

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет інформаційних технологій
Кафедра технологій управління

Спеціальність 122 «Комп'ютерні науки»

Освітньо-наукова програма «Управління проєктами»

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА на тему:

«Дослідження процесів управління проєктом з розробки інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб і таргетів»

Студентка 2-го курсу групи УП-21м

Вікторія ЗАРЕМБА

(підпис студента)

Науковий керівник:

кандидат технічних наук, професор

Віктор МОРОЗОВ

(дата)

(підпис)

Попередній захист:

(Висновок: "До захисту в Екзаменаційній комісії")

Завідувач кафедри
технологій управління

(підпис)

(прізвище, ініціали)

(дата)

Київ - 2024

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій**

Кафедра технологій управління
Освітній рівень Магістр
Спеціальність 122 Комп'ютерні науки
Освітня програма Управління проєктами

ЗАТВЕРДЖУЮ
Завідувач кафедри
професор Морозов В.В.

“3” жовтня 2023 року

**ЗАВДАННЯ
НА ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ**

Студентка Заремба Вікторія Валеріївна
Група УП-21

1. Тема кваліфікаційної роботи “Дослідження процесів управління проєктом з розробки інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб і таргетів”

Затверджена на засіданні кафедри технологій управління, протокол №5 від “3” жовтня 2023 р.

2. Строк подання студентом готової роботи - “13” травня 2024 р.

3. Цільова установка та вихідні дані до роботи: дослідження характеристики об'єкта управління, планування виконання проєкту (календарне планування, зміст, бюджет, ресурси, ризики, зміни) та планування управління іншими галузями управління проєктами

4. Зміст роботи: огляд методологій управління проєктами з розробки штучного інтелекту, аналіз фактичного стану досліджуваної проблеми, формулювання проблемної області, формулювання наукової новизни та інноваційності проєкту, проведення аналізу конкурентів, обґрунтування інвестиційної доцільності проєкту, постановка задачі дослідження, моделі та

методи управління проєктами з розробки штучного інтелекту, розробка моделі та моделювання ієрархічного опису проєкту, декомпозиція робіт проєкту, розрахунок моделі та моделювання розкладу робіт проєкту, розрахунок моделі та моделювання розподілу ресурсів та формування бюджету проєкту, розрахунок моделі та моделювання вартості ІТ проєкту, методи прийняття управлінських рішень в проєкті, розробка організаційної структури управління проєктом, формування команди проєкту, розробка матриці відповідальності, управління зацікавленими сторонами, розробка календарного плану, визначення та планування ресурсів, управління ризиками проєкту, використання платформи Jira для управління проєктом, розробка архітектурного рішення, організація роботи з великими даними для генерації молекул, проєктування та реалізація схеми бази даних, розробка інтерфейсів програмного забезпечення, розробка програмного забезпечення реалізації Веб частини ІТ проєкту, планування та проведення експериментів для III частини проєкту.

5. Перелік графічного матеріалу (слайдів): титульна сторінка, мета дипломної роботи, формулювання проблемної області, цілі та продукт проєкту, ієрархічна структура робіт, календарне планування, формування бюджету проєкту, організаційна структура, управління зацікавленими сторонами, управління ризиками, планування проєкту в Jira, розробка інформаційного та програмного забезпечення проєкту, висновки.

6. Календарний план виконання роботи:

№ з/п	Назва частини роботи	План виконання роботи
1	Вибір теми кваліфікаційної роботи магістра (КРМ)	До 15.11.23
2	Підготовка вступу	До 26.12.23
3	Підготовка розділу 1	08.01.24 - 18.02.24

4	Підготовка розділу 2	18.02.2 - 30.03.24
5	Підготовка розділу 3	01.04.24 - 20.04.24
6	Остаточне оформлення кваліфікаційної роботи	20.04.24 - 30.04.24
7	Передача КРМ в електронному вигляді на кафедру на перевірку роботи на плагіат	05.05.24 - 10.05.24
8	Передача КРМ (друк.) на рецензію керівнику	До 05.05.24
9	Презентація кваліфікаційної роботи магістра. Попередній захист роботи на кафедрі	10.05.24 – 11.05.24
10	Передача КРМ (друк.) на рецензію	21.05.24
11	Передача видрукованої та переплетеної роботи на кафедру	21.05.24
12	Захист магістерської кваліфікаційної роботи	27.05.24 - 29.05.24

Дата видачі завдання: “3” жовтня 2023 р.

Керівник роботи: кандидат техн. наук, професор Віктор МОРОЗОВ

(підпис)

Завдання прийняла до виконання:

студентка групи УП-21 Вікторія ЗАРЕМБА

(підпис)

Зміст

Зміст	5
АНОТАЦІЯ	7
Таблиця скорочень та пояснень	9
Вступ.....	10
РОЗДІЛ 1. ДОСЛІДЖЕННЯ ТА ОБҐРУНТУВАННЯ ДОЦІЛЬНОСТІ ТА ЖИТТЄЗДАТНОСТІ ПРОЄКТУ	14
1.1. Огляд методологій управління проєктами з розробки штучного інтелекту	14
1.2. Аналіз фактичного стану досліджуваної проблеми. Формулювання проблемної області.....	17
1.3. Формулювання наукової новизни та інноваційності проєкту.....	17
1.4. Проведення аналізу конкурентів.....	20
1.5. Обґрунтування інвестиційної доцільності проєкту.....	26
1.6. Постановка задачі дослідження, формулювання технічного завдання на розробку у вигляді паспорта проєкту	28
РОЗДІЛ 2. МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ.....	40
2.1. Розробка моделі та моделювання ієрархічного опису проєкту. Декомпозиція робіт проєкту.....	40
2.2. Розрахунок моделі та моделювання розкладу робіт проєкту.....	41
2.3. Розрахунок моделі та моделювання розподілу ресурсів	49
2.4. Розрахунок моделі та моделювання вартості ІТ проєкту	51
РОЗДІЛ 3. ПЛАНУВАННЯ ЕЛЕМЕНТІВ УПРАВЛІННЯ ПРОЄКТОМ.....	55
3.1. Розробка організаційної структури управління проєктом. Формування команди проєкту	55
3.2. Розробка матриці відповідальності.....	57

3.3. Управління зацікавленими сторонами	60
3.4. Створення комунікаційного плану проєкту.....	66
3.5. Управління ризиками проєкту.....	70
3.6. Використання платформи Jira для управління проєктом	75
РОЗДІЛ 4. РОЗРОБКА ІНФОРМАЦІЙНОГО ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПРОЄКТУ	78
4.1. Розробка архітектурного рішення.....	78
4.2. Організація роботи з великими даними для генерації молекул.....	79
4.3. Розробка інтерфейсів програмного забезпечення	82
4.4. Розробка програмного забезпечення реалізації Веб частини ІТ проєкту ..	85
4.5. Проєктування схеми бази даних	88
4.6. Планування та проведення експериментів для III частини проєкту	91
ВИСНОВКИ	97
ПЕРЕЛІК ВИКОРИСТАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ.....	99
ДОДАТОК А	102
ДОДАТОК Б.....	103
ДОДАТОК В	104
ДОДАТОК Г	106
ДОДАТОК І	107
ДОДАТОК Д.....	109

АНОТАЦІЯ

кваліфікаційної роботи магістра на тему

«Дослідження процесів управління проектом з розробки інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб і таргетів»

Студентка: Заремба Вікторія Валеріївна

Науковий керівник: Морозов Віктор Володимирович

Рік захисту - 2024

Тема кваліфікаційної роботи - “Дослідження процесів управління проектом з розробки інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб і таргетів”. Метою дослідження є планування та розробка ефективних процесів управління проектом з розробки інформаційної платформи з штучним інтелектом, спрямованих на дослідження ліків, хвороб і таргетів. Об’єкт дослідження — розробка інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб та таргетів. Предметом дослідження є процеси управління проектами з розробки штучного інтелекту, які впливають на ефективність розробки інформаційної платформи для дослідження ліків, хвороб і таргетів. Зокрема, це включає управління вартістю, часом, ризиками, якістю, змінами, а також механізми формування та координації роботи мультидисциплінарних команд, що беруть участь у розробці.

В рамках роботи було розглянуто існуючі підходи до управління проектами AI, проведено всебічний аналіз ринку для визначення потенціалу та викликів в галузі, та розроблено комплексний план проекту, включаючи його життєвий цикл та структуру робіт. Також змодельовано розклад робіт та формування бюджету проекту. Окрема робота проведена з розробкою організаційної структури, матриці відповідальності, управлінням зацікавленими сторонами, створенням комунікаційного плану, управління ризиками та використанням платформи Jira для управління проектом.

Перший розділ дипломної роботи висвітлює основні методології управління проєктами з розробки штучного інтелекту, аналізує фактичний стан досліджуваної проблеми, визначає наукову новизну та інноваційність проєкту, а також проводить аналіз конкурентів та обґрунтовує інвестиційну доцільність.

Другий розділ містить розробку математичної моделі та постановку задачі дослідження, декомпозицію робіт проєкту, розрахунок розкладу робіт, розподіл ресурсів та визначення вартості ІТ-проєкту.

У третьому розділі охоплено розробку організаційної структури управління проєктом, формування команди, матриці відповідальності, управління зацікавленими сторонами, створення комунікаційного плану та управління ризиками.

Розділ чотири описує розробку архітектурного рішення, організацію роботи з великими даними для генерації молекул, розробку інтерфейсів програмного забезпечення, проєктування бази даних та реалізацію веб-частини ІТ-проєкту.

Практична частина включає розробку програмного забезпечення проєкту, реалізації Веб-інтерфейсу, проєктування бази даних, а також планування та проведення експериментів для модулів штучного інтелекту з використанням методології DDS.

Результати роботи були успішно впроваджені на підприємстві blackthorn.ai і продемонстровані на міжнародних наукових конференціях, що свідчить про їхню актуальність та практичну значущість.

Робота містить 99 сторінок без додатків, 23 рисунки, 9 таблиць. Додатки розміщено на 7 сторінках.

Ключові слова: *управління проєктами, проєктний менеджер, штучний інтелект, управління науковими проєктами, розробка ліків, інформаційна платформа*

Таблиця скорочень та пояснень

Скорочення	Пояснення
ШІ	Штучний Інтелект
AI	Artificial Intelligence
ML	Машинне навчання
CRISP-DM	The Cross Industry Standard Process for Data Mining
SMILES	Simplified Molecular Input Line Entry Specification
ALS	Amyotrophic Lateral Sclerosis
ADME	абсорбція, дистрибуція, метаболізм, екскреція
QSAR	квантитативні відносини структури і активності
РНК	Рибонуклеїнова кислота
ICD-11	International Classification of Diseases 11th Revision
FDA	Food and Drug Administration
DE	Data Engineering
ПРЗ	Проти ризикові заходи
DDS	Data Driven Scrum – Скрам на основі даних
MVP	(minimum viable product) мінімально життєздатний продукт
MMP	(Minimal marketable product) мінімальний набір можливостей продукту, за який люди готові платити.

Вступ

У світі існує понад 7,000 захворювань без ефективного лікування, а майже чверть світового населення не має доступу до необхідних ліків, зокрема через недостатнє фінансування досліджень інфекційних захворювань та їхню низьку привабливість для фармацевтичної індустрії. Разом із цим вартість розробки нових ліків щороку збільшується, а прибуток від них зменшується. Це ставить перед світовою індустрією охорони здоров'я та фармацевтики нові виклики: як зробити доступ до лікування рівним та доступним для всіх, як знизити час та вартість розробки нових ліків, як інтегрувати науково-технологічний прогрес в класичні підходи до відкриття та розробки лікарських засобів.

Обрана тема є вкрай актуальною, враховуючи постійно зростаючу потребу у нових лікарських препаратах для лікування різноманітних захворювань, особливо тих, що наразі вважаються невиліковними. Розвиток мультиомічних платформ з інтеграцією штучного інтелекту відкриває нові горизонти у цьому дослідженні, пропонуючи більш ефективні та швидкі шляхи до відкриття та розробки медичних препаратів.

Метою дослідження є планування та розробка ефективних процесів управління проектом з розробки мультиомічної платформи з штучним інтелектом, спрямованих на дослідження ліків, хвороб і таргетів.

Об'єктом є розробка інформаційної платформи з штучним інтелектом, спрямованої на дослідження ліків для різних захворювань та їхніх молекулярних мішеней (таргетів).

Предметом дослідження є процеси управління проектами з розробки штучного інтелекту, які впливають на ефективність розробки інформаційної платформи для дослідження ліків, хвороб і таргетів. Зокрема, це включає управління вартістю, часом, ризиками, якістю, змінами, а також механізми формування та координації роботи мультидисциплінарних команд, що беруть участь у розробці.

Основними завданнями даної роботи є:

- Дослідження підходів управління проєктами з розробки штучного інтелекту
- Проведення аналізу предметної області: Детальне вивчення сучасних підходів, технологій та викликів у розробці платформи з штучним інтелектом для відкриття нових ліків.
- Проведення аналізу конкурентів та інвестиційних досліджень: Аналіз ринку, потенційних інвестиційних можливостей, ризиків та вимог до фінансування проєкту.
- Створення паспорту проєкту: Опис продукту, його завдань та цілей: Детальний опис мультиомічної платформи, включно з її ключовими функціями, завданнями, які вона вирішує, та цілями, на досягнення яких вона спрямована.
- Розробка життєвого циклу проєкту та декомпозиція робіт проєкту: Планування етапів розробки платформи та декомпозиція на окремі завдання та підзавдання.
- Моделювання розкладу робіт проєкту: Розробка часового плану виконання завдань та досягнення цілей проєкту.
- Проведення планування ресурсів та формування бюджету: Визначення необхідних ресурсів для виконання проєкту, включаючи людські, технічні та фінансові. Розробка бюджету проєкту.
- Розрахунок вартості проєкту: Оцінка загальних витрат на реалізацію проєкту.
- Побудова організаційної структури проєкту: Формування команди проєкту, визначення ролей та відповідальностей учасників.
- Розробка плану взаємодії з зацікавленими сторонами: Встановлення комунікації з усіма зацікавленими сторонами та забезпечення їх залученості у проєкт.
- Визначення ризиків та стратегії їх управління: Аналіз потенційних ризиків проєкту та розробка плану їх мінімізації або усунення.
- Розробка інформаційного та програмного забезпечення проєкту

- Моніторинг виконання проєкту: Використання показників методу освоєного обсягу для оцінки прогресу проєкту та своєчасного реагування на зміни.

У магістерській роботі використовується низка методів дослідження, кожен з яких обрано для вирішення конкретних аспектів дослідження. Аналітичний метод допоміг у систематизації та оцінці існуючих підходів до управління проєктами розробки штучного інтелекту та аналізу предметної області. Це дозволило ідентифікувати основні виклики та визначити потреби у вдосконаленні управлінських процесів.

Метод моделювання був застосований для створення імітаційних моделей процесів управління проєктами. Використання цього методу сприяло оцінці впливу різних управлінських рішень на ефективність розробки платформи. Це дало можливість адаптувати управлінські стратегії до специфіки проєктів у цій області.

Наукова новизна отриманих результатів полягає у запропонованій моделі управління проєктів з штучним інтелектом, а також у впровадженні теорії ігор в процес управління стейкхолдерами.

Розроблена платформа має *практичне значення* для фармацевтичних компаній та науково-дослідних інститутів, сприяючи прискоренню процесів дослідження нових лікарських засобів та зниженню витрат.

Результати досліджень на тему кваліфікаційної роботи було впровадження на підприємстві blackthorn.ai. Також в межах кваліфікаційної роботи було розроблено підхід до управління стейкхолдерами з застосуванням теорії ігор. Цей підхід апробовано:

- 1) 2022-2023, BLACKTHORN AI LTD, В компанії впроваджено оптимізований процес управління зацікавленими з застосуванням теорії ігор (Додаток А);

- 2) Результати дослідження доповідались на міжнародній конференції “Information Technology and Implementation” (IT & I-2023)[1].

3) Результати дослідження оформлені у вигляді статті на міжнародну конференцію 2024 IEEE International Conference on Smart Information Systems and Technologies Conference. Стаття пройшла рецензування і зараз на етапі формування збірника та його публікування

РОЗДІЛ 1. ДОСЛІДЖЕННЯ ТА ОБҐРУНТУВАННЯ ДОЦІЛЬНОСТІ ТА ЖИТТЄЗДАТНОСТІ ПРОЄКТУ

1.1. Огляд методологій управління проєктами з розробки штучного інтелекту

Управління науковими проєктами та проєктами з розробки штучного інтелекту вимагає спеціалізованих підходів, що враховують унікальний характер таких проєктів.

Управління проєктами машинного навчання є складним через непередбачуваність і високу складність завдань, а також через особливості тренувальних даних та їхнє застосування у продакшні. Проєкти часто зіштовхуються з викликами, як-от невизначеність ефективності алгоритмів на нових даних та потреба у великих обсягах релевантних тренувальних даних[2].

1.1.1. Застосування традиційних методологій

Традиційні методології, такі як Waterfall і Critical Path Method (CPM), довгий час застосовувалися у наукових дослідженнях та проєктах з чітко визначеними фазами та вимогами. Водоспадна методологія, з її послідовною моделлю розробки, дозволяє ретельно планувати та виконувати проєктні завдання. Але він є недостатньо гнучким до змін, що є критичним для проєктів розробки штучного інтелекту. Також, попри можливість запланувати виконання конкретних експериментів та досліджень, такий підхід не обіцяє отримати задовільний результат для замовника, що є критичним для успішності комерційних проєктів.

1.1.2. Застосування гнучких методологій

Методологія Scrum, особливо корисна для проєктів зі штучним інтелектом, забезпечує гнучке та адаптивне управління завдяки коротким робочим циклам - спринтам. Це дозволяє командам швидко реагувати на зміни, сприяє прозорості та відповідальності, що є ключовими для успішного керування проєктами ШІ [3].

Для адаптації Scrum під проекти зі штучного інтелекту важливо модифікувати деякі елементи цієї методології, включаючи ролі в команді, а також артефакти та події Scrum. Залучення спеціалістів з даних до складу команди, створення окремого переліку даних поряд з продуктовим беклогом, а також адаптація оглядів спринтів та ретроспектив можуть сприяти успішному управлінню проектами ШІ. Scrum може підвищити ефективність спілкування, прискорити доставку цінності, зменшити ризики та витрати, а також сприяти навчанню та адаптації проекту. Однак, серед викликів можуть бути управління залежностями даних, балансування між дослідженням та використанням, а також вимірювання цінності результатів ШІ. Для подолання цих викликів рекомендується планувати та пріоритизувати діяльність, пов'язану з даними, слідувати етичним принципам та визначати метрики цінності [4].

1.1.3. Специфічні методології для проектів розробки ШІ

Розробка ШІ представляє унікальні виклики, такі як висока непередбачуваність результатів, потреба у швидкому прототипуванні та ітерації, а також особливі вимоги до якості та надійності даних.

1.1.3.1. Методологія CRISP-DM

Методологія CRISP-DM [5, 6] є стандартним процесом для проектів з видобування даних та аналітики даних, включаючи шість послідовних фаз: *розуміння бізнесу, розуміння даних, підготовка даних, моделювання, оцінка та впровадження*. Цей процес допомагає командам ефективно організувати роботу з даними, від визначення бізнес-цілей до розгортання моделей. Він сприяє глибокому аналізу даних, правильному вибору та оцінці моделей, а також забезпечує, що результати відповідають бізнес-вимогам і є доступними для зацікавлених сторін.

CRISP-DM може бути розглянуто як гнучка та адаптивна методологія або як жорстка, схожа на Waterfall, залежно від способу її застосування. Якщо дотримуватися CRISP-DM дуже строго, плануючи детально кожен етап на

початку проєкту і не здійснюючи частих ітерацій, це нагадуватиме Waterfall-методологію. Водночас CRISP-DM може підтримувати принципи Agile, якщо дозволяти гнучкість в переходах між фазами і швидко ітерувати, інтегруючи при цьому інші гнучкі процеси. Важливим є те, як команда вибирає впроваджувати цю методологію в контексті свого проєкту. Діаграма методології CRISP-DM зображена на рисунку 1.1.

