# Project Software Engineering (DLMCSPSE01\_CF)

The goal of this project is to develop an AI based nutritional assistant. This assistant should help users to improve their diet and overall health.

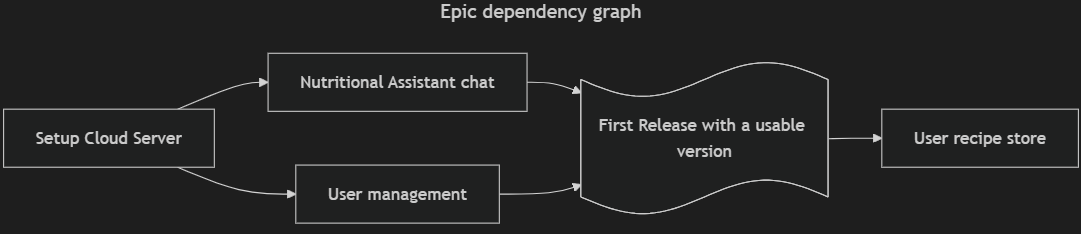
The assistant targets people how struggle with healthy diets because they lack the right information or people already have the basic knowledge but need special advise in their daily life. Since this is a digital offer, the early adopters are digitally savvy and probably young people.

## Project plan

The first milestone the app should only support basic chat functionalities between the user and the digital assistant. To demonstrate a first more advanced feature, the first assistant tool will be implemented. The first tool will allow the assistant to store recipes in the database. With this the first milestone will be implemented with those four epics and the scope of this assignment.



The following chart shows the dependencies of the epics to each other. It also shows the first usable version, that brings the first benefits to the user.



## Risks

As every software project also this projects has to mitigate certain risks which can be grouped in business related risks, technological risks and security risks.

### Business Risks

###### Data Privacy Risk

The solution experimented in this project is a 100% digital offer and the privacy concerns with those might scare off users. Especially because the assistant needs personal information about the user, to be able to support them with personalized advise.

**Mitigation:**

As a mitigation, data privacy should be a top priority. Users should be aware of every personal data, that is shared with the user. The User should also be able to view all data which the app collects and should be able to adapt / delete those.

In general the App should emphasize the importance of user privacy by providing good user assistance and the right data privacy features.

###### Product Market Risk

The problem of a healthy diet is huge and so are the solutions for it. It could be that the developed app is not a good solutions to the users dietary problems. This would lead to a high user churn.

**Mitigation:**

The first mitigation would be a user survey, but this will not be done as part of this assignment. Instead the first implementation will rely on the Problems of the Project member. After the first MVP is implemented, the app can be iterated based on customer feedback. This way the app can archi9ve product market fit over time.

###### AI Legal Risk

As outlined before, the goal is to use an AI to answer the questions to the user. The output of those generative models can only be controlled in a limited fashion. The risk is, that the AI provides some negative advice to a user.

**Mitigation:**

Implement clear user consent mechanism and make transparent, how generative AI works and that the recommendation’s must be considered carefully and that those are no medical advices. In addition the context definition of the assistant itself should contain guidelines to the AI, which recommendations are ok and what the AI should not do. In addition broadly adapted models should be used which are already shipped / trained with guidelines and content restrictions.

###### AI Wrapper Risk

An app, based on a GPT model enables problem solving of complex problems in a limited amount of time. The risk is that at some point a generic solution also based on GPT (e.g. Chat-GPT) can do all this on their one.

**Mitigation:**

Always provide additional features in the specific Problem space, that other solutions are not supporting. In the first milestone the more Problem focused App visualization and the feature to store generated recipes for later use are two differentiators. In addition the Assistant is already configured with an optimized Prompt, which a normal User will maybe not be able to write.

### Technical Risks

### Feature regression Risk

The software solution will be developed in multiple iterations and with small feature enhancements. It is important to prevent feature regressions, where some functionality is no-longer working as expected.

**Mitigation:**

Write automated tests and use CI/CD pipelines to deploy changes only when the tests are passed. The tests should be separated in low level unit tests covering critical functionalities (70% line coverage) and system tests, which test the backend service and backend client as a whole.

### GPT pricing risk

The app heavily depends on GPT models which already have a certain price. If a model increases the price significantly, this impacts the operational cost of the application.

**Mitigation:**

In Milestone one only the GPT model of OpenAI will be connected. It must be connected using an Adapter, that will allow an easy switch to a different model at a later point in time. That way the APP could switch the Model with low effort, if the used model changes the price or another model gets cheaper or leads to better results.

### Security risks

Since the app is handling critical user data, security plays an important role. All risks must be prioritized according to priority.

