5.1.3 Legal & Ethical Constraints

- A mandatory objective of AnnoHub is to stay observant of the Turkish Personal Data Protection Law (KVKK).
- User-provided information must not be leaked to any third-party organization unless explicitly approved by the user and is within KVKK's scope.
- The user's stored data should be destroyed within a time interval or when the user requests the data's destruction. The detailed policies on data storage and destruction must be provided to the user as stated in KVKK.
- For the data to be stored and processed, clarification and express consent texts must be given to the user. The methodologies used to store/process user data should be introduced to the user together with other legal and ethical considerations. The user must approve these texts with their consent before using the application.
- The data must be stored/processed exactly the way that was introduced to the user, which gets the user's approval.

5.2 Standards

5.2.1 Compliance with KVKK entails

- Registering with the Data Controllers' Registry (VERBIS) before processing any personal data.
- Obtaining explicit consent from data subjects for processing their personal data, except in specific circumstances outlined by the law.
- Ensuring the rights of data subjects are protected, including the right to be informed about data processing activities, the right to access their data, and the right to request correction or deletion of their data.
- Implementing measures for the secure transfer of personal data to foreign jurisdictions, which may involve ensuring adequate protection in the recipient country or obtaining authorization from the Turkish Data Protection Authority (DPA).
- Promptly notify the DPA and affected data subjects in the event of a data breach.

5.2.2 Data Security and Privacy Standards

- ISO/IEC 27001: This international standard outlines the specifications for an information security management system (ISMS). Compliance with ISO/IEC 27001 will help to safeguard sensitive client data as it flows through our system to labelers.
- General Data Protection Regulation (GDPR): Adherence to GDPR is critical for protecting the personal information of individuals within the European Union. Our system will implement necessary controls to manage user consent, data minimization, and individuals' rights over their data.

- Health Insurance Portability and Accountability Act (HIPAA): If dealing with health-related data, compliance with HIPAA will ensure the protection of personal health information, which could be part of the datasets being labeled.

5.2.3 Data Handling and Distribution Standards

- Data Minimization and Anonymization: Following best practices for data minimization and anonymization will ensure that labelers receive only the data necessary to perform their tasks without exposing sensitive information.
- End-to-end Encryption: The use of end-to-end encryption protocols for data transmission will ensure that client data is secure in transit, preventing unauthorized access.

5.2.4 Software and Web Development Standards

- OWASP Top 10: We will adhere to the Open Web Application Security Project's Top 10 list to protect against common web application security risks.
- W3C Web Standards: Our user interface will align with the World Wide Web Consortium (W3C) guidelines, including HTML5, CSS3, and Web Content Accessibility Guidelines (WCAG), to ensure compatibility and accessibility.

5.2.5 API and Data Interchange Standards

- RESTful API Security: Our APIs will follow REST security best practices, including authentication tokens, rate limiting, and proper session management.
- JSON Web Tokens (JWT): For secure transmission of information between parties, JWT will be used, allowing us to pass user and session data safely.

5.2.6 Database Design and Management Standards

- SQL:2016 / NoSQL Best Practices: We will follow the latest SQL standards or NoSQL best practices to ensure efficiency, reliability, and integrity in data storage and retrieval.
- ACID Properties: Transaction handling within our databases will conform to ACID properties to ensure that all data transactions are processed reliably.

5.2.7 Compliance with Industry-Specific Standards

- Payment Card Industry Data Security Standard (PCI DSS): If our platform handles transactions, we will comply with PCI DSS to secure cardholder data.

6. Teamwork Details

6.1 Contributing and functioning effectively on the team

Maintaining effective teamwork is one of our main concerns in the project's progress, for which we follow the principles below.

- Every member should attend the regular meetings.
- Every member should contribute to the decision-making process.
- Every member should take part in each step in the project's progress.
- Every member is responsible for doing the necessary research about their work.
- Every member should meet the deadlines that are determined in the regular meetings.
- The distribution of work in each step in the project's progress should be fair.

6.2 Helping create a collaborative and inclusive environment

In order to create a collaborative and inclusive environment, we utilize various tools and platforms to facilitate collaboration, such as Jira for task management and tracking progress. Regular meetings are conducted on Discord, providing opportunities for team members to share updates, discuss challenges, and brainstorm solutions collaboratively. Additionally, we encourage active participation and feedback from all team members, ensuring that everyone's perspectives and contributions are valued and considered.

6.3 Taking a lead role and sharing leadership on the team

Taking a lead role and sharing leadership on the team involves empowering team members to take ownership of tasks and initiatives while also keeping a collaborative leadership approach. Leadership responsibilities are distributed based on individual strengths and expertise, allowing team members to lead tasks aligned with their skills and interests.

7. References

[1] “Machine learning market size, share & growth report, 2030,” Machine Learning Market Size, Share & Growth Report, 2030, https://www.grandviewresearch.com/industry-analysis/machine-learning-market#:~:text=Report%20Overview,how%20businesses%20and%20people%20operate (accessed Nov. 16, 2023).

[2] “Access, storage, and security,” Labelbox docs, https://docs.labelbox.com/docs/access-storage (accessed Nov. 16, 2023).