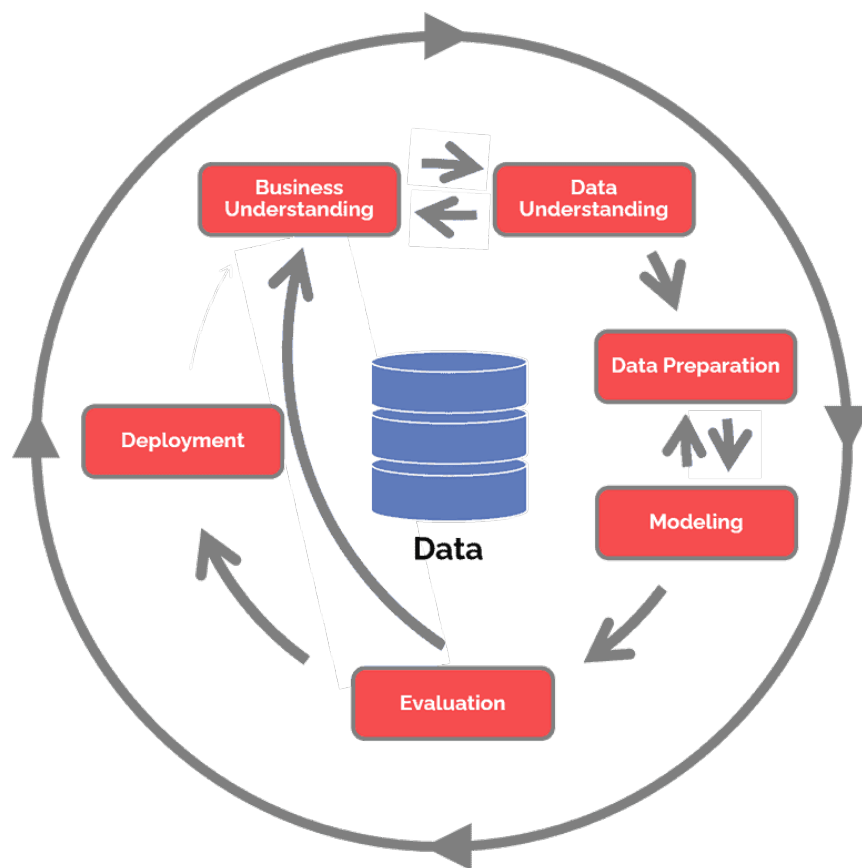


Рис.1.1. Діаграма методології CRISP-DM [5]

1.1.3.2. Data Driven Scrum

Data-Driven Scrum (DDS) [6] відрізняється від традиційного Scrum тим, що DDS адаптований під проєкти з обробки даних, де ітерації можуть мати різну тривалість. Це дозволяє командам ефективніше вирішувати завдання, які складно оцінити за часом. Замість точних оцінок завдань, DDS застосовує приблизні оцінки для визначення пріоритетів завдань. Також, у DDS

колективний аналіз інтегрований у основний робочий процес, що допомагає команді приймати обґрунтовані рішення, засновані на даних. Наради в DDS можуть проводитися незалежно від ітерацій, а перекривання ітерацій є звичайною практикою, особливо коли задачі потребують часу для збору даних.

1.2. Аналіз фактичного стану досліджуваної проблеми. Формулювання проблемної області

Традиційний процес відкриття ліків включає ідентифікацію таргетів для ліків, проєктування та скринінг молекул ліків, а також проведення клінічних випробувань, що є складним і тривалим [7, 8, 9]. Однак штучний інтелект (AI) трансформував цей процес, автоматизувавши та оптимізувавши різні аспекти. Обчислювальні методи революціонізували генерацію молекул, подібних до ліків, сполук зі специфічними фізико-хімічними властивостями, придатними для терапевтичного застосування. Нещодавні дослідження вивчали потенціал використання генних експресійних підписів та цільових рецепторів для керування проєктуванням цих молекул. Крім того, просторова конформація молекули, її тривимірна форма, відіграє вирішальну роль у взаємодії з біологічними мішенями (таргетами). Попри ці досягнення, у поточній літературі все ще існує потреба в більш всеосяжних дослідженнях, зокрема, потреба в більш комплексному дослідженні інтеграції генних експресійних підписів, цільових рецепторів та просторових конформацій для генерації молекул, подібних до ліків. Вирішення цих прогалин може призвести до значних прогресів у проєктуванні ліків, що в кінцевому підсумку принесє користь пацієнтам, швидше виводячи на ринок безпечні та ефективні ліки.

1.3. Формулювання наукової новизни та інноваційності проєкту.

Сучасний процес відкриття лікарських засобів характеризується високою складністю та тривалістю, що починається з ідентифікації цільових білків, проходить через етапи проєктування та скринінгу молекул, та завершується клінічними випробуваннями [10]. Втім, впровадження штучного інтелекту (AI)

та обчислювальних методів започаткувало трансформацію цього процесу, автоматизувавши його ключові етапи та оптимізувавши ресурси. Особливо помітним є вплив цих технологій на генерацію молекул з властивостями, аналогічними до лікарських засобів, дозволяючи значно прискорити пошук потенційних ліків [11].

Останні дослідження в галузі відкриття лікарських засобів фокусуються на вивченні можливостей використання генних експресійних підписів та цільових рецепторів у процесі проектування молекул. Також, значну увагу приділяється аналізу тривимірних конформацій молекул, адже саме просторова структура молекул визначає їхню здатність ефективно взаємодіяти з біологічними таргетами. Попри значний прогрес у цих дослідженнях, існує виразна потреба в більш комплексному вивченні інтеграції вищезгаданих параметрів для розробки нових молекул.

У даному контексті, проект, спрямований на розробку платформи для генерації drug-like молекул з використанням мультиоміксних даних, представляє важливий внесок у подолання існуючих прогалів у наукових знаннях. Ця платформа не лише спрощує ідентифікацію потенційних цільових білків за допомогою інтеграції генетичних і транскриптомних даних, але й забезпечує можливість глибокого аналізу взаємодій між молекулами та білками. Таким чином, проект сприяє розвитку методів швидкого відкриття ефективних та безпечних лікарських засобів, відкриваючи нові горизонти у фармацевтичних науках.

Перед початком розробки платформи було проведено кілька досліджень та опубліковано наукових робіт результати яких застосовуються у фінальному продукті.

Стаття "Toward efficient generation, correction, and properties control of unique drug-like structures" висвітлює прогрес у генерації молекул через використання глибоких нейронних мереж для мапінгу дискретних молекулярних репрезентацій у безперервний векторний простір. Дослідження

підкреслює важливість поєднання різних моделей машинного навчання для підвищення точності та ефективності процесу генерації нових молекул [12].

"Complex machine learning model needs complex testing: Examining predictability of molecular binding affinity by a graph neural network" представляє глибоке навчання, засноване на графах, для прогнозування афінності зв'язування між білками та лікарськими речовинами, наголошуючи на необхідності комплексного тестування та оцінки здатності моделі узагальнювати дані з метою вдосконалення процесу відкриття ліків [8].

"Advancing molecular graphs with descriptors for the prediction of chemical reaction yields" розкриває новаторську архітектуру нейронних мереж для передбачення виходів хімічних реакцій, що демонструє значення структурної інформації та описів на рівні молекул і реакцій для точнішого моделювання хімічних процесів [13].

Також на відміну від конкурентів, платформа ВТАІ виконує одразу кілька задач (сервісів), серед яких:

- Drug Design - Дизайн ліків. Розробка нових ліків з використанням мультиомічного знанневого графа (МОКГ), який включає дані про захворювання, цільові білки та ліки.
- Drug Repurposing - Перепрофілювання ліків. Виявлення нових медичних застосувань для існуючих ліків, зокрема пошук нових асоційованих цільових рецепторів (таргетів).
- Therapy development - Розробка терапій. Ідентифікація та ранжування генів, шляхів, медіаторів, які асоційовані з конкретним захворюванням

Враховуючи вищезазначене, наукова новизна дослідження полягає у створенні комплексної платформи, здатної аналізувати мультиоміксні дані для визначення оптимальних кандидатів у drug-like молекул, а це, в свою чергу, забезпечує прискорення процесу відкриття лікарських засобів, оптимізуючи взаємодію між молекулярними структурами та цільовими білками, що має вирішальне значення для розробки нових методів лікування.

1.4. Проведення аналізу конкурентів

Важливим аспектом дослідження на етапі ініціації проєкту є вивчення ринкового середовища та аналіз конкурентів у фармацевтичній сфері, зокрема компаній які використовують новітні технології відкриття ліків з застосуванням штучного інтелекту. Стратегічний аналіз конкурентного середовища має вирішальне значення для розуміння сильних і слабких сторін продуктів, пропонувані конкурентами, а також власних розробок. Це дозволяє визначити потенційні можливості для інновацій та розвитку, а також загрози, які створюють конкуренти на ринку.

У цьому розділі представлено поглиблений аналіз ключових компаній у сфері відкриття ліків з використанням технологій штучного інтелекту та машинного навчання, зокрема Insilico Medicine, Receptor.ai, Cyclica, Atomwise, Exscientia та Recursion Pharmaceuticals. Сильні та слабкі сторони цих компаній будуть проаналізовані для визначення стратегічних переваг платформи ВТАІ та сфер для подальшого розвитку та вдосконалення.

Аналіз конкурентів включатиме оцінку інноваційних технологій, аналітичних методів, а також можливостей генерації та оптимізації лікарських засобів, що використовуються кожною компанією. Особливу увагу буде приділено бенчмаркінгу з платформою ВТАІ для виявлення унікальних конкурентних переваг та можливостей для лідерства на ринку.

Результати цього аналізу стануть основою для розробки стратегічних рекомендацій щодо позиціонування платформи ВТАІ на ринку, розробки інноваційних рішень та розробки стратегій виходу на ринок. Він також допоможе визначити ключові аспекти для подальших досліджень та інновацій у сфері відкриття ліків з метою підвищення ефективності та безпеки майбутніх лікарських засобів.

Insilico Medicine

Insilico Medicines спеціалізується на використанні штучного інтелекту для розробки ліків і біомедичних досліджень. Підхід компанії зосереджений на

аналізі геномних, протеомних та інших біомедичних даних для виявлення потенційних цілей розробки ліків і створення нових молекул. Компанія також пропонує рішення для проєктування клінічних випробувань та автоматизованого виявлення кишень для розробки ліків [14, 15]. Виявлення кишеньок для генерації ліків є ключовим процесом у комп'ютерному моделюванні та дизайні лікарських засобів, особливо в контексті розробки інгібіторів, що взаємодіють з конкретними білками-мішенями. Кишеньки на поверхні білків - це уявні "пустоти" або "вдавнення", які можуть специфічно зв'язувати малі молекули, такі як потенційні лікарські засоби. Визначення цих структурних особливостей білка дозволяє вченим ідентифікувати, де малі молекули можуть ефективно зв'язуватися з білком, щоб модулювати його функцію, що є основою для розробки нових ліків. Детальніше підхід Insilico Medicine зображено на рисунку 1.2.

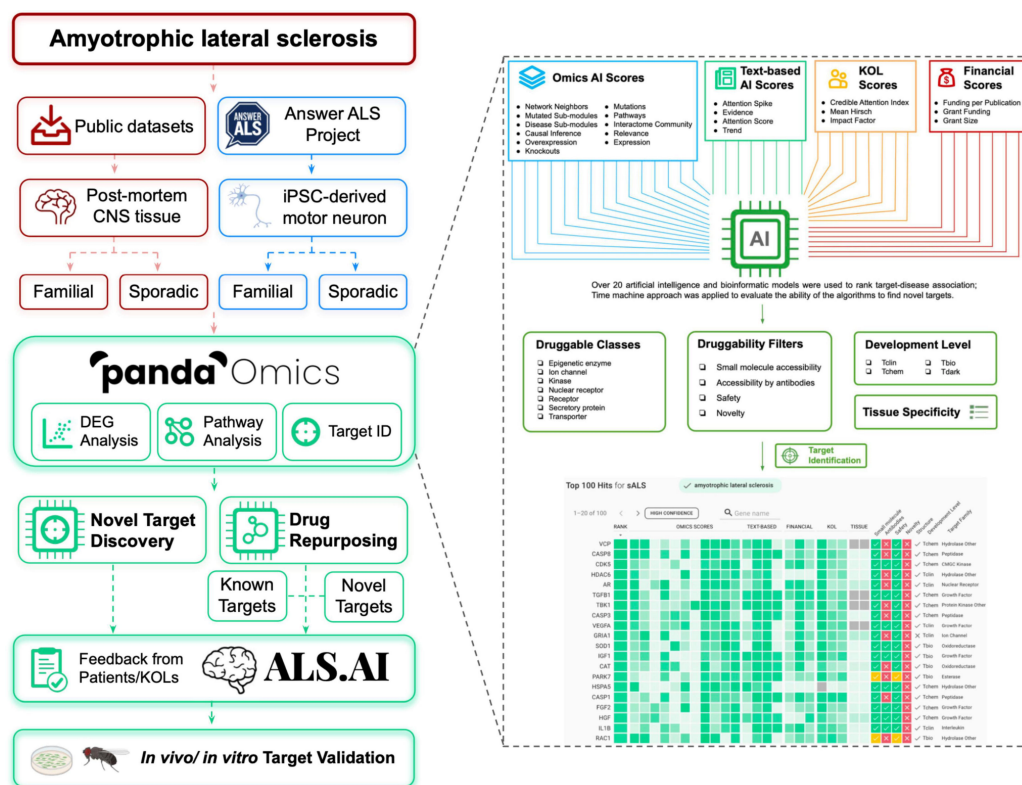


Рис. 1.2. Блок-схема для виявлення мішені ALS і перепрофілювання препарату [15]

Сильні сторони:

- Розширений аналіз геноміки і протеоміки з використанням штучного інтелекту для ідентифікації потенційних цілей для розробки ліків.
- Розробка клінічних випробувань з автоматизацією виявлення кишеньок для генерації ліків.
- Використання передових алгоритмів машинного навчання для прискорення процесу відкриття нових лікарських засобів.

Слабкі сторони:

- Можливі обмеження у варіативності методологій аналізу шляхів у порівнянні з платформою ВТАІ.
- Відсутність інтегрованих рішень для молекулярної симуляції на рівні, який пропонує ВТАІ.

Порівняння з платформою ВТАІ:

Платформа ВТАІ перевершує Insilico Medicine шляхом більш складного аналізу біологічних шляхів і використання молекулярної механіки для прогнозування властивостей лікарських засобів. Однак, Insilico Medicine має перевагу в дизайні клінічних випробувань. Проти компанія-розробник ВТАІ (blackthorn.ai) серед запропонованих сервісів має проведення клінічних випробувань.

Receptor.ai

Receptor.ai використовує квантові обчислення та інші передові технології для оптимізації кандидатів у лікарські препарати. Підхід компанії зосереджений на використанні квантової хімії та машинного навчання для покращення процесу розробки ліків, підвищення точності прогнозування та оптимізації процесів у лабораторії [16]. Пайплайн платформи зображено на рисунку 1.3.

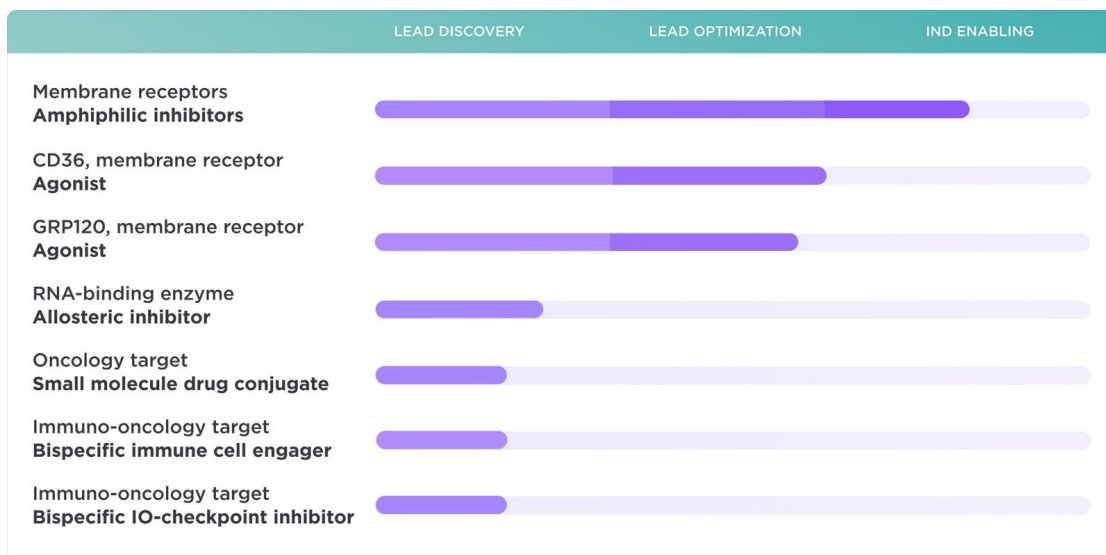


Рис. 1.3. Пайплайн розробки Resceptor.ai [16]

Сильні сторони:

- Використання квантових розрахунків для оптимізації кандидатів на ліки.
- Застосування передових технологій машинного навчання для вдосконалення процесу розробки ліків.

Слабкі сторони:

- Можлива вузька спеціалізація на квантових розрахунках, що обмежує широту застосування технологій.
- Обмежена інформація про інтеграцію метаболітного прогнозування.

Порівняння з платформою VTAI:

VTAI вирізняється ширшим спектром аналітичних можливостей, включаючи складний аналіз біологічних шляхів і молекулярну симуляцію. Водночас Resceptor.ai має сильні сторони у квантових розрахунках, що дозволяє виконувати обчислювальні задачі значно швидше, ніж традиційні комп'ютери.

Cyclica

Cyclica надає платформу, яка дає змогу дослідникам швидко визначати характеристики молекул, включно з ADME (абсорбція, дистрибуція, метаболізм, екскреція) та QSAR (квантитативні відносини структури і

активності). Технологія компанії спрямована на прогнозування молекулярних властивостей і оптимізацію лікарських кандидатів у лабораторних умовах [17].

Створення ліків від молекули до медичного препарату зображена на рисунку 1.4.

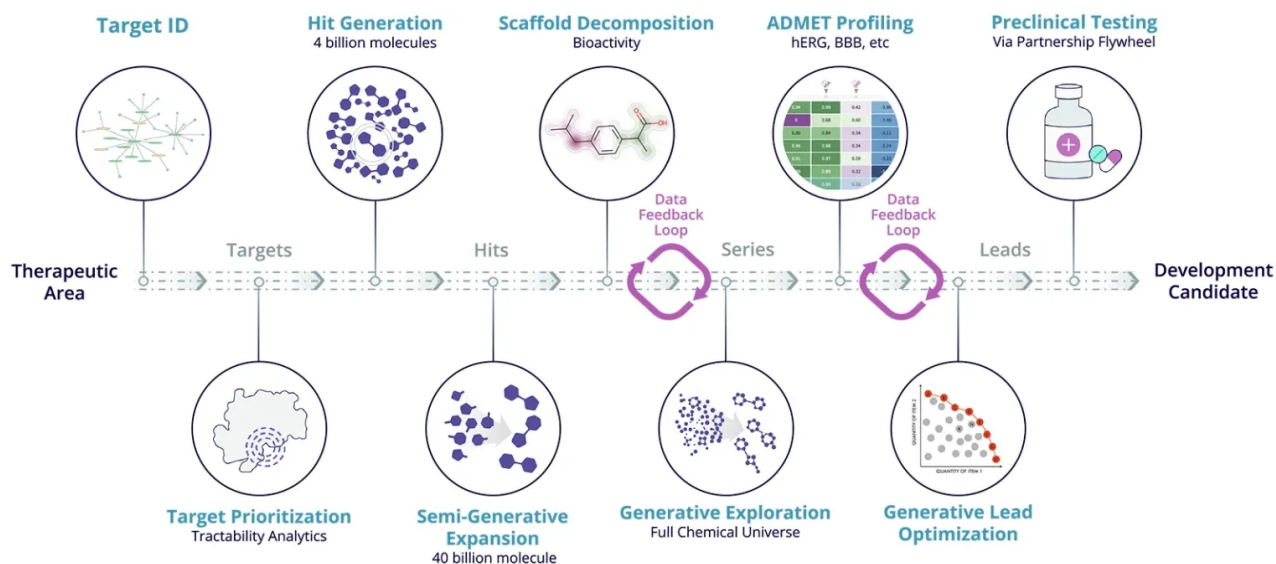


Рис. 1.4. Пайплайн розробки Сусліса [17]

Сильні сторони:

- Розширені можливості для оцінки молекулярних властивостей за допомогою ADME/QSAR.
- Швидка оцінка молекулярних властивостей для оптимізації лікарських кандидатів.

Слабкі сторони:

- Потенційні обмеження в глибині біологічного аналізу та молекулярного моделювання.
- Відсутність інтегрованого аналізу геноміки та протеоміки на рівні, що забезпечується платформою ВТАІ.

Порівняння з платформою ВТАІ:

ВТАІ має перевагу завдяки комплексному підходу до аналізу біологічних шляхів і використанню молекулярного моделювання. Сусліса, з іншого боку,

фокусується на швидкій фармакокінетичній характеристиці та доповнює аналітичні можливості ВТАІ в цій галузі.

Atomwise

Atomwise використовує штучний інтелект для поліпшення процесу відкриття ліків, включно з більш точним передбаченням молекулярної спорідненості до цільових білків. Платформа компанії прискорює відбір кандидатів у лікарські препарати, автоматизуючи процес виявлення високоафінних сполук.

Сильні сторони:

- Ефективне використання штучного інтелекту для покращення афінності молекул до цільових білків.
- Прискорення процесу відбору кандидатів на ліки за допомогою автоматизації.

Слабкі сторони:

- Потенційна вузька спеціалізація на афінності молекул, не охоплюючи інших аспектів розробки ліків.
- Обмежена інтеграція з біологічним аналізом шляхів на рівні, що пропонує ВТАІ.

Порівняння з платформою ВТАІ:

Платформа ВТАІ пропонує ширший спектр аналітичних інструментів, включаючи детальний біологічний аналіз, молекулярну симуляцію та дослідження афінності білків. Atomwise має сильні позиції у сфері оптимізації афінності, що може бути корисно для певних задач в рамках комплексних проєктів розробки ліків.