#### Unauthorized Access to Sensitive Health Data

If the User Data is not properly secured, attackers can get unauthorized access to the sensitive personal and health-related information. This can result from database misconfigurations, weak encryption or inadequate access controls.

|  |  |  |  |
| --- | --- | --- | --- |
| **Thread Event** | **Likelihood** | **Impact** | **Risk Level** |
| **Loss of Confidentiality** | **Possible** | **Significant** | **High** |
| **Loss of Integrity** | **Possible** | **Moderate** | **Moderate** |
| **Loss of Availability** | **Possible** | **Significant** | **High** |
|  |  | **Overall Risk** | **High** |

**Mitigations:**

1. Encrypt Data at rest and in transit
2. Strict access controls
3. Secure storage using zero trust model (avoid database exposure)
4. Apply rate limiting

#### Data leakage from AI Models

If the Users Health Data is shared with the AI, this can imply critical data protection issues. Personal health data is very sensitive and special mechanisms must be applied.

|  |  |  |  |
| --- | --- | --- | --- |
| **Thread Event** | **Likelihood** | **Impact** | **Risk Level** |
| **Loss of Confidentiality** | **Likely** | **Moderate** | **Moderate** |
| **Loss of Integrity** | **Unlikely** | **Moderate** | **Moderate** |
| **Loss of Availability** | **Very Unlikely** | **Negligible** | **Low** |
|  |  | **Overall Risk** | **Moderate** |

**Mitigations:**

1. Do not share sensitive health data with the assistant
2. Inform the user, that no personal health data should be shared
3. Instruct the AI model to not ask for personal health information
4. Use model with Health Insurance Portability and Accountability Act (HIPAA) support
   1. Might be important for later features
   2. Those models also allow zero retention policies

#### Weak or compromised Authentication

Weak authentication mechanisms or poor session management could allow unauthorized access.

|  |  |  |  |
| --- | --- | --- | --- |
| **Thread Event** | **Likelihood** | **Impact** | **Risk Level** |
| **Loss of Confidentiality** | **Possible** | **Significant** | **High** |
| **Loss of Integrity** | **Unlikely** | **Moderate** | **Moderate** |
| **Loss of Availability** | **Possible** | **Moderate** | **Moderate** |
|  |  | **Overall Risk** | **High** |

**Mitigations:**

1. Strict password policy
2. Support multifactor authentication
3. Use short sessions and use refresh tokens for session security
4. Secure App access with biometric encryption

#### API Abuse and Unauthorized Data Access

Attacker can use not properly secured APIs to exploit them and to extract user data or perform unintended actions

|  |  |  |  |
| --- | --- | --- | --- |
| **Thread Event** | **Likelihood** | **Impact** | **Risk Level** |
| **Loss of Confidentiality** | **Possible** | **Moderate** | **Moderate** |
| **Loss of Integrity** | **Unlikely** | **Minor** | **Low** |
| **Loss of Availability** | **Possible** | **Moderate** | **Moderate** |
|  |  | **Overall Risk** | **Moderate** |

**Mitigations:**

1. All APIs secured using JWT tokens
2. Rate limiting and request throttling
3. Use a web application firewall (WAF) to filter and block malicious requests
4. Strong input validation to prevent injection attacks
5. Use secure TLS versions
6. Auditlog suspicious activity

#### Insecure Communication Between Components

The communication between components uses Protobuf and HTTP protocols. If these communications are not secured, attackers can intercept and or manipulate data.

**Ease of Attack**: **Medium**

|  |  |  |  |
| --- | --- | --- | --- |
| **Thread Event** | **Likelihood** | **Impact** | **Risk Level** |
| **Loss of Confidentiality** | **Possible** | **Significant** | **Moderate** |
| **Loss of Integrity** | **Possible** | **Moderate** | **Moderate** |
| **Loss of Availability** | **Possible** | **Moderate** | **Moderate** |
|  |  | **Overall Risk** | **Moderate** |

**Mitigations:**

1. Use TLS 1.3 where possible and do not allow versions lower than 1.2
2. Implement mTLS between system components
3. Apply HSTS headers to prevent protocol downgrade attacks

## Software development methodology

This project will use the Kanban software development methodology. Kanban is very lightweight which makes it suitable for this small one person project. It does not require complex planning and does not force some cycles, but still allows to be flexible and to change priorities.

### Kanban Board

The core focus in Kanban is the Kanban-Board. This project will be managed via Github-Projects and the Kanban board has the six columns Backlog, Refined, Ready, In progress, In review and Done.