Exscientia

Exscientia - компанія, що займається розробкою лікарських препаратів на основі штучного інтелекту, орієнтована на швидкість і ефективність. Вона

використовує штучний інтелект для аналізу зрізів тканин і прогнозування ефективності ліків, забезпечуючи інноваційний підхід до розробки молекул.

Recursion Pharmaceuticals, Inc.

Recursion Pharmaceuticals аналізує великі обсяги біологічних і хімічних даних для пошуку нових ліків. Компанія використовує автоматизовані експериментальні системи для отримання даних, які потім аналізуються за допомогою машинного навчання для виявлення нових терапевтичних мішеней.

Підсумовуючи аналіз конкурентів можна зробити висновок, що платформа ВТАІ демонструє значні переваги в контексті комплексного аналізу біологічних шляхів, молекулярної симуляції та прогнозування властивостей лікарських засобів. Це ставить її в один ряд з провідними розробками у сфері відкриття ліків. Конкуренти, зі свого боку, пропонують унікальні рішення, які можуть доповнювати або конкурувати в певних аспектах, але загальна інтегрованість та глибина аналізу, що надається ВТАІ, забезпечує їй лідируючі позиції на ринку.

1.5. Обґрунтування інвестиційної доцільності проєкту

Згідно з даними, представленими на Економічному форумі в Давосі у 2024 році, розробка одного лікарського засобу може коштувати від 1.3 до 4 мільярдів доларів і займати від 10 до 15 років. Величезна кількість сполук проходить лабораторні дослідження щороку, але лише мала частина досягає клінічних випробувань. У Сполучених Штатах приблизно 88% препаратів, що досягають стадії клінічних випробувань, не отримують схвалення[18].

Ситуація ускладнюється тим, що рентабельність інвестицій у дослідження та розробку щороку знижується. Фармацевтичні компанії витрачають більше, ніж будь-коли, на відкриття та розробку ліків, а розробки все менш успішні. Десять років тому один долар, інвестований у дослідження та розробки, приносив прибуток у 10 центів. Сьогодні він становить менш ніж 2 центи [19].

Втім, застосування штучного інтелекту (ШІ) відкриває нові перспективи для подолання цих викликів. ШІ може сприяти розв'язанню цих проблем, допомагаючи визначити найперспективніші сполуки з необхідним ефектом і мінімізувати непередбачувані токсичні ефекти та побічні дії. Це може призвести до створення безпечніших, ефективніших лікувань і прискорити їх вихід на ринок. ШІ дозволяє ідентифікувати потенційні цільові білки та генерувати оптимізовані молекули для лікарських засобів набагато швидше, ніж це можливо традиційними методами, значно скорочуючи час дослідження та розробки.

Це підтверджується успішним досвідом компанії Insilico Medicine. Компанія оголосила, що її кандидат на ліки, розроблений за допомогою ШІ для лікування ідіопатичного легеневого фіброзу, захворювання, яке викликає рубцювання легенів, завершив етапи відкриття та передклінічних випробувань за всього 30 місяців. Це підкреслює значний потенціал ШІ у прискоренні розробки ліків, хоча і потребує подальшої перевірки та публікації у рецензованих виданнях[20].

Глобальний інститут McKinsey підрахував, що рішення штучного інтелекту, які застосовуються у фармацевтичній промисловості, можуть приносити майже 100 мільярдів доларів на рік – і це в системі охорони здоров'я лише в Сполучених Штатах [19]. Така перспектива підкреслює значення інвестицій у розробку інноваційних платформ для генерації ліків за допомогою ШІ як стратегічно важливого напрямку, здатного кардинально змінити парадигму фармацевтичної індустрії.

Розглядаючи викладені дані та приклади, стає очевидною доцільність інвестицій у розробку платформ на основі штучного інтелекту для генерації ліків. Такі інвестиції не тільки обіцяють значне скорочення часу і витрат на дослідження та розробку нових лікарських засобів, але й відкривають шлях до інноваційних рішень для лікування складних захворювань. Саме тому подальша підтримка та розвиток таких платформ є ключовим для досягнення прогресу в охороні здоров'я та покращення якості життя населення на глобальному рівні.

1.6. Постановка задачі дослідження, формулювання технічного завдання на розробку у вигляді паспорта проєкту

1.6.1. Опис проєкту

Проєкт (ВТАІ) - це інноваційна платформа для відкриття лікарських засобів, хворіб і таргетів, яка використовує штучний інтелект для аналізу великих масивів даних про взаємодію відомих молекул з цільовими білками та відповідними генетичними сигнатурами (РНК). Наш проєкт допомагає науковцям та фармацевтичним компаніям прискорити процес відкриття нових лікарських засобів, знижуючи час та витрати на дослідження.

Централізована платформа дозволяє дослідникам використовувати передові алгоритми машинного навчання та штучного інтелекту для генерації потенційних лікарських молекул та оцінки їх ефективності. Проєкт вирішує декілька ключових проблем у процесі відкриття ліків:

- Висока складність і тривалість процесу відкриття нових лікарських засобів. Традиційний процес може займати до десятків років та коштувати мільярди доларів. Платформа ВТАІ здатна значно скоротити цей час і витрати, використовуючи аналіз великих даних та штучний інтелект.

- Висока вартість розробки нових лікарських засобів. Великі витрати на класичну розробку нових ліків створюють значні фінансові та інноваційні виклики для фармацевтичної індустрії. Платформа ВТАІ дозволяє скоротити витрати в десятки разів.

- Обмеженість потенційних цільових білків та молекул для дослідження. ШІ може виявити нові, раніше не розглядані таргети для лікування різних захворювань, розширюючи горизонти потенційних лікарських засобів.

- Потреба в оптимізації процесу дослідження та розробки. Платформа ВТАІ надає дослідникам інструменти для ефективнішого відбору молекул та швидшого проведення передклінічних випробувань, зменшуючи кількість неефективних досліджень.

1.6.2. Мета, місія та візія проєкту

Мета: Метою проєкту ВТАІ є революційне трансформування процесу відкриття нових лікарських засобів за допомогою штучного інтелекту, щоб прискорити розробку та знизити витрати на виведення нових препаратів на ринок. Це досягається шляхом оптимізації вибору потенційних молекул для лікування різноманітних захворювань, а також швидкого і ефективного тестування цих молекул з мінімальними ризиками і побічними ефектами. Окрім того, платформа дозволяє виявити нові медичних застосування для існуючих ліків.

Місія: Місія проєкту ВТАІ полягає у наданні науковцям, дослідникам і фармацевтичним компаніям потужного інструменту, що дозволяє розширити можливості для відкриття нових, ефективних і безпечних лікарських засобів. Платформа прагне зробити процес відкриття ліків більш доступним і ефективним, що сприятиме пошуку лікування для захворювань, які на сьогодні вважаються невиліковними або важкими для лікування.

Візія: Візією проєкту ВТАІ є створення майбутнього, де процес відкриття і розробки нових лікарських засобів є значно прискореним, економічно ефективним і менш ризикованим. Ми прагнемо до світу, де кожне захворювання має потенційне лікування, а доступ до цих лікувань є широким і справедливим. Проєкт ВТАІ має на меті сприяти поліпшенню якості життя людей по всьому світу, надаючи науковій спільноті інструменти для розробки нових терапевтичних засобів швидше, ніж коли-небудь раніше.

1.6.3. Цінності проєкту

Відданість пацієнту (Patient Dedication)

Ми прагнемо, щоб кожне наше рішення сприяло поліпшенню добробуту і здоров'я людей. Кінцева мета кожного проєкту, кожної ініціативи та кожного дослідження — це покращення якості життя пацієнтів, надаючи їм доступ до безпечних, ефективних та інноваційних лікувань.

Невпинні інновації (Unceasing Innovation)

В основі нашої роботи лежить постійне прагнення до інновацій. Ми віримо, що немає меж для того, як штучний інтелект може трансформувати відкриття та розробку лікарських засобів. Від пошуку нових підходів до лікування невиліковних на сьогодні захворювань та розробки технологій, які можуть прискорити та оптимізувати наукові дослідження, ми невпинно розвиваємось для того, щоб втілити майбутнє медицини вже сьогодні.

Відкритість та співпраця (Openness and cooperation)

Ми прагнемо об'єднати наукову спільноту навколо спільної мети. Ми підтримуємо створення міцних партнерських відносин між дослідниками, лабораторіями та фармацевтичними компаніями для спільної роботи над відкриттями, які можуть змінити життя мільйонів людей.

1.6.4. Цілі проєкту

- Розширити базу даних мультиомічних інтеракцій: Інтегрувати в платформу ВТАІ дані по щонайменше 20 різних баз даних типу «drug-target» впродовж перших 3 місяців розробки.
- Забезпечити високу доступність та користувацьку зручність платформи ВТАІ: Розробити і запустити інтуїтивно зрозумілий користувацький інтерфейс, тестування якого покаже мінімум 90% задоволеності користувачів у рамках пілотного проєкту.
- Залучити мінімум 1000 активних користувачів з наукової спільноти та фармацевтичної індустрії протягом першого року функціонування платформи: Сформувати базу користувачів, які активно використовують платформу для дослідження і розробки нових лікарських засобів.
- Сприяти розвитку наукових досліджень у сфері біофармацевтики на національному та міжнародному рівнях: Створити партнерства з мінімум п'ятьма провідними науковими інституціями для спільних досліджень і розробки нових методик використання мультиомічних даних.

- Забезпечити повну функціональність платформи ВТАІ з високою надійністю: Реалізувати всі заплановані технічні функції, забезпечивши при цьому 99% часу безперебійної роботи платформи після її комерційного запуску.

1.6.5. Функціонал проєкту

Увесь проєкт складається з трьох сервісів, залежно від потреб і можливостей майбутніх користувачів. Тому буде доцільно розділити функціонал на 3 категорії, залежно від сервісу.

1.6.5.1 Функціонал для категорії дизайну ліків (Drug Design)

Цілі цієї категорії будуть актуальні для користувачів, які володіють інтелектуальною власністю на рецептор. Термін "рецептор" відноситься до специфічного білка або молекули в організмі, які можуть взаємодіяти з певними хімічними речовинами, включаючи лікарські засоби. Володіння інтелектуальною власністю на рецептор означає, що компанія або особа мають ексклюзивні права на використання конкретного рецептора в дослідженнях та розробці нових лікарських засобів.

Тут будуть актуальні такі цілі:

1. Визначити захворювання та/або стани (згідно з класифікацією ICD-11 або вказівками на Drugbank) які містять даний рецептор, для яких можуть бути розроблені нові лікувальні стратегії.

2. Виявити потенційних кандидатів серед існуючих портфелів лікарських засобів зареєстрованих FDA, або тих, що знаходяться на ранніх стадіях розробки, для повторного використання у лікуванні інших захворювань.

3. Забезпечити генерацію та відкриття нових, селективних лікарських засобів. Розробити та впровадити новітні методики з використанням штучного інтелекту для створення нових, селективних лікарських препаратів, спрямованих на конкретні молекулярні мішені, з метою проведення перших лабораторних тестів.

1.6.5.2 Функціонал для категорії перепрофілювання ліків (Drug Repurposing)

Перепрофілювання ліків це виявлення нових медичних застосувань для існуючих ліків, зокрема пошук нових асоційованих цільових рецепторів (таргетів). Це цілі для користувачів, які вже мають ліки чи drug-like молекули.

1. Виявити нові асоційовані рецептори (таргети): Визначити нові молекулярні мішені, зокрема рецептори, з якими може взаємодіяти розроблювана молекула, провівши аналіз і встановлення принаймні трьох потенційних нових мішеней.

2. Виявити нові показання (асоційовані захворювання/стани): Ідентифікувати асоційовані захворювання або стани, для яких потенційно може бути застосована розроблювана молекула, з встановленням конкретного списку нових показань.

Тут важливо зазначити, що захворювання рідко є результатом лише 1 рецептора. Більшість хворіб є результатом складної взаємодії між різними молекулами, клітинами та біологічними шляхами в організмі. Тому дана ціль проекту стосується пошуку набору умов для якої ліки можуть бути ефективними.

1.6.5.2 Функціонал для категорії розробки терапій (Therapy development)

Розробка терапій — це ідентифікація та ранжування генів, шляхів, медіаторів, які асоційовані з конкретним захворюванням. Ці цілі актуальні, коли рецептори та ліки є частково або повністю невідомі (наприклад коли відомо певний набір умов хвороби).

1. Оцінити доцільність розробки лікування певного захворювання. Провести аналіз потенційних лікувальних стратегій на основі ідентифікованих цільових мішеней для оцінки інвестиційної привабливості та економічної доцільності розробки нового лікування захворювання.

2. Ідентифікувати цільові мішені (рецептори) захворювань: Визначити та ранжувати асоційовані гени, біологічні шляхи, медіатори, кофактори тощо, з метою виявлення ключових молекулярних мішеней.

1.6.6. Зацікавлені сторони

На етапі ініціації проєкту ВТАІ, важливим аспектом є виявлення та аналіз зацікавлених сторін. Цей процес є вирішальним для розуміння інтересів, очікувань та потреб усіх груп, чиї внески або інтереси можуть бути залучені в проєкт.

Під час ініціації проєкту ВТАІ важливо зосередитися на встановленні ефективних комунікаційних каналів, визначенні ролей та відповідальностей, а також на встановленні ясних очікувань та цілей. Успіх проєкту значною мірою залежить від вміння керівника проєкту забезпечити високий рівень залученості та задоволення інтересів ключових стейкхолдерів.

Один із найвпливовіших стейкхолдерів, VP of Chemistry, James Wilson, як ініціатор проєкту, відіграє критичну роль у наданні стратегічного напрямку та фінансуванні. З його зацікавленістю у високому рівні повернення інвестицій, стає очевидним необхідність розробки бізнес-моделі, здатної максимізувати віддачу від вкладених коштів.

Консультанти проєкту, Sarah Brown та Robert Smith, внесуть вагомий вклад у формування тактичних та оперативних аспектів проєкту, а також у процес оцінки його успішності. Особлива увага приділяється врахуванню їх рекомендацій щодо стратегії та ризиків, асоційованих з проєктом.

Юрій Коваленко, як науковий консультант та партнер, відіграє ключову роль у висвітленні проєкту у наукових колах та публікації релевантних статей. Його участь забезпечує наукову обґрунтованість та високий рівень довіри до результатів проєкту.

На етапі ініціації проєкту вирішальним є залучення енд-юзерів, оскільки вони безпосередньо використовуватимуть продукт. Важливо забезпечити, щоб їх потреби та відгуки були враховані в процесі розробки продукту.

Детальніше процес управління зацікавленими сторонами буде розглянуто у наступних розділах.

1.6.7. Продукт проєкту

Продукт:

Інформаційна платформа для дослідження нових ліків, хвороб і таргетів за допомогою впровадження методів штучного інтелекту.

Цільова група:

- Фармацевтичні компанії, які прагнуть прискорити розробку нових ліків;
- Біотехнологічні стартапи, зосереджені на інноваційних дослідженнях у сфері медицини;
- Науково-дослідні інститути та академічні центри, що спеціалізуються на дослідженні захворювань;
- Інвестори та фонди що спеціалізуються на проєктах у сфері охорони здоров'я та фармацевтики.

Потреби, які задовольняє продукт:

- Оптимізація процесу розробки ліків: Фармацевтичні компанії мають потребу в засобах для зниження часу та витрат на дослідження та розробку нових лікарських засобів. Використання штучного інтелекту для прогнозування ефективності та безпеки потенційних медикаментів на ранніх стадіях дослідження може значно прискорити виведення ліків на ринок.
- Швидкий доступ до наукових даних про молекулярні мішені: Дослідники та розробники ліків мають потребу в оперативному доступі до інформації про потенційні молекулярні мішені для розробки нових препаратів, включаючи гени, білки, рецептори тощо. Вони прагнуть зрозуміти їхні функції, зв'язки з конкретними захворюваннями та можливості для цілеспрямованого лікування.
- Ідентифікація можливостей для репозиціонування існуючих ліків: Фармацевтичні компанії шукають способи розширити використання вже затверджених лікарських засобів (а отже і збільшення доходів з цих ліків)

шляхом виявлення нових показань для лікування. Вони зацікавлені в аналізі та пошуку доказів, які допоможуть знайти нові застосування для вже відомих молекул.

- Обмін знаннями та співпраця у галузі досліджень: Науковці та медичні дослідники прагнуть до спілкування та обміну інформацією з колегами по галузі для покращення розуміння механізмів захворювань, обговорення новітніх досліджень та розробки спільних проєктів. Вони цінують можливість взаємодіяти з однодумцями та спільно працювати над інноваційними рішеннями у сфері охорони здоров'я.

Опис цінності:

- Підвищення ефективності досліджень завдяки здатності швидко аналізувати великі обсяги даних та ідентифікувати важливі біологічні взаємодії;
- Зниження часу та витрат на розробку нових препаратів, оптимізуючи шлях від концепції до клінічних випробувань;
- Відкриття нових можливостей для лікування складних захворювань, включаючи ті, для яких раніше не було ефективних методів терапії;

1.6.8. Оцінка вартості

Для розробки пілотного проєкту необхідно 6 місяців.

Для ефективного управління вся команда поділена на 4 напрямки, а саме: в менеджмент, машинне навчання, Для ефективного управління вся команда поділена на 4 напрямки, а саме: менеджмент, машинне навчання, веб розробка, інжиніринг великих даних. Детальний склад команди, та оцінка вартості розписані в таблиці 1.1.

Важливою складовою проєктів розробки штучного інтелекту також є використання обчислювальних ресурсів, та пам'яті. Ми використовуємо можливості Google Cloud для зберігання даних, а також Google Cloud GPU.

Крім того, при плануванні також був закладений бюджет на ризики. Він складає 10 відсотків від загального бюджету проєкту. Ми наперед

повідомляємо замовника про те, що ці кошти повинні бути доступні в разі потреби. Але у випадку, якщо проєкт іде по плану — ці кошти залишаються замовником

Таблиця 1.1. Бюджет проєкту

Команда				
Назва витрат	Price, \$/h	Amount	Monthes	Price
PM	75	1	7	84000
SolutionArchitect	100	0,2	7	22400
Data Engineer	45	1	6	43200
BigData Technical Leader	75	0,5	6	36000
Junior Machine Learning Engineer	30	2	6	57600
Senior Machine Learning Engineer	60	1	6	57600
Sr. Bioinformatics Engineer	75	0,5	6	36000
Fullstack Web TechLead	75	0,3	6	21600
UI/UX Designer	30	1	6	28800
Frontend Developer	30	2	6	57600
Back-end dev	45	1	6	43200
DevOps	45	0,7	6	30240
Разом				518240
Обладнання				
	Price, \$/month	Amount	Monthes	Price
Google Cloud GPU	300	3	3	2700
10 Tb on Google Cloud	200		3	600
software licenses				1000
Разом				4300
Додаткові витрати				
Бюджет на ризики				52254
Загальна вартість				574794

1.6.9. Постановка цілей, задач та визначення основних робіт на всіх етапах життєвого циклу проєкту

В таблиці 1.2 детально описаний кожен етап життєвого циклу проєкту.

Таблиця 1.2. Життєвий цикл проекту

Визначення проекту	
Ціль:	Визначити концепцію проекту, створити паспорт проекту
Задачі	Основні роботи
<ul style="list-style-type: none"> • визначити ідею та мету проекту • визначити зацікавлені сторони проекту • визначити цілі та задачі проекту • опис основної ідеї • визначення цілей проекту та критеріїв їх досягнення • опис результатів досягнення цілей • розгляд альтернатив • визначення задач проекту • розробка концепції проекту • затвердження концепції • розрахунок бюджету проекту • Визначення вимог 	<ul style="list-style-type: none"> • опис основної ідеї • визначення цілей проекту та критеріїв їх досягнення • визначення зацікавлених сторін • опис бачення досягнення цілей • визначення задач проекту • розробка концепції проекту • затвердження концепції • створення дерева цілей • створення плану віх проекту
Планування проекту	
Ціль:	Підготуватись до старту проекту та розробити детальний план
Задачі	Основні роботи
<ul style="list-style-type: none"> • Узгодження процесів розробки та управління • Визначення програмного засобу для управління проектом • Організація робочих місць • Забезпечення необхідними технологіями для реалізації проекту • Формування вимог до команди, опис складу команди • Створення плану комунікації • Високорівневі архітектурні рішення • Створення проектної документації 	<ul style="list-style-type: none"> • Створити для майбутньої команди сторінку з інформацією про проект • Створення борди в системі управління проектами • Створити план комунікації, зібрати для неї усю необхідну інформацію (імена, пошти, зони відповідальності...) • Розробити та задокументувати майбутню архітектуру проекту • Розробити та задокументувати процеси розробки • Розробити WBS • Розробити організаційну структуру команди. Створити матрицю відповідальності • Уточнити за затвердити бюджет

<ul style="list-style-type: none"> • Узгодити вимоги до кінцевого продукту • Визначити обмеження проекту (resources, scope, time) • Провести аналіз ризиків проекту • найняти людей до команди • розподілити ролі на проєкті • визначити обов'язки членів команди • визначити схему підпорядкування в команді • Встановити цілі і задачі проєкту • Розробити детальний план проєкту і встановити “віхи проєкту” • Спланувати бюджет проєкту • Розробити план комунікації 	<ul style="list-style-type: none"> • Визначення вимог до продукту • Узгодження вимог із зацікавленими сторонами • Визначення ресурсів та обмежень проєкту • побудувати матрицю ризиків • Створення реєстру ризиків • Заповнення елементів організаційної структури проєкту і підпорядкування • Визначення обов'язків членів команди • розписати обов'язки та умови для посад • заповнення елементів організаційної структури • Набір співробітників з компанії-виконавця • Призначення ключових членів команди • Розробка цільового плану • Розробка бюджету • Розробка дерева робіт • Розробка плану комунікацій
Управління виконанням	
Ціль:	розробити кінцевий продукт
Задачі	Основні роботи
<ul style="list-style-type: none"> • інтегрувати та виконати операції проєкту відповідно до плану управління проєктом • забезпечити ефективну взаємодію між зацікавленими сторонами проєкту • налагодити ефективні робочі процеси • ретельно стежити за прогресом команди • Розробка дизайну • Тестування продукту • організація бази даних • управління відхиленнями • Вирішення конфліктів • Моніторинг бюджету та ресурсів 	<ul style="list-style-type: none"> • Проведення закупівлі ресурсів • Виконання робіт • Координація робіт, моніторинг просування проєкту. • Оперативне планування робіт • Розв'язання проблемних питань • Організація виконання робіт • Моніторинг бюджету і ресурсів • Мотивація команди • Управління командою розробників • Розробити дизайн • Дослідження предметної області • Розробка схеми БД • Імплементация БД • Наповнення БД • Тестування

Відслідковування проєкту	
Ціль:	забезпечити виконання цілей та результатів проєкту
Задачі	Основні роботи
<ul style="list-style-type: none"> • Відслідкувати вкладені зусилля та вартість • Відслідкувати прогрес проєкту • Переконатися у відповідності плану • управління відхиленнями по вартості і часу 	<ul style="list-style-type: none"> • Супровід інформаційної системи • Відслідковування виконання завдань проєкту • Ведення дошки з завданнями проєкту • Перевірка витрачених коштів під час виконання проєкту • Розв'язання проблем з розподілом ресурсів • Перевірка звіту про стан проєкту • Розгляд ризиків проєкту • Проведення демо сесії на інших зустрічах для інформування замовника про стан проєкту • Управління відхиленням по вартості та часу • Тестування
Закінчення проєкту	
Ціль:	Закінчити підтримку продукту згідно з угодою
Задачі	Основні роботи
<ul style="list-style-type: none"> • Закрити всі фази ЖЦ проєкту • Узгодити результати проєкту із замовником • Задokumentувати результати проєкту • Здати проєкт 	<ul style="list-style-type: none"> • Виконання дослідження готового продукту проєкту • Оцінка результатів проєкту та підбивання підсумків, порівняння очікуваних результатів та отриманих • аудит проєкту (зведення бюджету) • Здача проєкту замовнику • передати усі ресурси замовнику • Виконання підсумкових розрахунків • Розформування команди проєкту • Закриття проєкту • Організація івенту для команди та зацікавлених сторін

РОЗДІЛ 2. МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

2.1. Розробка моделі та моделювання ієрархічного опису проєкту.

Декомпозиція робіт проєкту

Ієрархічна структура робіт (WBS). Ієрархічна декомпозиція всього обсягу робіт, який команда проєкту повинна виконати для досягнення цілей проєкту та створення необхідних доробків [21]. WBS виступає основоположним інструментом для побудови архітектури управління проєктом. Він забезпечує чітку ієрархічну декомпозицію усіх робіт на керовані компоненти, що є критичним для забезпечення прозорості і точного відстежування прогресу.

Розробка WBS розпочинається з визначення ключових частин проєкту та подальшого розбиття кожної на більш дрібні, керовані завдання. Цей процес вимагає уважного аналізу вимог проєкту та тісної співпраці між усіма зацікавленими сторонами для забезпечення всебічного покриття проєктних цілей.

Для проєкту, який досліджується у магістерській роботі, декомпозиція робіт була здійснена на основі методу розбиття на продукти. Основою для цієї композиції служить функціонал платформи, який було визначено у 1-му розділі.

Результат до композиції робіт наведено на рисунку 2.1. Повний аналіз (декомпозиція на 5 рівнів) можна переглянути у додатку Б.

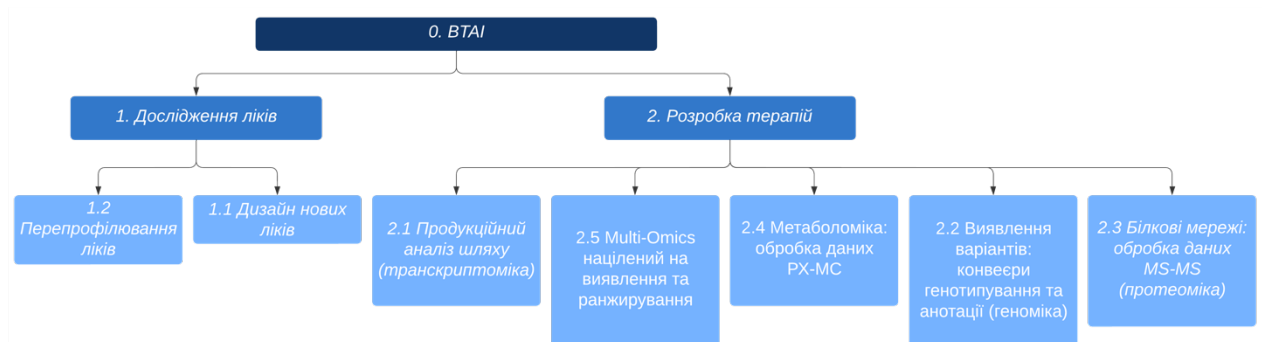


Рис.2.1. WBS проєкту на 3 рівні

Проект ВТАІ розбивається на ключові розділи, кожен з яких описує окрему частину роботи, необхідну для реалізації кінцевого продукту. Розробка платформи поділяється на наступні великі блоки: дизайн нових ліків, перепрофілювання ліків та розробка терапій, кожен з яких далі декомпонується на дрібніші задачі та підзадачі.

У розділі дизайну нових ліків включено задачі з UI/UX дизайну, розробки фронтенду та бекенду, налаштування інфраструктури та інтеграції систем, що покриває як візуальний так і технічний аспекти платформи. Цей розділ також включає створення баз даних для зберігання згенерованих сполук та управління залежностями ліків, мішеней та захворювань.

Розділ перепрофілювання ліків зосереджується на ідентифікації нових потенційних застосувань існуючих лікарських препаратів. Це включає розробку графу знань, прогнозування зв'язків за допомогою алгоритмів машинного навчання, а також розробку інтерфейсів та бекенду для підтримки цих процесів.

Останній великий розділ, розробка терапій, охоплює більш спеціалізовані аналітичні та обчислювальні завдання, такі як транскриптоміка, геноміка, протеоміка та метаболоміка. Ці задачі включають розробку спеціалізованих конвеєрів обробки даних, дизайн інтерфейсів для кінцевих користувачів та налаштування інфраструктури для обробки та аналізу великих наборів даних.

WBS проекту ВТАІ забезпечує детальний опис і організацію всіх завдань, необхідних для досягнення кінцевих цілей проекту, і є основою для подальшого планування, розподілу ресурсів, оцінки бюджету та відстеження прогресу в роботі.

2.2. Розрахунок моделі та моделювання розкладу робіт проекту

У проекті ВТАІ було застосовано гібридний підхід до розробки, який об'єднує елементи водоспадної моделі та Agile-методології [21]. Цей підхід дозволяє ефективно поєднати стратегічне планування на вищому рівні з гнучкістю в роботі над конкретними завданнями. Водоспадний метод характеризується послідовним переходом від одного етапу до іншого, що дає

змогу чітко структурувати ініціацію, планування та закриття проєкту. З іншого боку, проєкти з розробки штучного інтелекту характеризуються високим рівнем невизначеності. В таких випадках гнучкі методології надають можливість швидко реагувати на зміни, відповідати на непередбачувані виклики та забезпечувати неперервний зворотний зв'язок зі стейкхолдерами.

В рамках дипломної роботи розглядається в основному розробка MVP продукту, але йому передував довгий етап проведення експериментів виключно для дослідження ліків. Загалом послідовність всіх етапів життєвого циклу проєкту та етапів проєкту ВТАІ можна виділити в такі частини каскадної моделі:

- Ініціація експериментальної частини Drug Discovery. На основі попередніх досліджень та публікацій було прийняте рішення провести експерименти в сфері Drug Discovery.

- Планування Drug Discovery. Розробка детальної стратегії досліджень, вибір методів та технологій дослідження.

- Проведення експериментів. Включає емпіричні дослідження, які направлені на тестування гіпотез і збір даних для розвитку нових лікарських препаратів.

- Ініціація створення платформи ВТАІ. На основі отриманих результатів експериментів було прийнято рішення розробляти повноцінну платформу. Ініціація проєкту передбачає визначення його масштабів, цілей та очікувань зацікавлених сторін, а також алокацію ресурсів.

- Планування створення платформи ВТАІ. Планування етапів розробки включає розробку детального розкладу робіт (WBS) та ресурсного планування, яке базується на методології PMBOK (Project Management Body of Knowledge) від PMI (Project Management Institute). Ця методологія наголошує на важливості інтеграції різних сфер знань управління проєктами для оптимізації процесів.

- Розробка MVP платформи (функціонал платформи з WBS індексом 1), моніторинг та контроль. Під час розробки MVP використовується принцип ітеративного покращення продукту, де кожен спринт забезпечує прогрес та

внесення корективів на основі зворотного зв'язку. Також кожен спринт включає експериментальну частину, розробку дизайн, фронтенду та бекенду і роботу з великими даними. Такий підхід сприяє мінімізації ризиків, оскільки команда може швидко адаптувати розробку до нових вимог та змін у проєктному середовищі. Детальніше організацію спринтів описано в наступних розділах магістерської роботи.

- Запуск платформи на ринку. Етап виходу на ринок включає не тільки технічне впровадження продукту, а й маркетингові заходи, а також підготовку підтримки користувачів та збір відгуків. На даному етапі використовується аналіз SWOT для оцінки ринкових можливостей та загроз, а також методики бізнес-аналітики для прийняття стратегічних рішень.

- Розробка MMR платформи (функціонал платформи з WBS індексом 2). Ця частина також використовує принцип ітеративного покращення, але включає більше експериментів, ніж розробка MVP.

- Закриття проєкту. Ця фаза проєкту передбачає перегляд здобутих результатів, узагальнення отриманого досвіду та його документування, що є критично важливим для висновків та рекомендацій.

Для моделювання розкладу робіт проєкту було використано платформу GanttPro[22]. Це інтерактивний онлайн-інструмент для створення діаграм Ганта, що дозволяє планувати, координувати завдання та управляти проєктами. Він пропонує готові шаблони для різноманітних бізнес-проєктів, інтуїтивні засоби для розподілу ресурсів, налаштування залежностей та визначення пріоритетів завдань.

Для початку необхідно створити проєкт. При створенні вказую дату початку проєкту – 12.09.2022. Також налаштовую календар на 5-тиденний робочий тиждень і додаю всі додаткові вихідні дні. На час воєнного стану в компанії blackthorn.ai їх лише 4 на рік: Новий Рік, Великодній понеділок, День Незалежності та Різдво (25 грудня). Частина календаря з додатковими вихідними зображено на рисунку 2.2.

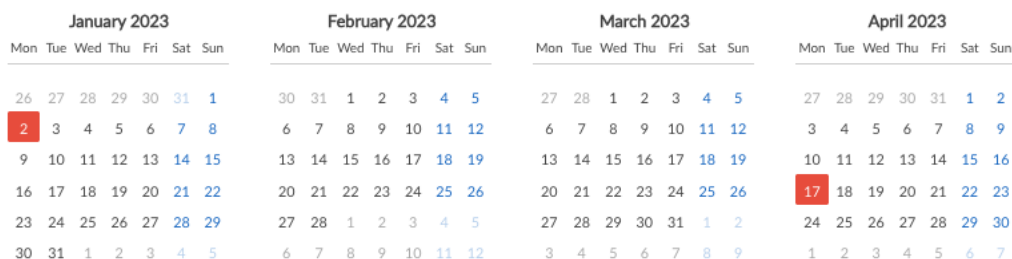


Рис.2.2. Календар з вихідними на 2023 рік.

Наступним етапом необхідно додати усі роботи (етапи проекту), тривалість та залежності між завданнями. За поточних налаштувань дати початку кожного етапу будуть проставлені автоматично, з врахуванням вихідних. Критичний шлях відображається червоним. Діаграма Ганта для всіх етапів зображена на рисунку 2.3.

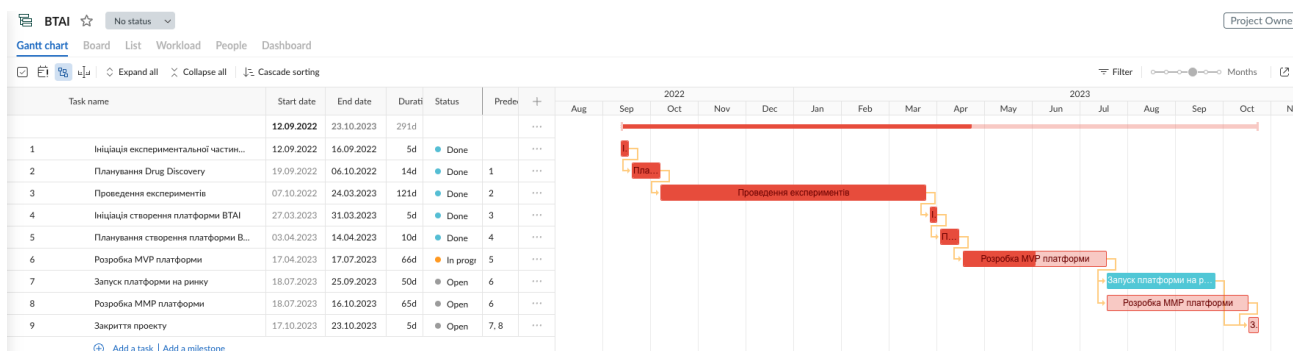


Рис. 2.3. Моделювання розкладу робіт проекту

Важливо зазначити, що управління послідовністю та залежностями між завданнями в проекті вимагає точного і грамотного підходу. Проект розпочинається з ініціації експериментальної частини Drug Discovery (задача 1), яка безпосередньо впливає на планування цієї фази (задача 2). Після планування негайно розпочинається проведення експериментів (задача 3), що є критичним для валідації наукових даних та підготовки до розробки платформи.

Після з завершення експериментів, оскільки було отримано хороші результати, ініціюється процес створення платформи BTAI (задача 4), що

зумовлює детальне планування архітектури та функціональності майбутньої платформи (задача 5). На основі цього плану розробляється MVP (мінімально життєздатний продукт) платформи (задача 6), який є вирішальним для подальшого впровадження продукту на ринок.

Завершальні фази проєкту включають паралельне запускання MVP на ринку (задача 7) та одночасну розробку розширеної версії платформи, відомої як ММР (задача 8). Обидві ці задачі залежать від успішної реалізації MVP і можуть початись лише після закінчення розробки. Завершенням проєкту є закриття проєкту (задача 9), що здійснюється після успішного завершення і запуску обох версій платформи. Цей етап включає аналіз досягнутих результатів, оцінку впливу проєкту на цільові аудиторії та підготовку звітності для інвесторів та зацікавлених сторін.

Оскільки використання адаптивних методологій передбачає гнучке реагування на зміни у процесі розробки, традиційне відображення таких етапів на діаграмі Ганта не застосовується. Натомість, оцінювання тривалості кожного з цих етапів відбувається з використанням методу максимальних та мінімальних оцінок, на основі якого обраховується середнє значення часу виконання.

Для кожної задачі визначено відповідних виконавців та необхідні ресурси, що дозволяє забезпечити точне планування ресурсів та ефективне розподілення відповідальностей серед команди проєкту.

Оцінки для етапу MVP представлено у таблиці 2.1 документації проєкту. Для ММР викладено відповідні дані в додатку 3, що дозволяє забезпечити глибше розуміння та контроль за використанням ресурсів та часу, що витрачається на кожен етап.

Така структура планування сприяє забезпеченню ефективності управління проєктом та оптимізації процесів розробки, водночас дозволяючи команді швидко адаптуватися до змін у проєктних вимогах і умовах зовнішнього середовища.

Таблиця 2.1. Оцінка часу та ресурсів для MVP та MMP

№	Задача	Оцінка часу, дн	Відповідальні
1	2	3	4
1	Дослідження ліків		
1.1	Дизайн нових ліків		
1.1.1	UI/UX дизайн		UI/UX Designer
1.1.2	Фронтенд	50	Frontend Developer
1.1.2.1	Флоу автентифікації		
1.1.2.2	Інформаційна панель з статистикою генерації ліків		
1.1.2.3	Інформаційна панель з фільтрами для згенерованих ліків		
1.1.2.4	Сторінка завантаження та перегляду рецептора		
1.1.2.5	Сторінка огляду стикованої складної структури		
1.1.2.6	Сторінка перегляду результатів моделювання		
1.1.2.7	Параметризація та запуск генерації		
1.1.2.8	Застосування фільтрів для згенерованих ліків		
1.1.2.9	Застосування попарних фільтрів препарат-рецептор		
1.1.2.10	Налаштування та запуск моделювання		
1.1.2.11	Налаштування та запуск стикування		
1.1.3	Розробка та реалізація схеми рівня даних	20	Backend developer
1.1.3.1	Розробка та реалізація схеми рівня даних		
1.1.3.2	БД згенерованих сполук		

Продовження таблиці 2.1

1	2	3	4
1.1.3.3	Двійкові файли та інші проміжні сховища (Storage)		
1.1.3.4	Neo4j для графу зв'язків ліків-мішеней-захворювання		
1.1.4	Бек-енд		
1.1.4.1	бекенд+API для фільтрів подібності до ліків та інших	40	Backend developer
1.1.4.2	бекенд+API для перегляду стикованих комплексів		
1.1.4.3	бекенд+API для перегляду результатів генерацій		
1.1.5	Налаштування CI/CD	25	DevOps Engineer
1.1.6	Terraform: розгортання інфраструктури	30	DevOps Engineer
1.1.7	Логування та моніторинг	10	DevOps Engineer
1.1.8	Розгортання та інтеграція	40	Sr. Bioinformatician, ML Engineer (1)
1.1.9	Тестування і виправлення помилок	20	ML Engineer, full-stack
1.1.10	Оптимізація продуктивності веб-сервісу	10	ML Engineer, full-stack
1.1.11	Підготовка чистого коду та документації, презентація результатів, обмін знаннями	5	ML Engineer, full-stack
1.2	Перепрофілювання ліків		
1.2.1	Проектування графу знань вибір ознак лікарських засобів, цілей та показань	20	Sr. Bioinformatician, ML Engineer (2)
1.2.2	Прогнозування зв'язків на основі евристик	10	
1.2.3	Генерація та інтеграція ознак вузлів графу	20	

Закінчення таблиці 2.1

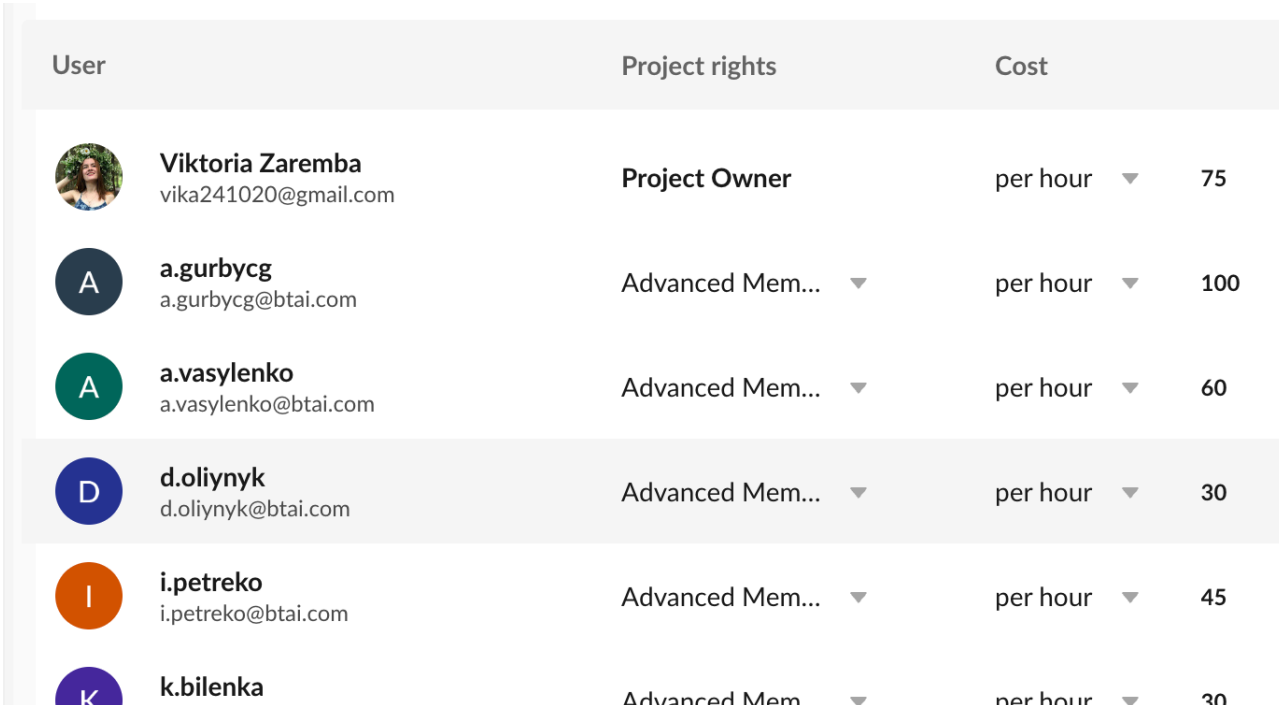
1	2	3	4
1.2.4	Рекомендація захворювань для повторного використання ліків Прогнозування зв'язків на графі за допомогою ML	20	
1.2.5	Рекомендація таргетів для повторного використання ліків Прогнозування зв'язків на графі за допомогою ML	20	
1.2.6	Рекомендація асоціацій захворювань та цілей Прогнозування зв'язків на графі за допомогою ML	20	
1.2.7	UI/UX Дизайн для перепрофілювання ліків	40	UI/UX Designer
1.2.8	Розробка фронтенду		Frontend Developer
1.2.9	Розробка бекенду		Backend Developer
1.2.10	Налаштування CI/CD	20	DevOps
1.2.11	Terraform: розгортання інфраструктури		
1.2.12	Логування та моніторинг		
1.2.13	Тестування та виправлення помилок, оптимізація продуктивності веб-сервісу	20	ML Engineer, full-stack
1.2.14	Підготовка чистого коду та документації, презентація результатів, обмін знаннями	5	ML Engineer, full-stack

Таким чином, впровадження гібридного підходу на проєкті ВТАІ дозволило поєднати переваги як традиційного, так і гнучкого управління проєктами, сприяючи ефективній розробці інноваційного продукту.