The Backlog contains a list of all tasks for a bigger timeframe. It has a very big limit of 20, because in this area all items are collected for the future. Items in the Backlog are only roughly described and are not estimated and not prioritized.

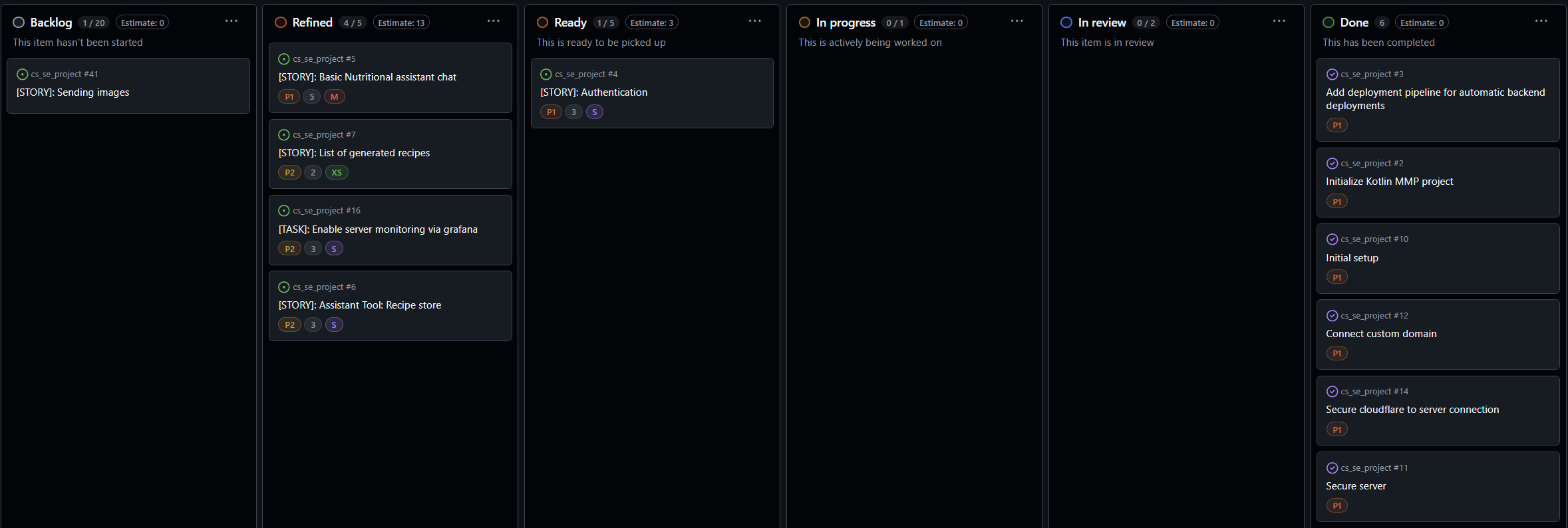
Items which are well described and which will be tackled in the next weeks, are in the Refined column. The column is limited to five prioritized and estimated items, which are ready for development as soon as all dependent requirements are fulfilled.

Once the refined items are completely unblocked, they can move in the ready column. These items could be started at any point in time. This column is also limited to five items and combined with the five refined items, they show the next up to ten items that will be done.

Started items are moved to in progress. Since this is a one person development project, the progress column is restricted to one item. With the multi-tasking will be prevented.