2.3. Розрахунок моделі та моделювання розподілу ресурсів

Розрахунок моделі та моделювання розподілу ресурсів на проєкті ВТАІ проводиться з використанням онлайн-платформи GanttPRO, яка дозволяє візуалізувати розподіл ресурсів за допомогою діаграм Ганта. Використання цієї платформи сприяє чіткому плануванню і наочному представленню розподілу ресурсів між завданнями і етапами проєкту, що є критично важливим для забезпечення ефективності і своєчасного виконання робіт.

На першому етапі відбувається визначення основних ресурсів проєкту, які включають людські, технічні, інформаційні та фінансові компоненти. Більшість ресурсів вже визначені на етапі ініціації проєкту, але тут важливо не лише знати яких спеціалістів і в якій кількості потрібно, а й вже знати конкретних людей, які будуть виконувати проєкт. Усі ресурси додано у відповідній вкладці платформи GanttPro (Рис. 2.4.)









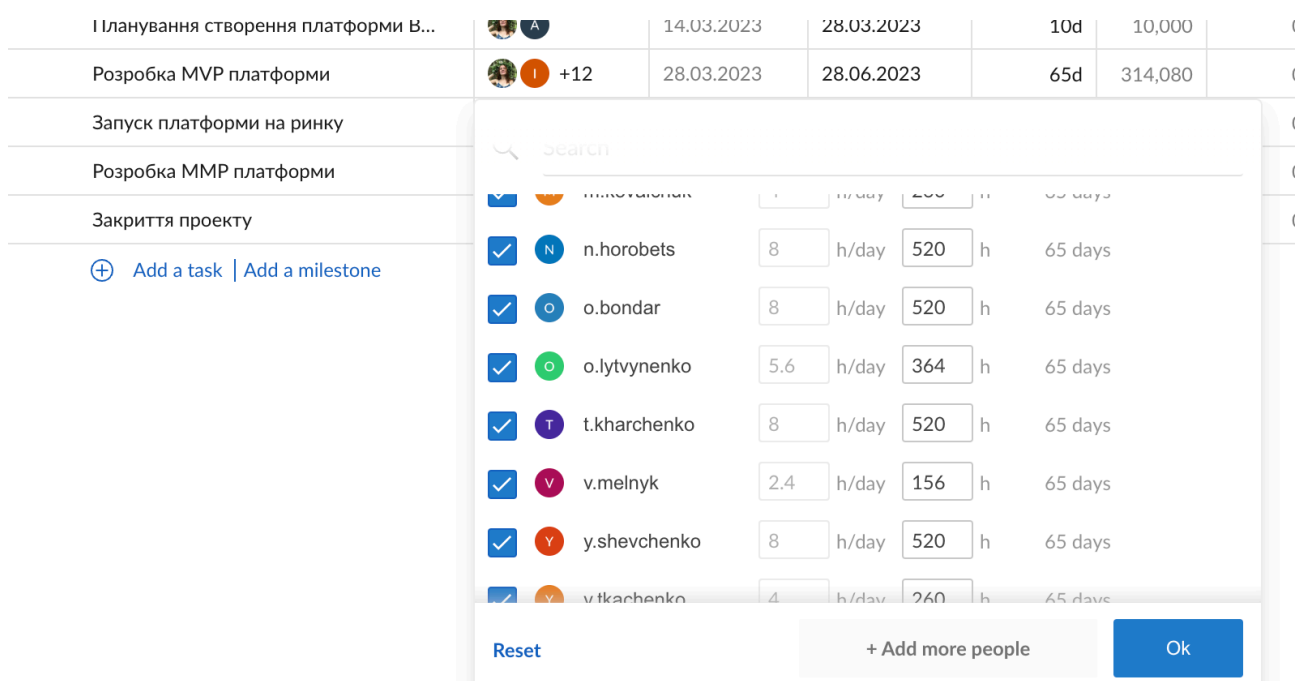
User	Project rights	Cost
 Viktoria Zarembo vika241020@gmail.com	Project Owner	per hour ▼ 75
 a.gurbycg a.gurbycg@btai.com	Advanced Mem... ▼	per hour ▼ 100
 a.vasylenko a.vasylenko@btai.com	Advanced Mem... ▼	per hour ▼ 60
 d.oliynyk d.oliynyk@btai.com	Advanced Mem... ▼	per hour ▼ 30
 i.petreko i.petreko@btai.com	Advanced Mem... ▼	per hour ▼ 45
 k.bilenka	Advanced Mem... ▼	per hour ▼ 30

Рис. 2.4. Додавання ресурсів на платформі GanttPro

Кожен ресурс оцінюється з точки зору його необхідності для виконання конкретних завдань проєкту. Це дозволяє оптимізувати використання кожного ресурсу і уникнути перевантаження ресурсів або їх недостатнього використання.

Далі на кожному з етапів задачі призначаються ресурси. GanttPro дає можливість присвоювати кілька ресурсів за раз, а також для кожного з них вказувати навантаження. Завантаження ресурсів для етапу розробки MVP вказане на рисунку 2.5.



Task	Start Date	End Date	Duration	Cost
Планування створення платформи В...	14.03.2023	28.03.2023	10d	10,000
Розробка MVP платформи	28.03.2023	28.06.2023	65d	314,080
Запуск платформи на ринку				
Розробка ММР платформи				
Закриття проєкту				

Resource	h/day	Total h	Days
n.horobets	8	520	65 days
o.bondar	8	520	65 days
o.lytvynenko	5.6	364	65 days
t.kharchenko	8	520	65 days
v.melnyk	2.4	156	65 days
y.shevchenko	8	520	65 days
v.tkachenko	4	260	65 days

Рис.2.5. Призначення ресурсів на етапі розробки MVP

Завдяки детальному моніторингу навантаження ресурсів (Рис.2.6.), GanttPRO сприяє своєчасному виявленню і вирішенню проблем, пов'язаних з розподілом робочого часу, що дозволяє командам уникати перевантаження і підтримувати високу продуктивність. Інструмент надає функції відстеження прогресу задач, розрахунку бюджету та витрат на ресурси, а також дозволяє формувати звіти, що включають аналіз використання ресурсів та ефективності їхнього використання в рамках проєкту. Навантаження ресурсів можна переглянути в різному масштабі – від днів до років.

Resource	June 2023			July 2023						August 2023				September 2023			
	12-18 (24w)	19-25 (25w)	26-2 (26w)	3-9 (27w)	10-16 (28w)	17-23 (29w)	24-30 (30w)	31-6 (31w)	7-13 (32w)	14-20 (33w)	21-27 (34w)	28-3 (35w)	4-10 (36w)	11-17 (37w)	18-24 (38w)	25-	
Marketing	0	0	22	40	40	40	40	40	40	40	32	40	18	0	0		
Viktoria Zaremba	40	40	38.2	36.7	36.7	36.7	36.7	36.7	36.7	36.7	29.3	36.7	33.2	30	30		
a.gurbycg	40	40	27.2	16.7	16.7	16.7	16.7	16.7	16.7	16.7	13.3	16.7	13.2	10	10		
a.vasylenko	40	40	18	0	0	0	0	0	0	0	0	0	0	0	0		
d.olymyk	40	40	40	40	40	40	40	40	40	40	32	40	40	40	40		
i.petreko	40	40	40	40	40	40	40	40	40	40	32	40	40	40	40		
k.bilenka	40	40	40	40	40	40	40	40	40	40	32	40	40	40	40		
m.kovalchuk	20	20	20	20	20	20	20	20	20	20	16	20	20	20	20		
n.horobets	40	40	40	40	40	40	40	40	40	40	32	40	40	40	40		
o.bondar	40	40	40	40	40	40	40	40	40	40	32	40	40	40	40		
o.lytvynenko	28	28	17	7.95	7.95	7.95	7.95	7.95	7.95	7.95	6.36	7.95	7.95	7.95	7.95		
t.kharchenko	40	40	40	40	40	40	40	40	40	40	32	40	40	40	40		
v.melnyk	12	12	12	12	12	12	12	12	12	12	9.56	12	12	12	12		
y.shevchenko	40	40	29	20	20	20	20	20	20	20	16	20	20	20	20		
y.tkachenko	20	20	9	0	0	0	0	0	0	0	0	0	0	0	0		

Рис. 2.6. Моніторинг навантаження ресурсів

Підсумовуючи, інтеграція інструментів розрахунку та моделювання ресурсів з використанням GanttPRO відіграє ключову роль у плануванні і виконанні проєктів ВТАІ, забезпечуючи високий рівень управління проєктними ресурсами і оптимізацію робочого навантаження.

2.4. Розрахунок моделі та моделювання вартості ІТ проєкту

Після завершення процедури призначення ресурсів до конкретних задач та встановлення ставки вартості за годину праці кожного ресурсу, платформа GanttPro автоматично проводить розрахунок загальної вартості кожного етапу проєкту. Це дозволяє отримати детальну інформацію щодо фінансових витрат проєкту, сприяючи точному бюджетуванню та ефективному розподілу ресурсів. Використання такого підходу є важливим для підтримки фінансової дисципліни та забезпечення відповідності загальним фінансовим планам проєкту.

На етапі ініціації проєкту, який був детально описаний у першому розділі документації, вже було проведено розрахунок бюджету, який охоплює планування, ініціацію, а також розробку MVP та MMR. Результати цих розрахунків у GanttPro підтвердили первісні оцінки (Рис 2.7.), що свідчить про

високу точність попередньої фінансової моделі та ефективність використаних інструментів для планування.





















Task name	Budget	Assigned
	725,383	
1 Ініціація експериментальної частини Drug Discovery.	2,000	 a.gurbycg
2 Планування Drug Discovery	5,950	 
3 Проведення експериментів	163,030	  +3
4 Ініціація створення платформи BTAI	3,500	 
5 Планування створення платформи BTAI	10,000	 
6 Розробка MVP платформи	314,080	  +12
7 Запуск платформи на ринку	31,033	  
8 Розробка MMP платформи	192,065	  +10
9 Закриття проекту  	3,725	  +12

Рис. 2.7. Розрахунок бюджету різних етапів проєкту в Gantt Pro

Адекватне бюджетування та ретельний контроль за витратами є ключовими для успіху проєкту, особливо в умовах, коли ресурси обмежені та необхідно максимально ефективно використовувати кожну інвестицію. Платформа GanttPro надає можливість не тільки планувати витрати, але й вчасно вносити корективи в процесі реалізації проєкту, що дозволяє оперативно реагувати на будь-які зміни в проєктному середовищі, забезпечуючи краще управління ризиками та фінансовою стійкістю проєкту.

Для вимірювання вартості та ефективності виконання проєктних робіт у відповідності до бюджетних та часових рамок важливим є обрахунок показників BCWP, ACWP та BCWS. Ці показники забезпечують можливість моніторингу фінансових і оперативних ризиків, допомагають виявляти відхилення від плану на ранніх стадіях та вносити необхідні корективи для забезпечення виконання проєкту в межах затверджених параметрів.

- BCWS (Budgeted Cost of Work Scheduled) для ВТАІ показує, скільки коштів було б заплановано витратити на виконання запланованих робіт до поточного моменту. Цей показник допомагає зрозуміти, чи рухається проєкт згідно з встановленим бюджетом та графіком, і вказує на потенційні зони, де можливе перевищення витрат.

- BCWP (Budgeted Cost of Work Performed) для проєкту ВТАІ визначається шляхом порівняння бюджету, який був запланований на виконанні роботи станом на поточну дату, з фактично виконаними роботами. Такий підхід дозволяє зрозуміти, чи є проєкт відповідним до бюджетних рамок і чи ефективно використовуються ресурси.

- ACWP (Actual Cost of Work Performed) вимірює реальні витрати, понесені на виконання робіт проєкту. Для проєкту ВТАІ це означає зіставлення фактичних витрат із запланованими, що допомагає управлінцям виявляти та аналізувати причини фінансових відхилень та вживати заходів для корекції в майбутньому.

Оскільки платформа GanttPro не надає можливості безпосередньо обраховувати ці показники, необхідно було знайти альтернативу для ефективного управління бюджетом та прийняття управлінських рішень. В рамках проєкту ВТАІ обрахунок всіх необхідних показників здійснено вручну за допомогою Excel. Це забезпечує необхідну гнучкість та адаптивність у визначенні фінансового стану проєкту та дозволяє керувати ресурсами більш точно та ефективно.

Для кожного етапу проєкту були створені окремі таблиці, в яких для кожного місяця заповнено дані для BCWP, ACWP та BCWS. Це дозволяє систематично відстежувати фінансові показники та виявляти можливі відхилення від запланованих параметрів на ранніх етапах проєкту, що є ключовим для своєчасного впровадження необхідних корективних дій. Приклад для BCWP наведено в таблиці 2.2.

Аналіз даних здійснюється шляхом візуалізації на графіках (Рис. 2.8.), що сприяє кращому розумінню динаміки проєкту та забезпечує більш ефективне управління змінами.

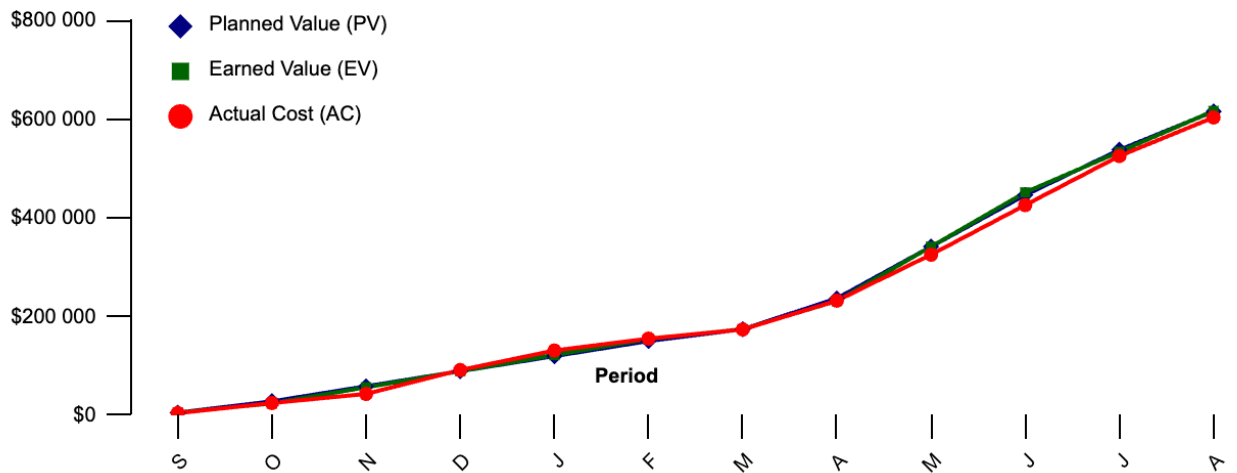


Рис.2.8. Аналіз заробленої вартості проєкту

Фактичні витрати, ACWP, спочатку були нижчими за планові, але поступово, до кінця 2022 року, зросли і значно перевищили планову вартість у грудні. Це може бути свідченням непередбачених витрат або збільшення обсягу робіт. Протягом 2023 року ця тенденція тривала, але різниця між планом і фактичними витратами зменшилася, свідчачи про поліпшення контролю витрат.

Заробітна вартість, BCWP, показує вартість вже виконаних робіт і за більшу частину проєкту трималася в межах плану або незначно його перевищувала, особливо в останній період. Це говорить про те, що виконання проєкту було в цілому ефективним і відповідало плановим очікуванням.

Таким чином, проєкт ВТАІ продемонстрував значну динаміку у бюджетному управлінні та ефективність виконання робіт, що дозволяє зробити висновок про загальний успіх проєкту з точки зору вартісного планування та управління.

РОЗДІЛ 3. ПЛАНУВАННЯ ЕЛЕМЕНТІВ УПРАВЛІННЯ ПРОЄКТОМ

3.1. Розробка організаційної структури управління проєктом. Формування команди проєкту

Важливим етапом при плануванні проєкту є розробка організаційної структури управління проєкту. Це забезпечує ефективну розробку та впровадження проєктних ініціатив, особливо у складних проєктів з

Великою командою та різноманітними технологіями. Ефективна організаційна структура забезпечує чітке розподілення ролей, відповідальностей та повноважень між членами проєктної команди, а також уможливорює якісне спілкування та координацію зусиль.

Оскільки для реалізації проєкту використовуються різні технології, а саме інжиніринг великих даних, розробка штучного інтелекту (машинне навчання, науковий відділ) , та веб розробка — було прийняте рішення розподілити роботу між декількома спеціалізованими командами. Кожна з команд має свої унікальні завдання та відповідальності:

Команда управління

Ця команда відповідає за загальне керівництво проєктом, включаючи стратегічне планування, прийняття рішень, координацію з усіма учасниками та зацікавленими сторонами, а також за забезпечення відповідності проєктних цілей бізнес-вимогам. Вони також надають технічне керівництво та підтримку командам розробників.

- Вікторія Заремба - Project Manager (PM). Відповідає за загальне керівництво проєктом, стратегічне планування та координацію між усіма учасниками.

- Олександр Гурбич - Solution Architect. Розробляє архітектурні рішення, забезпечує впровадження машинного навчання в проєктні рішення та відповідає за інтеграцію систем.

Команда розробки даних (DE)

Ця команда спеціалізується на обробці та аналізі великих обсягів даних. Вони розробляють та оптимізують процеси збору, зберігання, обробки та

аналізу даних, забезпечуючи високу якість та доступність даних для інших команд проєкту, зокрема для наукового відділу, адже їм потрібні дані для тренування моделей штучного інтелекту.

- Марія Ковальчук - Big Data Technical Leader. Веде команду, що працює з великими обсягами даних, оптимізує процеси обробки даних.
- Іван Петренко - Data Engineer. Займається структуруванням, обробкою та аналізом даних, важливих для проєкту.

Команда машинного навчання (ML)

Команда машинного навчання відповідає за розробку, впровадження та оптимізацію алгоритмів штучного інтелекту. Вони працюють над моделями, що можуть аналізувати біомедичні дані, передбачати потенційні лікарські мішені та засоби, а також покращувати точність діагностичних інструментів.

- Юлія Ткаченко - Senior Bioinformatics Engineer. Спеціалізується на біоінформатичному аналізі, зокрема у геномній інженерії та протеоміці.
- Олександр Бондар, Яна Шевченко, Андрій Василенко - Machine Learning Engineers. Розробляють і впроваджують алгоритми машинного навчання, які допомагають в аналізі біомедичних даних.

Веб-команда (Web)

Веб-команда відповідає за розробку інтерфейсу користувача, а також за бекенд і серверну частину платформи. Вони забезпечують, що платформа є функціональною, зручною та візуально привабливою для кінцевих користувачів, а також займаються інтеграцією всіх компонентів проєкту в єдине ціле.

- Василь Мельник - Fullstack Web TechLead. Відповідає за розробку веб-інтерфейсу і взаємодію переднього та заднього кінця платформи.
- Наталія Горобець - UI/UX Designer. Розробляє інтерфейси користувача, забезпечуючи їх зручність та естетичність.
- Дмитро Олійник, Катерина Біленька - Frontend Developers. Створюють клієнтську частину веб-додатків.

- Тарас Харченко - Back-end Developer. Розробляє серверну частину платформи, забезпечуючи обробку даних та логіку додатку.
- Олена Литвиненко - DevOps. Відповідає за оптимізацію робочих процесів, налаштування серверів та підтримку інфраструктури.

Ця організаційна структура (Рис.3.1., Додаток Г) забезпечує ефективне розподілення завдань та відповідальності серед учасників проєкту, сприяє координації та злагодженій взаємодії між різними підрозділами, що є критично важливим для досягнення цілей проєкту у встановлені терміни.

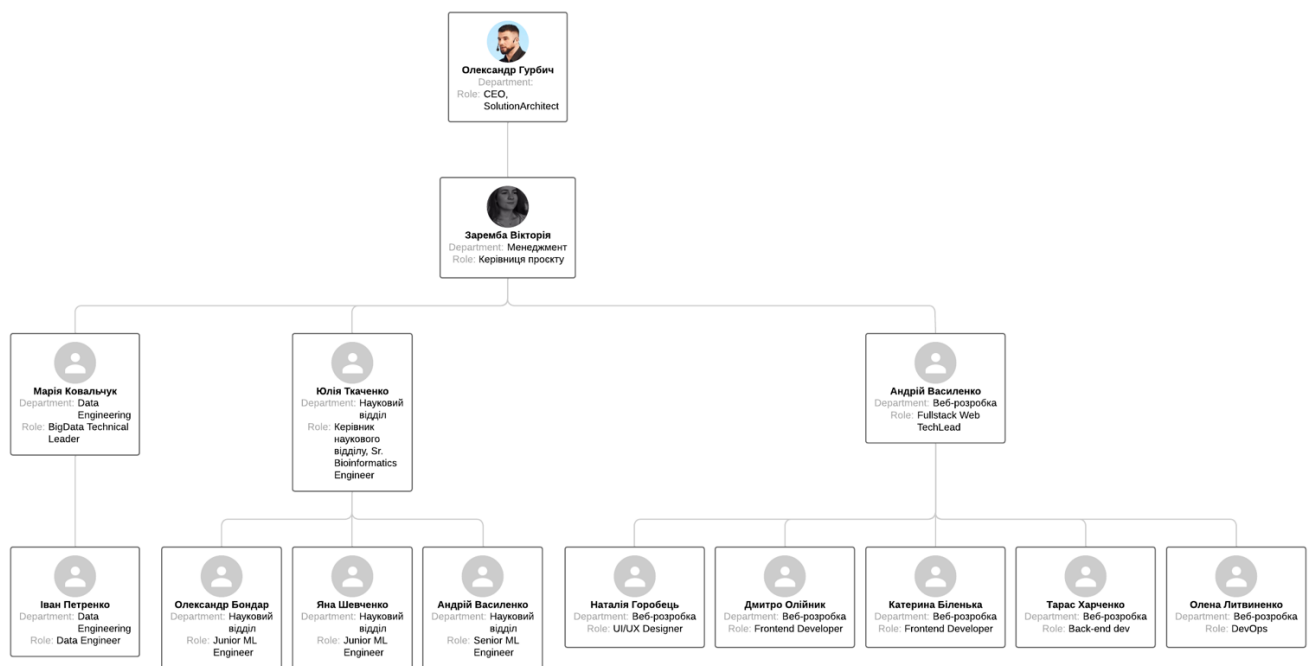


Рис.3.1. Організаційна структура проєкту

3.2. Розробка матриці відповідальності

Маючи сформовано команду проєкту, та розроблено організаційну структуру, важливо правильно та якісно розподілити задачі між членами команди. Чітке розподілення відповідальності усіх учасників проєкту може бути основним аспектом успішності.

Матриця відповідальності RACI є інструментом, який забезпечує систематизацію розподілу завдань і ролей між членами команди, що сприяє

оптимізації робочих процесів та підвищенню ефективності проєктного менеджменту.

Розробка матриці відповідальності RACI

Матриця відповідальності RACI (Responsible, Accountable, Consulted, Informed) використовується для визначення і документування ролей та відповідальностей за кожне завдання проєкту. Основні елементи матриці включають:

- **Responsible (Виконавець):** Особа або група, яка безпосередньо виконує завдання.
- **Accountable (Відповідальний):** Особа, яка несе кінцеву відповідальність за завершення завдання та повинна схвалювати його результати.
- **Consulted (Консультовані):** Особи або групи, чії знання та досвід використовуються для виконання завдань (залучені до обговорення).
- **Informed (Інформовані):** Особи або групи, які потребують інформації про хід та результати завдання.

Застосування матриці RACI в проєкті ВТАІ забезпечило чітке розуміння розподілу ролей та обов'язків, зменшило конфлікти та непорозуміння у командах, сприяло більшій залученості учасників та ефективному вирішенню проблем.

Процес розробки матриці RACI включав наступні кроки:

1. **Ідентифікація завдань:** Визначення всіх ключових завдань проєкту, які потребують виконання.
2. **Класифікація завдань та ролей:** Розподіл завдань за категоріями та присвоєння відповідних ролей учасникам.
3. **Консультації з командами:** Обговорення проєктних завдань з кожним відділом та уточнення їх ролей та відповідальностей.
4. **Документування матриці:** Формування офіційного документа, який фіксує відповідальності за всі завдання.

Таблиця 3.1. RACI Матриця відповідальностей

Завдання	Project Manager	Solution Architect	Data Engineer	BD Tech Leader	ML Engineer	Sr. Bioinf. Engineer	UI/UX Designer	Fullstack Web	Frontend Developer	DevOps
1	2	3	4	5	6	7	8	9	10	11
Визначення вимог проєкту	A	R	C	C	C	C	C	C	C	C
Розробка архітектури системи	C	A			R	R		R		
Розробка алгоритмів ML	I	C			A	R				
Інтеграція баз даних	C	R	R	A	C	R		R		
Розробка UI/UX	I	C					A	R	R	
Розробка фронтенду	I	C					R	A	A	
Налаштування серверного середовища	I	R						R		A
Тестування та оптимізація	I	C	C	C	C	C	C	A	A	C
Забезпечення безпеки даних	I	R	R	A		R				R
Підготовка документації	A	C	C	C	C	R	C	C	C	C
Комунікація зі стейкхолдерами	A	R								
Розробка бекенду	I	R						A		
Розгортання інфраструктури	C	A								R

Закінчення таблиці 3.1.

1	2	3	4	5	6	7	8	9	10	11
Моніторинг та логування	I	R				R				A
Оцінка проєктних ризиків	A	R		R	R	R				
Розробка детального проєктного плану	A	R	C	C	C	C	C	R	R	C
Проведення ринкових досліджень	R	A					C	C	C	
Планування ресурсів проєкту	A	R	R	R	R	R		R	C	C

3.3. Управління зацікавленими сторонами

3.3.1. Класичний підхід до управління зацікавленими сторонами

Управління зацікавленими сторонами є невід'ємною частиною успішного управління проєктом, оскільки залучення та збалансоване управління інтересами різних груп впливає на прийняття рішень та, в кінцевому підсумку, на успішне завершення проєкту.

Цей процес передбачає ідентифікацію усіх стейкхолдерів, аналіз їхніх інтересів, впливу та очікувань, а також розробку та втілення стратегій для ефективної взаємодії з ними на протязі всього проєкту.

Реєстр зацікавлених сторін, представлений у формі таблиці 3.2., був складений з метою надання вичерпної інформації про кожного учасника проєкту. Він містить відомості про ім'я, прізвище, компанію та займану посаду, а також ретельно описує ролі кожного стейкхолдера в проєкті, їх контактні дані та особливості, пов'язані з відповідальністю та впливом на проєкт. Важливість даного документу полягає у можливості ідентифікувати та аналізувати потенційні впливи та очікування кожного стейкхолдера, що дозволяє врахувати їх при плануванні та реалізації проєктних завдань.

Таблиця 3.2. Реєстр стейкхолдерів

ПІБ	Посада	Роль в проєкті	Контактні дані	Опис	Вплив	Зацікавленість
1	2	3	4	5	6	7
James Wilson	VP of Chemistry, BTAI	Project Sponsor	jwilson@btai.com	ініціатор проєкту. вклад у фінансуванні проєкту, зацікавлений в отриманні високого рівня прибутку	90%	100%
Sarah Brown	VP of Strategy and Early Discovery, BTAI	Project consultant	sbrown@btai.com	Консультує проєкт. Також оцінює успішність проєкту	80%	80%
Robert Smith	VP of Therapeutics, BTAI	Project consultant	rsmith@btai.com	Консультує проєкт. Також оцінює успішність проєкту	80%	80%
Юрій Коваленко	CEO / ChemSpace	науковий консультант та партнер	yurii.kovalenko@chsp.com	Залучений до написання та публікацій наукових статей, що стосуються теми проєкту. Іноді надає необхідні дані.	20%	30%

Закінчення таблиці 3.2

1	2	3	4	5	6	7
Заремба Вікторія	blackthorn.ai	PM	v.zaremba@b lackthorn.ai	відповідальна за управління проектом	70%	70%
Олександр Гурбич	blackthorn.ai	Solution Architect, SEO	ohur@blackt horn.ai	Відповідальний за технічний успіх проекту та розробку архітектурних і наукових рішень	100%	100%
Команда веб- розробки	blackthorn.ai				40%	40%
Команда ML- розробки	blackthorn.ai				40%	40%
DE Команда	blackthorn.ai				40%	30%
Користувачі				використання продукту	40%	80%

Крім того, реєстр включає кількісну оцінку зацікавленості та впливу стейкхолдерів у процентах, що допомагає визначити, які стейкхолдери потребують більше уваги та ресурсів з боку керівництва проєкту. Ця інформація дозволяє ефективно розподілити зусилля команди проєкту для забезпечення максимальної підтримки та залученості тих стейкхолдерів, які мають значний вплив на рішення або можуть бути ключовими для досягнення проєктних цілей.

Такий підхід забезпечує можливість більш ефективно управляти взаємодією з усіма зацікавленими сторонами, оптимізувати комунікацію та співпрацю між ними, а також знаходити баланс між різними інтересами для досягнення загальних цілей проєкту.

Далі на основі інформації про вплив та зацікавленість розроблено матрицю стейкхолдерів (Рис.3.2.). В матриці сторони умовно поділені на 4 групи залежно від рівня впливу та зацікавленості. Для кожної з груп є різна стратегія роботи з ними.

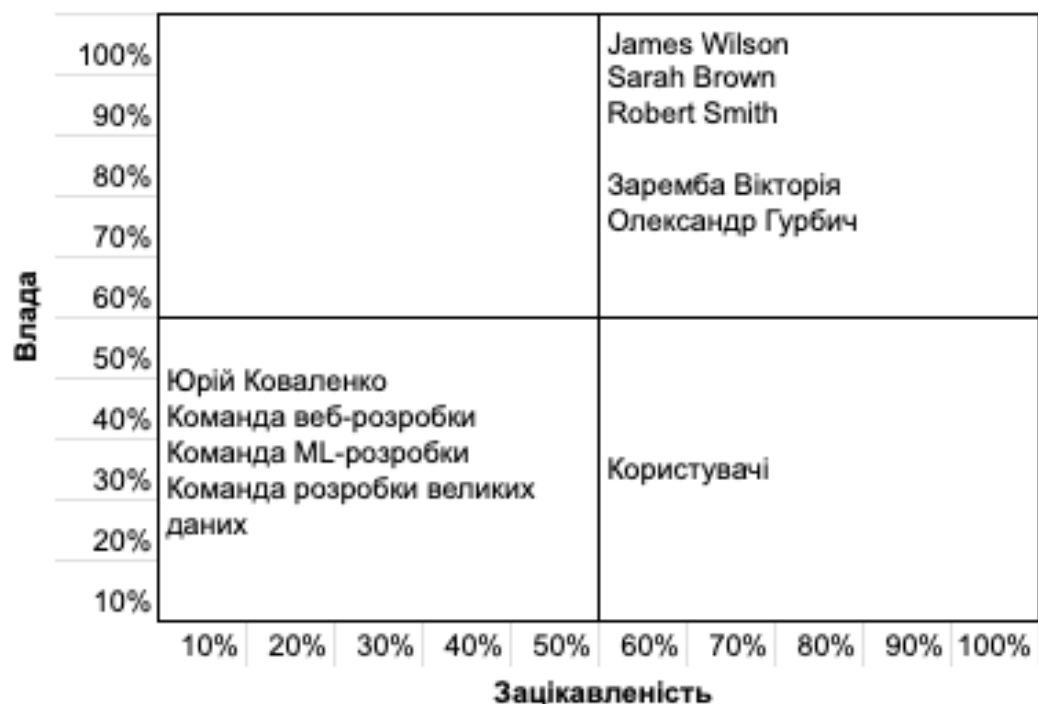


Рис.3.2. Матриця стейкхолдерів

Згідно матриці стейкхолдерів видно, що команда виконавців має помірний інтерес до проєкту і обмежену владу, ми можемо застосувати стратегію інформування. Також в цьому секторі є партнер проєкту Юрій Мороз, який іноді допомагає, але немає великої зацікавленості чи влади на проєкт. Тому до нього також буде застосована стратегія інформування. Вища ланка керівництва, а також інвестори і спонсори мають найбільший інтерес до проєкту і значні повноваження, тож їхнє активне залучення буде позитивно впливати на розвиток проєкту.

3.3.2. Застосування теорії ігор в управлінні стейкхолдерами

У сфері управління проєктами теорія ігор надає потужний інструментарій для аналізу взаємодій між стейкхолдерами, зокрема у контексті вирішення конфліктів, прийняття рішень та оптимізації співпраці. Використання цієї теорії дозволяє проєктній команді ВТАІ моделювати стратегічні взаємодії з ключовими стейкхолдерами, визначати потенційні коаліції, передбачати реакції та формувати оптимальні тактики для досягнення проєктних цілей.

Надалі буде розглянуто конкретну ситуацію яка була на проєкті.

Проєкт розробляється з застосуванням гнучких методології, а отже на різних етапах розробки можлива зміна або вибір пріоритетів. Перед плануванням нового спринта з розробки машинного навчання основні стейкхолдери мали різні пріоритети щодо того, яку частину функціоналу необхідно брати в роботу:

- James Wilson (VP of Chemistry, ВТАІ) (s_1) - хоче розвивати функціонал для підвищення точності прогнозів, оскільки це може забезпечити високу конкурентоспроможність продукту та відповідає науковим цілям компанії (f_1).
- Robert Smith (VP of Therapeutics, ВТАІ) (s_2) - зацікавлений у початку роботи над розробкою терапій, оскільки це відповідає стратегічним цілям його підрозділку компанії (f_2).

- Юрій Коваленко (науковий консультант та партнер, CEO / ChemSpace) (s₃) - бачить необхідність у роботі над генерацією датасету з новими drug-like молекулами, адже це буде вагомий внесок в науку (f₃).

Ця ситуація є класичним випадком застосування теорії ігор, де кожен стейкхолдер має свої стратегічні інтереси. Ми можемо моделювати цю ситуацію, застосовуючи ядро гри для визначення спільних інтересів і розробки стратегії, яка буде оптимальною для всіх учасників.

Модель ядра гри дозволяє кожному стейкхолдеру висловити свої переваги та оцінити потенційні вигоди від кожного з можливих рішень. Наприклад, можливо створити систему оцінки для кожного з запропонованих функціоналів, де бали присвоюються на основі важливості для кожного стейкхолдера. Після збору даних за допомогою групових дискусій або опитувань можна визначити, які функціонали забезпечать "виграш" для всіх стейкхолдерів, тобто опиняться в ядрі гри.

Нехай S - це набір стейкхолдерів $S = \{s_1, s_2, \dots, s_n\}$, а F - множина можливих функціоналів для розробки на проєкті $F = \{f_1, f_2, \dots, f_n\}$.

Кожен стейкхолдер s_i має вектор пріоритетів P_i щодо кожного функціоналу, який визначається як $P_i = \{p_{1i}, p_{2i}, \dots, p_{ni}\}$, де p_{ij} - пріоритет стейкхолдера s_i для функціоналу f_j .

Можливі вигоди від реалізації кожного функціоналу для кожного стейкхолдера можуть бути представлені як утиліта $U(s_i, f_j)$, яка зазвичай вимірюється як кількість ресурсів, часу, або інших вимірюваних факторів, що стейкхолдер готовий витратити для реалізації f_j .

Задача полягає у визначенні коаліції $C \subseteq F$ функціоналів так, щоб максимізувати загальну утиліту всіх стейкхолдерів:

$$U(C) = \sum_{s_i \in S} \sum_{f_i \in C} U(s_i, f_j) \quad (1)$$

Цілі ігри визначаються як досягнення найбільшої сумарної утиліти для всіх стейкхолдерів, тобто максимізація $U(C)$.

Ядро гри $Core(F)$ визначається як набір коаліцій функціоналів, що не можуть бути покращені шляхом перерозподілу:

$$Core(F) = \{C \subseteq F \mid \nexists C' \subset F \text{ таке, що } U(C') > U(C)\} \quad (2)$$

Робота з ядром гри полягає у визначенні оптимального розподілу ресурсів і уваги серед функціоналів так, щоб жоден стейкхолдер не міг отримати більшу вигоду, переключившись на іншу коаліцію функціоналів.

Для ситуації з пріоритезацією, якщо ми маємо три стейкхолдери s_1, s_2, s_3 та два функціонали f_1, f_2 , то можлива утиліта може бути представлена так:

$$U(s_1, f_1) = 10, U(s_1, f_2) = 4 \quad (3)$$

$$U(s_2, f_1) = 5, U(s_2, f_2) = 8 \quad (4)$$

$$U(s_3, f_1) = 6, U(s_3, f_2) = 7 \quad (5)$$

Загальна утиліта для коаліції тільки з f_1 буде $U(\{f_1\}) = U(s_1, f_1) + U(s_2, f_1) + U(s_3, f_1) = 21$, а для коаліції тільки з f_2 буде $U(\{f_2\}) = U(s_1, f_2) + U(s_2, f_2) + U(s_3, f_2) = 19$.

Таким чином, вибір f_1 як пріоритетного функціоналу є оптимальним у контексті цього прикладу.

3.4. Створення комунікаційного плану проєкту

Комунікаційний план виступає як фундамент для забезпечення ефективного обміну інформацією між усіма учасниками проєкту. Він допомагає координувати дії команди, інформує зацікавлені сторони про хід реалізації проєкту та забезпечує зворотний зв'язок, що є критично важливим для успішного завершення проєкту. У розділі описано процес розробки та впровадження комунікаційного плану для проєкту ВТАІ.

3.4.1. Визначення основних цілей комунікації:

- Забезпечення прозорості та відкритості всіх аспектів проєкту.
- Своєчасне інформування стейкхолдерів про прогрес і зміни в проєкті.
- Забезпечення двосторонньої комунікації для ефективної взаємодії.

3.4.2. Відповідальні особи:

- Проектний менеджер (РМ) – координатор всіх комунікаційних заходів.
- Члени команди – відповідальні за комунікацію в рамках своїх обов'язків.

3.4.3. Визначення груп стейкхолдерів та їх інтересів

Основним завданням комунікаційного плану є гарантування, що всі зацікавлені сторони мають доступ до потрібної інформації в потрібний час. Особливу увагу в проєкті ВТАІ приділено різним групам стейкхолдерів, що мають різні інформаційні потреби.

- Перша група — керівництво компанії ВТАІ, прямі замовники проєкту. Ця група включає керівників відділів компанії, а також вони є ключовими спонсорами та консультантами проєкту. Вони зацікавлені у стратегічних звітах, що відображають загальний прогрес, фінансові вкладення і потенційні ризики. Для цієї групи плануються щотижневі демо-сесії та щомісячні звіти, що надають детальний аналіз стану проєкту.

- Друга група включає зовнішніх консультантів і партнерів, які можуть надавати спеціалізовані послуги чи знання, Їх залученість має консультативний характер, оскільки їх рівень зацікавленості та вплив на проєкт є відносно низьким. Ці стейкхолдери втручаються у проєктні процеси лише за необхідності. Щоб забезпечити їхню інформованість про хід виконання проєкту та здатність своєчасно реагувати на запити консультації, їм надсилаються щомісячні письмові звіти. Це дозволяє підтримувати постійний зв'язок та інформаційну взаємодію з цією групою стейкхолдерів.

- Третя група включає команду проєкту — розробників, інженерів-біоінформатиків та інших технічних спеціалістів, які працюють безпосередньо над реалізацією проєкту. Їхні щоденні комунікації включають стендапи, технічні обговорення та демонстрації розроблених функціональностей. Крім того, розроблено систему керування завданнями для слідкування за прогресом і вирішенням проблем в системі Jira, що дозволяє забезпечувати своєчасне обговорення та вирішення задач.

Кожна з цих груп має визначені канали та частоту комунікацій, що забезпечує ефективність взаємодії та підтримує загальну координацію проєктних зусиль. Враховуючи різноманітність зацікавлених сторін і широкий

спектр інформаційних потреб, створений комунікаційний план є ключем до успішної реалізації проєкту ВТАІ.