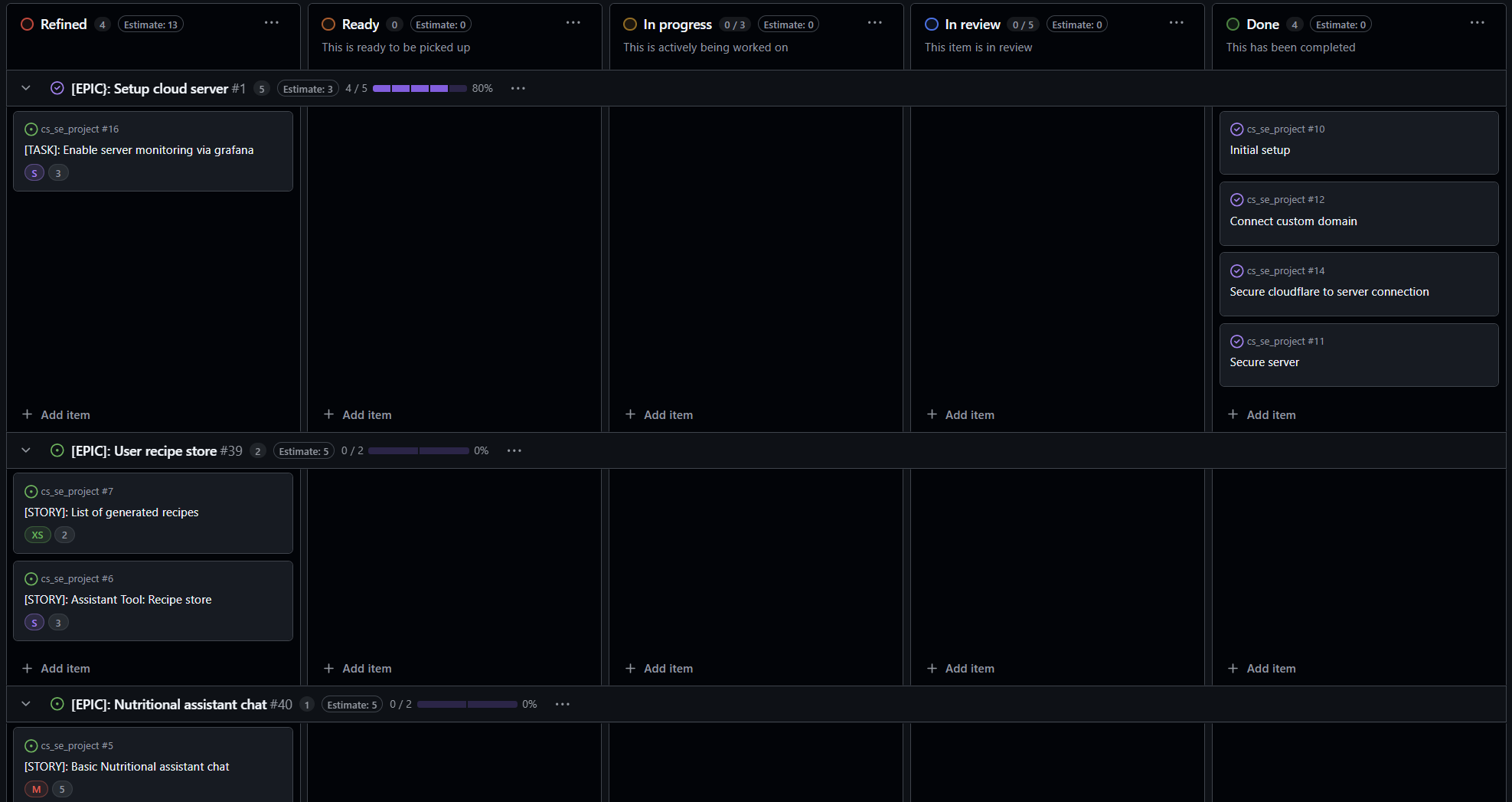
Once a PR is opened, items are automatically moved to In review. There are no second eyes, that could review the code, but there are automated checks, which prevent security breaches and software regressions.

After a PR is merged, items are moved to Done. The Done column has no limit configured.



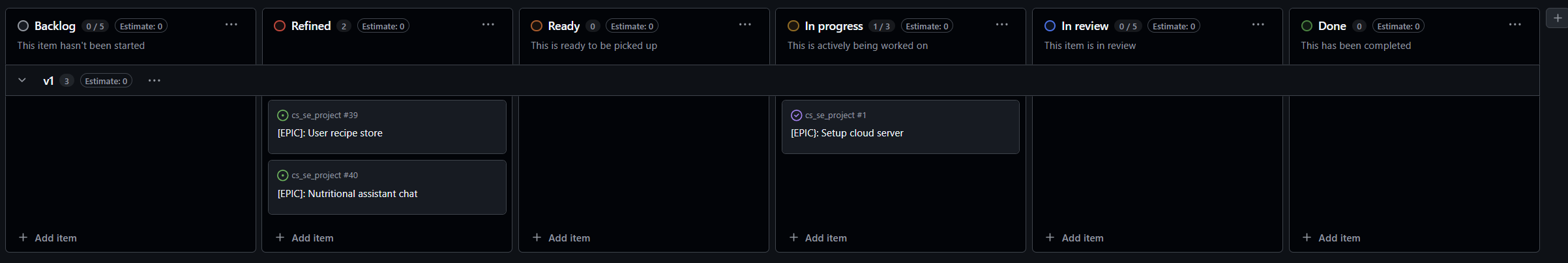
### Epic Board

The epic board shows a high level overview, how far the epics have progressed and which tasks/stories there are left to complete an epic.



### Milestone View

The milestone view show the status of the overall milestone and which epics are open until the milestone is completed.



### Roles

The Kanban project requires multiple roles which will all be filled with the on person in the exceptional project.