3.4.4. Способи, частота, та інші деталі комунікації

Для досягнення максимальної прозорості та чистоти комунікації важливо дотримуватися певних стандартів та практик. Електронні листи повинні бути структурованими та містити важливу інформацію у вступі. Щотижневі зустрічі та щомісячні звіти вимагають складеного наперед розкладу. Системи управління проєктами, такі як Jira, та документаційні системи, як Notion, забезпечують централізоване сховище інформації, що є критично важливим для збереження історії проєкту та знань.

Для того, щоб вести якісну комунікацію, необхідно задокументувати всі комунікаційні процеси. На початку проєкту ВТАІ було розроблено окремий документ, який відображає всю необхідну інформацію. Тут міститься інформація про часову зону зацікавлених сторін, очікуваний час відповіді на повідомлення, робочі години які перетинаються для усіх стейкхолдерів, частота доступності для мітингів, основні джерела комунікації, бажана частота репорту. Відповідний документ є в таблиці 3.3.

Таблиця 3.3. Комунікаційний план

Часова зона стейкхолдерів	GMT-7
Очікуваний час відповіді на повідомлення	24 години
Робочі години які перетинаються для усіх стейкхолдерів	18:00 - 21:00 (Київ) = 08:00 – 11:00 (Тусон, Арізона)
Частота доступності замовників для мітингів (щоденно, щотижнево...)	Двічі на тиждень
Основні джерела комунікації	Microsoft Teams, Emails
Бажана частота репорту	Щотижневий демо-мітинг, письмовий репорт раз на місяць

Оскільки визначено кілька груп зацікавлених сторін, то для них також будуть різна частота джерела та мета комунікації. Тому з початком активної роботи над проектом необхідно також розробити розширений комунікаційний план. Він включає перелік стейкхолдерів, їх роль у проекті, тип та частоту комунікації, обрані методи та інструменти для обміну інформацією та специфічні цілі кожного типу комунікації. (Таблиця 3.4.).

Таблиця 3.4. Розширений комунікаційний план

Група зацікавлених сторін	Тип комунікації	Частота	Метод/Інструмент	Мета
1	2	3	4	5
Керівництво компанії ВТАІ (замовники)	Стратегічні звіти	Щомісячно, 15 число	Електронні листи	Інформування про загальний прогрес, фінансові показники та потенційні ризики
Керівництво компанії ВТАІ (замовники)	Демо-сесія ML частини	Раз в 2 тижні (парні тижні)	Презентація, Microsoft Teams	Інформування про результати експериментів за 2 тижні.
Керівництво компанії ВТАІ (замовники)	Демо-сесія веб та DE частини	Раз в 2 тижні (НЕ парні тижні)	Презентація, Microsoft Teams	Інформування про результати експериментів за спринт.
Зовнішні консультанти та партнери	Місячний звіт	Щомісячні звіти	Електронні листи	Підтримання обізнаності з ходом проекту

1	2	3	4	5
Зовнішні консультанти та партнери	Онлайн-консультації	За запитом	Microsoft Teams	Консультавання щодо наукової частини проєкту
Команда проєкту	Щоденні дейлі мітинги	Щодня	Jira, Microsoft Teams	Координація щоденних завдань та вирішення проблем. Для 3 команд розробників мітинги проводяться окремо.
Команда проєкту	Письмова комунікація щодо задач	Щодня	Jira, чат в Microsoft Teams	Координація щоденних завдань та вирішення проблем. Для 3 команд розробників є 1 спільний на 3 окремі чати.

3.5. Управління ризиками проєкту

Проєкти з розробки штучного інтелекту завжди пов'язані з надзвичайно високим рівнем невизначеності, а це несе за собою багато ризиків. Тому критично важливо починати процеси управління ризиками ще навіть на етапі перед продажів, обдумуючи можливі ризики та попереджаючи потенційних замовників про них.

Процес управління ризиками на проєкті розробки платформи складався з усіх основних етапів: ідентифікація ризиків, оцінка ризиків, розробка трьох стратегій реагування на ризиків, та моніторинг і контроль ризиків.

На етапі ідентифікації усі можливі ризики поділили на 5 категорій: програмні ризики, апаратні ризики, внутрішні ризики, зовнішні ризики, там форс-мажори. До кожної з цих категорій віднесли від трьох до шести можливих ризиків.

Далі для кожного з ризиків була проведена якісна та квазікількісна оцінка для п'яти параметрів: сила впливу, керованість, ймовірність, затримки у часі, фінансові втрати. Також оцінений загальний показник важливості ризику, який є середнім арифметичним квазікількісних оцінок всіх параметрів.

Шкала квазікількісної оцінки виглядає наступним чином:

Сила впливу:

- 1 - Мінімальний вплив
- 2 - Низький вплив
- 3 - Середній вплив
- 4 - Високий вплив
- 5 - Дуже високий вплив

Затримки у часі:

- 1 - Мінімальна затримка (1-2 дні)
- 2 - Незначна затримка (3-5 днів)
- 3 - Помірна затримка (6-10 днів)
- 4 - Значуща затримка (11-20 днів)
- 5 - Критична затримка (більше 20 днів)

Керованість ризику:

- 1 - Легко керований
- 2 - Можливий, потребує зусиль
- 3 - Середньо важко керований
- 4 - Важко керований
- 5 – Майже неможливо керований

Фінансові втрати:

- 1 - Мінімальні фінансові втрати
- 2 - Незначні фінансові втрати
- 3 - Помірні фінансові втрати
- 4 - Значущі фінансові втрати
- 5 - Критичні фінансові втрати

Ймовірність ризику:

- 1 - Дуже низька ймовірність
- 2 - Низька ймовірність
- 3 - Середня ймовірність
- 4 - Висока ймовірність
- 5 - Дуже висока ймовірність

У таблиці 3.5. представлено кількісну оцінку ризиків.

Таблиця 3.5. Кількісна оцінка ризиків

№	Тип	Ризикова подія	Сила Впливу	Керова ність	Ймовір ність	Затрим ки	Фін. втрати	Важли вість
1	2	3	4	5	6	7	8	9
1	Програмні ризики	Затримка у розробці програмного забезпечення	4	3	3	3	2	3
2		Ризик несумісності з різними браузерами	3	4	4	3	4	3,6
3		Помилки в коді, що призведуть до неправильної роботи	4	2	3	2	3	2,8
4		Проблеми з інтеграцією сторонніх сервісів	3	3	3	3	3	3
5		Проблеми в обробці великих обсягів даних	5	1	4	3	2	3
6	Апаратні ризики	Проблеми з забезпеченням необхідної обчислювальної потужності	5	1	2	5	4	3,4
7		Відмова серверів або мережевих з'єднань, що може призвести до недоступності сайту для користувачів	4	2	3	1	3	2,6
8		Недостатня масштабованість серверів	3	4	3	3	3	3,2
9		Збої в мережевому обладнанні	3	3	3	1	1	2,2
10		Відсутність необхідних ресурсів для масштабування сайту під зростаюче навантаження	4	3	3	4	4	3,6

Закінчення таблиці 3.5

1	2	3	4	5	6	7	8	9
11		Неможливість масштабувати інфраструктуру платформи під нові вимоги	3	3	4	4	4	3,6
12	Внутрішні ризики	Недостатній досвід команди розробників	4	3	3	3	3	3,2
13		Втрата ключових розробників	4	2	2	5	3	3,2
14		Ризик конфліктів в команді	3	3	2	1	2	2,2
15		Відсутність ефективної моделі монетизації сайту	4	3	4	5	5	4,2
16		Незацікавленість інвесторів	4	2	3	5	5	3,8
17		Ризик зміни вимог та обсягу	4	2	4	3	3	3,2
18		Зовнішні ризики	Ризик змін у законодавстві про захист особистих даних	4	2	2	4	1
19	Ризик високого рівня конкуренції на ринку		3	4	4	1	4	3,2
20	Можливість хакерських атак		4	3	3	2	2	2,8
21	Форс-мажори	Природні катастрофи	4	5	1	2	4	3,2
22		Пандемія або епідемія	4	4	1	2	5	3,2
23		Мобілізація членів команди	3	5	3	5	3	3,8

Наступним етапом в управлінні ризиками є розробка проти ризикових заходів:

- Проти ризикові заходи (ПРЗ) 1 - Профілактика
- Проти ризикові заходи (ПРЗ) 2 - при симптомі
- Проти ризикові заходи (ПРЗ) 3 - при проблемі

Далі буде розглянути приклад ПРЗ для ризику змін вимог та обсягу

Ризик:

Ризик зміни вимог та обсягу

ПРЗ 1 (Профілактика):

- Запровадження гнучких методологій розробки
- Ретельний аналіз вимог на різних рівнях (бізнес-рівень, рівень користувача, функціональний рівень)

- Детально прописати в договорі що команда робить і що НЕ робить

• Прописати в договорі кількість ітерацій правок які входять в бюджет. наступні правки за додаткову оплату. Забезпечити наявність ресурсів на цей випадок

- З новими замовниками працювати за time&material контрактом
- Постійний моніторинг і аналіз змін
- Регулярна ревізія плану проєкту

Симптоми (рання ознака):

- Стейкхолдери не можуть між собою домовитись щодо вимог
- Постійні зміни в вимогах замовника або збільшення обсягу робіт

ПРЗ 2 (при симптомі):

- Постійний моніторинг та аналіз змін
- Збереження структурованих комунікацій з замовником та регулярна перевірка вимог.

• Для кожної зміни чи додаткового запиту створювати change request та погоджувати з замовником

- Ведення change log проєкту

ПРЗ 3 (при проблемі):

- Аналіз можливостей прискореного виконання та внесення додаткових ресурсів у проєкт
- Виокремлення окремого проєкту для врахування усіх змін
- Підписання нового договору

3.6. Використання платформи Jira для управління проєктом

Для реалізації для забезпечення ефективної координації, планування та контролю за реалізації завдань було використано платформу Jira. Вона служить як централізована інструмент, що дозволяє всім учасникам проєкту отримувати актуальну інформацію про статус з робіт, пріоритети завдань, і відслідковувати прогрес роботи.

Робота з Jira починається зі створення проєкту в системі, де кожен елемент проєкту визначений як завдання, історія користувача чи помилка. Кожен такий елемент супроводжується докладним описом, який включає в себе вимоги, критерії прийняття, пріоритетність, відповідальних осіб та очікувані терміни. Така деталізація забезпечує кожному учаснику проєкту чітке розуміння своїх обов'язків і сприяє належному виконанню завдань.

Важливим аспектом використання Jira є налаштування робочого процесу (workflow), що максимально відповідає потребам проєкту ВТАІ (Рис.3.3). Окрім звичайних статусів для проєкту додано також кілька які забезпечують потреби різних команд. Наприклад задачі по або спочатку переходять в статус перевірки тех-лідом, потім в статус готовності до релізу, і лише після релізу закриваються. Задачі по машинному навчання часто потребують етапу тренування моделі, бо так задача одночасно і в прогресі, але розробник при цьому не залучений оскільки тренування моделі машинного навчання може тривати від кількох годин до кількох тижнів. Тому для таких задач створено окремий статус “training”.



Рис. 3.2. Робочий процес проекту ВТАІ

Наступним етапом є створення задач. Для зручності та ефективності використання необхідно всі задачі створювати за однаковим шаблоном: аббревіатура напрямку, дефіс, назва задачі з великої букви. Усі задачі повинні мати заповненими обов'язкові поля, такі як опис естимейти, батьківський епік, відповідальний, тощо (Рис. 3.3.).

Backlog (132 issues)		461	0	Create sprint
HTG-691 API - Target discovery - Implement DELETE endpoint for pathway analysis	TARGET DISCOVERY	OPEN	5	[Avatar]
HTG-692 API - Target discovery - Implement GET endpoints for pathway analysis	TARGET DISCOVERY	OPEN	5	[Avatar]
HTG-867 DE - Create DAG to load HGNC to BQ	HGNC	OPEN	8	[Avatar]
HTG-863 DE - Create Terraform Dataset and tables for HGNC	HGNC	OPEN	8	[Avatar]
HTG-138 UI - General - Implement toggle dark light mode	GENERAL	OPEN	13	[Avatar]
HTG-824 UI - Datasets - Integrate pagination with API	DATASETS	OPEN	5	[Avatar]
HTG-697 ML - Cardiotoxicity prediction of metabolites (from BioTranformer) using H2O	CARDIOTOXICITY	OPEN	5	[Avatar]
HTG-280 ML - select top-1000 and bottom-1000 molecules from ChemSpace - SIRT1	REAL WORLD VALIDA...	IN REVIEW	8	[Avatar]

Рис. 3.3. Наповнення беклогу проекту

Наступний етап – створення спринтів. Оскільки над проектом працює 12 людей, а це забагато для ефективних процесів SCRUM. Тут доцільно скористатись методологією Scrum Of Scrum[23] та розділити усю команду на 2 підкоманди: команда машинного навчання, команда веб розробки, та команда розробки великих даних. Також важливо зазначити, що команда машинного навчання працює за методологією Data Driven Scrum (DDL) – Скрам на основі даних, а отже спринт базується на можливостях (а не на календарних подіях, обмежених у часі) [6].

Ліга дозволяє мати відкритими одразу кілька спринтів паралельно, що ідеально підходить для проекту ВТАІ (рис. 3.4.).

Backlog

YZ Y +11
Version Epic More

- > HTG - ML Team Sprint 10 5 Jul – 19 Jul (30 issues) 0 93 8 Complete sprint
- > HTG - DE Team Sprint 11 12 Jul – 26 Jul (26 issues) 0 0 86 Complete sprint
- > HTG - WEB Team Sprint 11 14 Jul – 28 Jul (30 issues) 52 21 108 Complete sprint

Рис. 3.4. Відкриті спринти для різних команд паралельно

Приклад активного спринта зображено на рисунку 3.5.

Projects / HTG Molecular

HTG - WEB Team Sprint 11

0 days remaining Complete sprint

YZ Y +6
Add people
Epic
Label
Type
Sprint
Clear filters
GROUP BY None
Insights
View settings

OPEN 5 OF 7	IN PROGRESS	BLOCKED	IN REVIEW 2 OF 19	READY FOR RELEASE 3 OF 3	DONE 20 OF 57
<p>UI - General - Implement toggle dark light mode</p> <p>GENERAL UI</p> <p>HTG-138 13</p>			<p>API - Drug discovery - Implement molecular-simulation config file endpoint.</p> <p>DRUG DISCOVERY API</p> <p>HTG-922 13</p>	<p>API - Drug discovery - Implement GET endpoints for docking</p> <p>DRUG DISCOVERY Endpoint API</p> <p>HTG-684 5</p>	<p>API - Drug discovery - Implement DELETE endpoint for docking</p> <p>DRUG DISCOVERY Endpoint API</p> <p>HTG-689 5</p>
<p>UI - Drug discovery - Implement pdb file viewer</p> <p>DRUG DISCOVERY UI</p> <p>HTG-834 21</p>			<p>API - Drug discovery - Implement GET report endpoint for drug likeness</p> <p>DRUG DISCOVERY Endpoint API</p> <p>HTG-840 8</p>	<p>API - Update API documentation</p> <p>API</p> <p>HTG-910 3</p>	<p>API - Move GET /v1/drug-discovery/{dataset}/drug-likeness endpoint from drug generation to drug-likeness controller</p> <p>HTG-912 1</p>
				<p>API - Drug discovery - Implement DELETE endpoint for drug likeness</p> <p>DRUG DISCOVERY Endpoint API</p> <p>HTG-690 5</p>	<p>API - Drug discovery - Implement creation endpoint for docking</p> <p>DRUG DISCOVERY Endpoint API</p> <p>HTG-679 5</p>

Рис. 3.5. Приклад активного спринта для команди веб-розробки

РОЗДІЛ 4. РОЗРОБКА ІНФОРМАЦІЙНОГО ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПРОЄКТУ

4.1. Розробка архітектурного рішення

Створення архітектури сервісу є ключовою частиною будь-якого програмного проєкту. Вона повинна забезпечити не лише функціональність, ай ефективність, масштабованість та надійність системи.

Для виконання задач генерації ліків, що вимагають значної обчислювальної потужності, пропонується архітектурне рішення з використанням Google Cloud Function, активованої через API Gateway, як зображено на рисунку 4.1. Таке рішення забезпечує автономність та незалежність кожного завдання генерації, дозволяючи ефективно масштабувати систему.

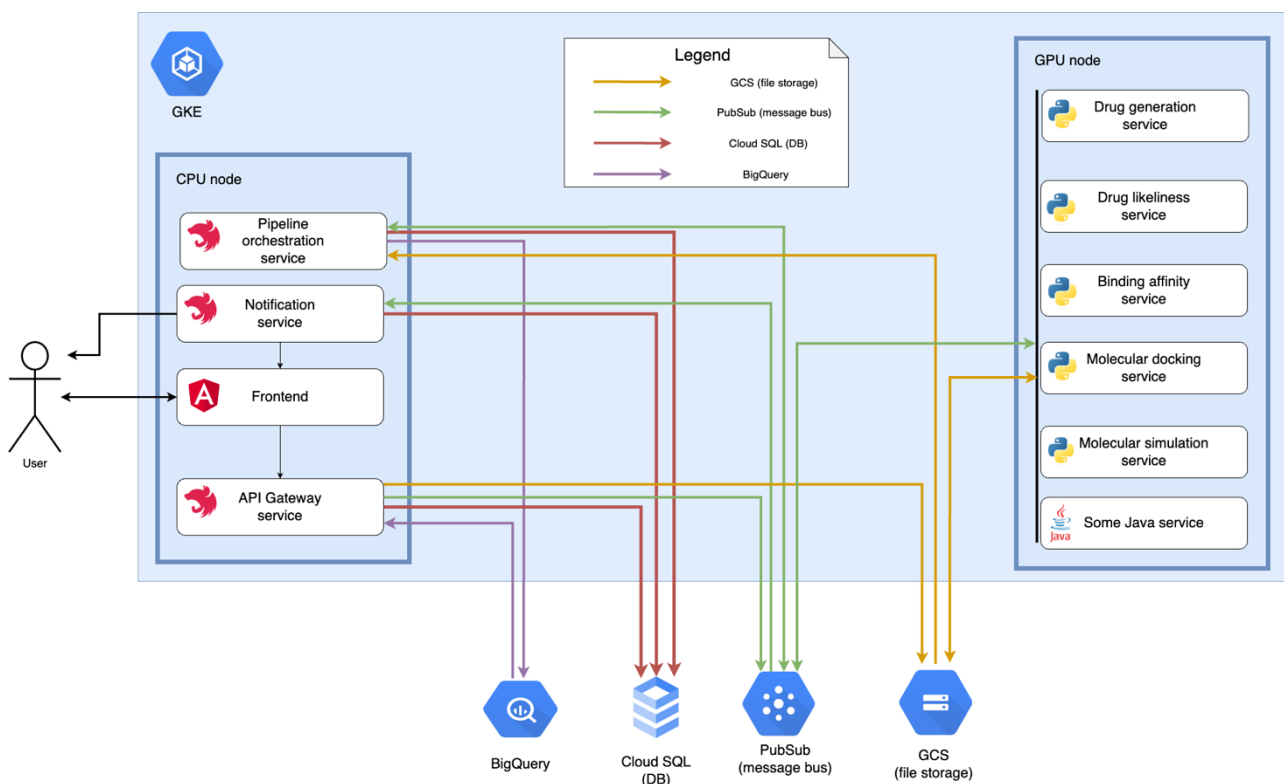


Рис.4.1. Архітектурна пропозиція для платформи VTAI

Окрема увага приділяється оркестрації процесів за допомогою Airflow, що дозволяє ефективно управляти робочими потоками і підвищує швидкість

розробки. Також обрано контейнеризацію і оркестрацію завдань через Kubernetes, що є оптимальним рішенням для керування великою інфраструктурою.

Всі застосунки повинні бути контейнеризовані та використовувати PubSub або інший брокер повідомлень для зв'язку, з обов'язковим доступом до баз даних чи Google Cloud Storage.

Запропонована архітектура роботи з великими даними вимагає створення відповідних схем баз даних для зберігання результатів генерації, а також розробки сповіщень для користувачів через електронну пошту чи веб-сокети. Це вимагає створення окремого сервісу, який би виконував цю роль.

Для реалізації вищезазначених планів потрібно буде виконати наступні завдання:

- Створення обгортки Python для моделей генерації ліків.
- Докеризація Python сервісу.
- Локальне розгортання конфігурації за допомогою Docker-compose та інших інструментів.
- Розробка API через API Gateway для ініціювання генерації.
- Розробка схеми бази даних для зберігання спроб генерації.
- Розгортання інфраструктури для обміну повідомленнями, Google Functions та Kubernetes за допомогою Terraform.

Комплексний підхід до розробки системи, який включає аналіз вимог, вибір технологій, безпеку, інфраструктуру та процеси розробки, є вирішальним для створення стабільного та ефективного програмного продукту.