The first role is the Product Owner, who is responsible for creating the backlog and maintaining priorities based on user requirements. The Epics and User stories which the product owner defines, are further refined by the development team. This team consist of architects, developers and DevOps Engineers. The Architect is responsible for the system design and application architecture. The DevOps-Engineers create a continuous integration and continuous development pipeline. The pipeline ensures that new features are release often and only if the quality stages are passed.

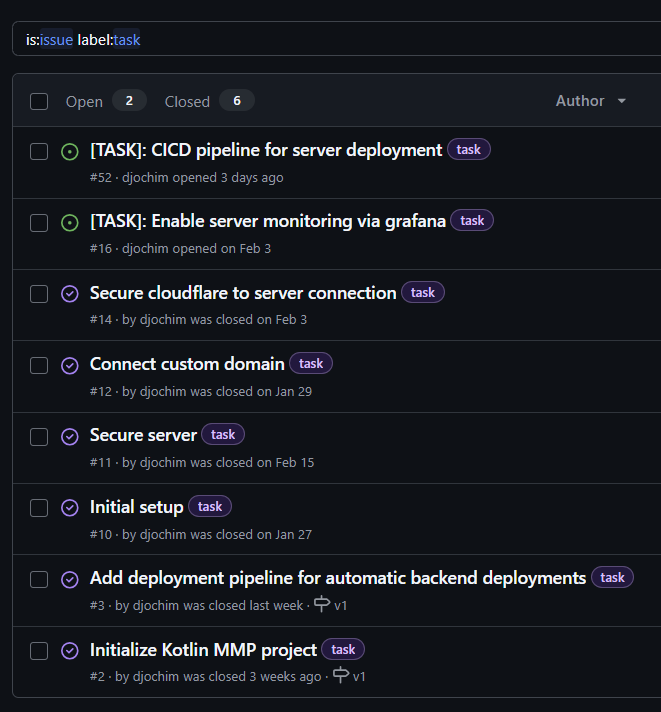
Developers are responsible for the implementation of the tasks and stories. They are also responsible to write good/automated tests that ensure that new releases don’t introduce regressions an that the manual testing effort is low.

## Requirements

The requirements can be grouped into non-functional and functional requirements. The non-functional requirements are required to deliver a good software solutions with high quality.

### Non-functional

In alignment with a lean development process, a continuous integration and deployment pipeline is required for a continuously flow of new items to the user. A backend server must be stetted up as a proxy to the OpenAI assistant API, but also to implement additional functionalities. Technical functionalities are maintained as tasks.

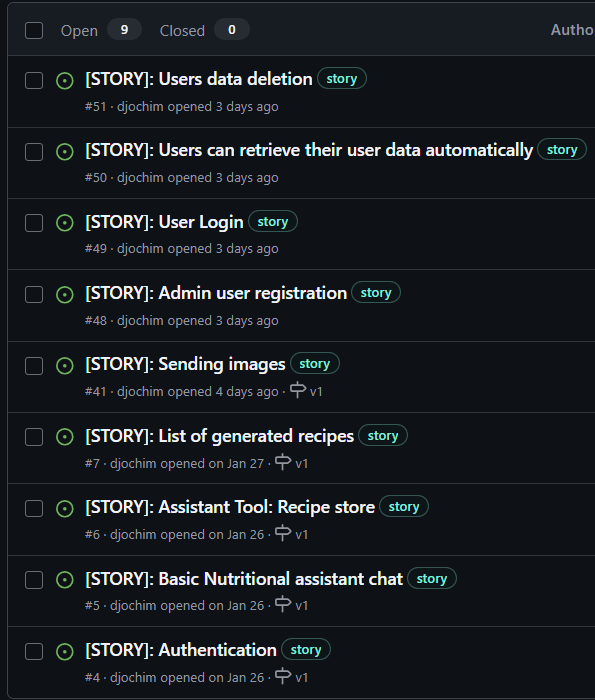


### Functional

For the first milestone the health coach has a very limited scope. First of all, the basic communication will be enabled between the Assistant and the user. This communication is based on text messages and the user can also send images.

In addition there will be a first toll implementation that allows the user to store recipes that are generated or extracted from text, images or websites.

Feature enhancements are tracked via User Stories.



## System Design

The application will be develop as Kotlin Multi-Platform project. There the server, Android App and iOS APP is develop with Kotlin and all three are also able to share some of the code parts with each other. The backend server uses Ktor as a server framework and connects to a MongoDB as storage.

The server will be deployed on a Hetzner Cloud Server as Docker container. For enhanced security the server will be behind a Ngix proxy and the whole Hetzner server will only be reachable over Cloudflare.