4.2. Організація роботи з великими даними для генерації молекул

Робота з великими обсягами даних вимагає ґрунтовного планування та чіткої структури. В процесі розробки використовується обширний пул даних, який включає тренувальні набори даних, відфільтровані для навчання моделей, та публічні датасети. Для використання в проєкті застосовано близько 30 різних публічних датасетів типу «drug-target». Серед найважливіших:

- ChEMBL32 - база даних хімічних з'єднань, що містить велику колекцію даних про молекули з медичними властивостями. Вона зосереджена на з'єднаннях, що мають потенціал в якості лікарських засобів, та включає інформацію про їх біологічну активність, хімічну структуру, цільові білки та дані про ефективність та безпечність [24].

- PDBeBank - PDB Bank банк білкових 3D структур [25]

- ICD11 – Принципи міжнародних кодів хвороб. Ця інформація лише для нас, щоб зрозуміти, як вибирати екземпляри в діапазонах захворювань [26].

- GWAS - Дослідження геномної асоціації з SNP, хворобами та коефіцієнтами шансів. Необхідні для прогнозування зв'язків цільових захворювань та персоналізованої медицини [27].

Ключовим аспектом роботи з великими даними є створення та використання BigQuery Dataset через Terraform, реалізація PySpark Job для завантаження даних з джерела до BigQuery та реалізація DAG для завантаження ресурсів джерел даних. Ці інструменти та методика забезпечують автоматизацію та оптимізацію роботи з великими обсягами інформації, а також дозволяють команді швидко адаптуватися до мінливих потреб проєкту.

Застосування таких інструментів, як BigQuery та PySpark, дозволяє гнучко масштабувати інфраструктуру для задоволення потреб у роботі з даними та генерації молекул. Автоматизація забезпечується за допомогою Terraform, який впроваджує інфраструктуру як код, покращуючи тим самим ефективність розгортання та управління ресурсами.

Враховуючи розміри та складність даних у таких проєктах, як BTAI, необхідно розробити ретельно сплановану архітектуру даних, яка включає стратегічне розміщення баз даних та оптимізацію шляхів доступу. Все це вимагає не лише технічної експертизи, а й глибокого розуміння біоінформатики та геноміки.

Нижче наведений код для імплементації DAG Airflow для «ChEMBL32 General -> BigQuery».

```

-- START chembl_id_lookup
TRUNCATE TABLE {{resource_name('landing_chembl32',
divider='_')}}.chembl_id_lookup
;

INSERT INTO {{resource_name('landing_chembl32',
divider='_')}}.chembl_id_lookup (
  chembl_id,
  entity_type,
  entity_id,
  status,
  last_active
)
SELECT
  chembl_id,
  entity_type,
  CAST(entity_id AS INT64) AS entity_id,
  status,
  CAST(last_active AS INT64) AS last_active
FROM {{resource_name('staging_chembl32', divider='_')}}.chembl_id_lookup
;
-- END chembl_id_lookup

-- START version
TRUNCATE TABLE {{resource_name('landing_chembl32', divider='_')}}.version
;

INSERT INTO {{resource_name('landing_chembl32', divider='_')}}.version (
  name,
  creation_date,
  comments
)
SELECT
  name,
  creation_date,
  comments
FROM {{resource_name('staging_chembl32', divider='_')}}.version
;
-- END version

```

Також на рисунку 4.2. зображене графічне представлення DAG'у.

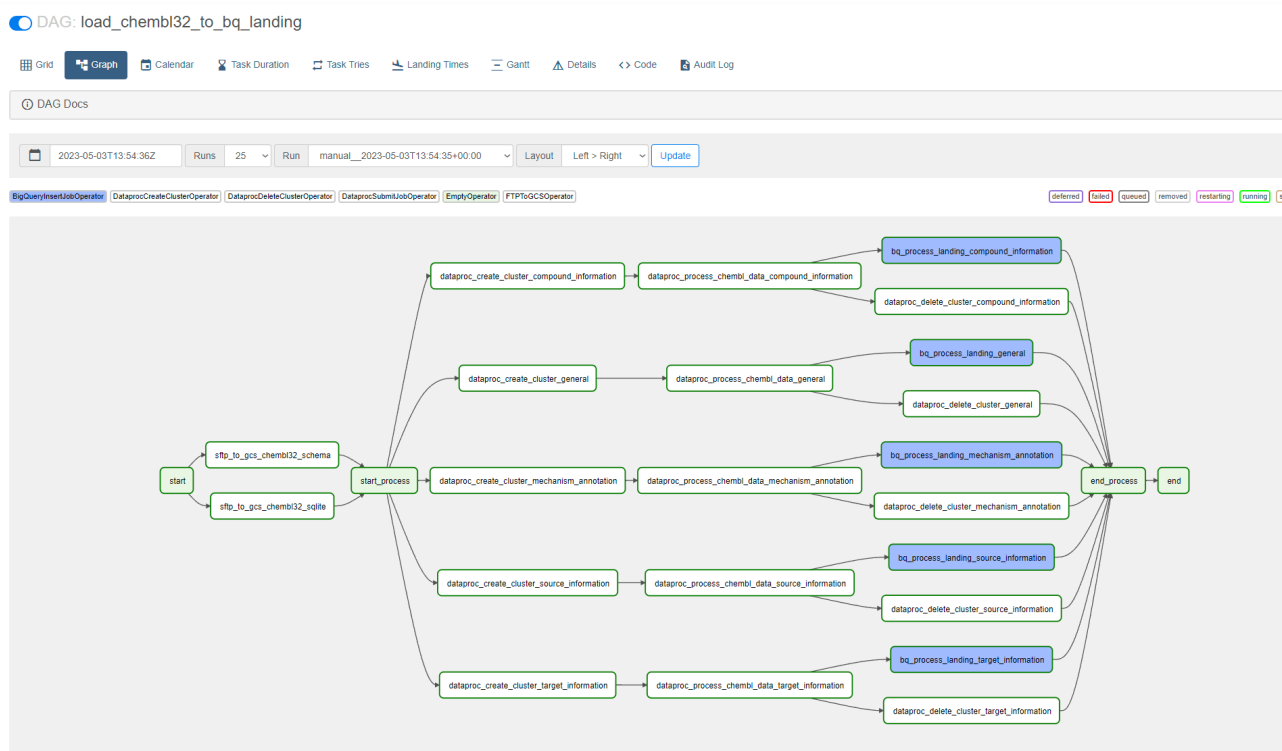


Рис.4.2. DAG Airflow для «Chembl32 General -> BigQuery»

4.3. Розробка інтерфейсів програмного забезпечення

Розробка інтерфейсів програмного забезпечення включає створення та документування всіх аспектів взаємодії користувачів з програмним забезпеченням проєкту. Основна мета полягає у забезпеченні зручності користування, інтуїтивно зрозумілого дизайну, та забезпеченні ефективного зв'язку між користувачем та системою.

Для створення ефективного та інтуїтивного користувацького інтерфейсу в платформі ВТАІ, усю систему можна розділити на шість основних потоків користувацького досвіду. Розглянемо кожен з них і як вони взаємодіють у межах платформи:

- **Auth (Аутентифікація).** Цей потік забезпечує процеси ідентифікації користувачів та їхньої авторизації в системі. Аутентифікація та реєстрація включають стандартний логін та пароль. Також забезпечено функціонал у випадку забутого паролю. На рисунку 4.3. зображено сторінку реєстрації нового користувача.

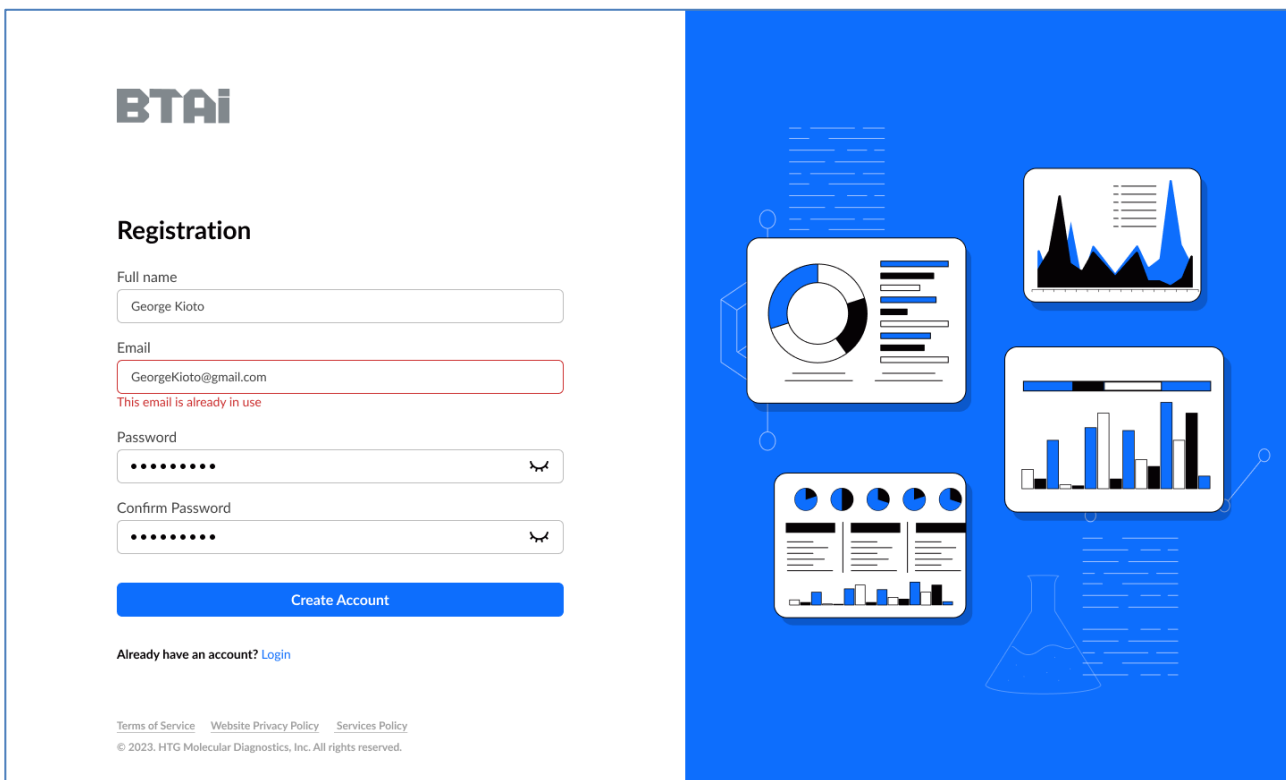


Рис. 4.3. Сторінка реєстрації користувача

- **User Profile (Профіль користувача).** Профіль користувача включає управління особистою інформацією, налаштуваннями системи та персоналізацію. Користувачі можуть змінювати свої особисті дані, налаштування приватності та інтерфейсу, що дозволяє адаптувати систему під індивідуальні потреби.

- **Target Discovery (Виявлення мішеней).** Цей потік спрямований на ідентифікацію молекулярних мішеней для нових лікарських засобів. Включає засоби для аналізу генетичних, білкових і клітинних шляхів, що можуть бути потенційними цілями для лікування. Користувач завантажує власний датасет, переходить на сторінку налаштування аналізу нового шляху. Далі переходить на сторінку усіх аналізів, після чого можна окремо переглянути звіт по кожному.

- **Drug Generation (Генерація лікарських засобів).** Цей потік відповідає за створення нових молекул з потенціалом стати лікарськими засобами. Цей потік включає додавання нового датасету на основі якого буде відбуватись генерація,

сторінка вибору типу конфігурації, налаштування обраної конфігурації та сторінку усіх генерацій. З сторінки генерацій є можливість перейти на звіт по ній. На рисунку 4.4. зображений дизайн для сторінки списку генерацій.

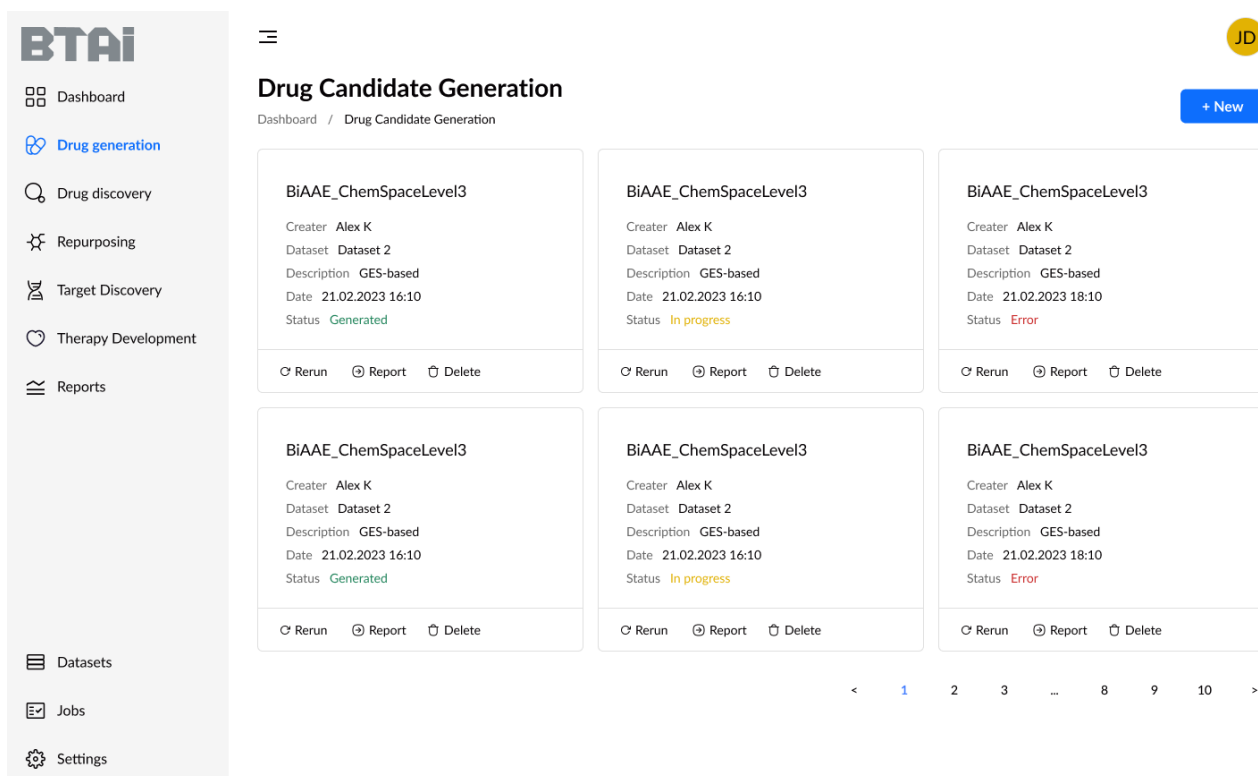


Рис. 4.4. Сторінка списку генерацій

- **Datasets (Набори даних).** У цьому потоці користувачі можуть управляти збереженими датасетами, завантажувати нові та проводити їх аналіз для подальшого використання у генерації та відкритті лікарських засобів.

- **Drug Discovery (Дослідження лікарських засобів).** Цей потік надає інструменти для детального дослідження згенерованих кандидатів. Можливо кілька різних видів аналізу та відповідних звітів: деталі датасету, лікарські властивості, молекулярний докінг, молекулярна симуляція та прогнозування спорідненості. Також імплементовано фільтрування молекул за різними параметрами. На рисунку 4.5. зображено приклад сторінки молекулярної симуляції.

Рис. 4.5. Налаштування молекулярної симуляції

4.4. Розробка програмного забезпечення реалізації Веб частини ІТ проекту

В рамках сучасних вимог до розробки високофункціонального та надійного програмного забезпечення, розробка веб-частини ІТ проекту вимагає злагодженої роботи команди, чіткої архітектури та вибору оптимальних технологічних рішень. Значення цих аспектів значно підвищується, коли йдеться про інтеграцію великих даних та їх застосування для розробки штучного інтелекту який застосовується для генерації молекул у фармацевтичному секторі.

Розробка веб-частини ІТ проєкту ВТАІ включає створення комплексної інфраструктури для обробки та аналізу великих даних з метою генерації молекул за допомогою штучного інтелекту. Цей процес охоплює не тільки розробку серверної частини але й клієнтського інтерфейсу. У зв'язку з цим, було обрано Angular CLI версії 15.2.6 для генерації проєкту, що дозволяє створювати оптимізовані та ефективні веб-додатки.

Для запуску проєкту ВТАІ необхідно встановити Node.js версії 18.16.0 і npm, що є стандартним менеджером пакетів для Node.js. Використання цих інструментів забезпечує надійне середовище для розробки веб-додатків. Проєкт реалізовано на фреймворку NestJS версії 9.0.0, відомому своєю ефективністю та підтримкою TypeScript, що сприяє підвищенню якості коду та полегшенню його підтримки. TypeORM версії 0.3.15 використовується для зручної роботи з базами даних, дозволяючи ефективно керувати даними в об'єктно-орієнтованому стилі. Ці інструменти та технології формують міцну основу для створення надійних та масштабованих веб-додатків, важливих для проєкту ВТАІ.

Використання цих технологій та інструментів сприяє розробці високопродуктивної веб-частини проєкту, що може ефективно масштабуватись та обробляти великі об'єми даних, необхідних для генерації молекул. Забезпечення сумісності та використання перевірених версій програмного забезпечення гарантує стабільність робочого процесу та високу якість кінцевого продукту.

Програмний код, який дозволяє взаємодіяти з ресурсами проєкту, розміщений у репозиторій на GitHub: окремо для API (бекенду) і окремо для UI (фронтенду). Це спрощує процес розробки, тестування та розгортання. Зокрема, реєстрація нових користувачів, авторизація, управління профілями, запуск задач з генерації молекул – всі ці функціональні можливості мають бути доступними через визначені API ендпоінти.

Для імплементації всього функціоналу проєкту було розроблено 33 API ендпоінти, які загалом можна погрупувати на 5 потоків:

- `/api/auth` - для автентифікації користувачів.
- `/api/profile` - для управління профілями користувачів.
- `/api/drug-generation` - для ініціації процесів генерації та дослідження молекул.
- `/api/datasets` - для доступу та керування наборами даних.
- `/api/target-discovery` - для автоматизованого виявлення та аналізу потенційних біомолекулярних мішеней.

Для кожного з цих API було розроблено докладні специфікації та документацію, яка описує методи HTTP, параметри запиту, формат відповідей та можливі помилки. Такий підхід дозволяє стандартизувати взаємодію з системою і надає чіткі інструкції для розробників. На рисунку 4.6. зображено приклад документації для ендпойнту який використовується для отримання оцінки "подібності до ліків" для вказаного набору даних.

```

List dataset drug likeness
GET /drug-discovery/{datasetId}/drug-likeness

Request ( query paramteres )
1 datasetId

Response
1 {
2   "id": "string",
3   "filters": [
4     {
5       "name": "string",
6       "type": "BOOL",
7       "value": {}
8     }
9   ],
10  "user": {
11    "id": "string",
12    "firstName": "string",
13    "lastName": "string",
14    "email": "string"
15  },
16  "name": "string",
17  "totalNumberOfSmiles": 0,
18  "createdAt": "2023-07-26T13:18:12.210Z"
19 }

```

HTG-683: API - Drug discovery - Implement GET List endpoint for drug likeness DONE

Рис. 4.6. приклад документації для ендпойнту

Розробка веб-частини також включає засоби для моніторингу та аналізу статистики використання ресурсів, які допомагають оцінити загальну продуктивність та виявити потенційні проблеми на ранніх етапах. Створення звітів, документації та інших матеріалів, які дозволяють користувачам та розробникам краще розуміти та використовувати систему, є не менш важливою частиною розробки.

Як зазначалось вище, на етапі планування інфраструктури роботи з великими даними для генерації молекул було прийнято рішення використовувати Google BigQuery для зберігання і аналізу великих обсягів даних. Це передбачає використання PySpark для завантаження даних з різних джерел і їх обробку перед введенням в BigQuery. Terraform застосовувався для автоматизації створення та управління ресурсами в хмарній інфраструктурі.

Розробка веб-частини ІТ проєкту має багатовимірний характер, що вимагає інтеграції технічних, бізнесових та дизайнерських аспектів зі створенням інтуїтивно зрозумілих інтерфейсів для користувачів. Все це в сукупності спрямоване на забезпечення високого рівня задоволеності користувачів і ефективного досягнення бізнес-цілей проєкту.

4.5. Проєктування схеми бази даних

Схема бази даних у проєкті ВТАІ є фундаментальним елементом, який забезпечує структурування та організацію даних, необхідних для ефективної роботи системи. Основна мета проєктування схеми бази даних полягає у створенні оптимальної структури для зберігання, обробки та аналізу великих обсягів даних, які використовуються у процесах генерації та відкриття нових ліків за допомогою штучного інтелекту.

Проєкт використовує сучасні технології для реалізації бази даних, зокрема, TypeORM як об'єктно-реляційний відображувач (ORM), що дозволяє інтегрувати базу даних з кодом аплікації, написаної на NestJS. TypeORM підтримує різні системи управління базами даних та надає гнучкі можливості

для опису сутностей і відносин між ними в кодї, що значно спрощує процес розробки та підтримки проєкту.

Для проєкту ВТАІ було імплементовано кілька основних моделей даних:

- **AuthModel:** Відповідає за аутентифікацію та управління користувацькими обліковими записами. Забезпечує безпеку та контроль доступу до системи.

- **ProfileModel:** Управляє користувацькими профілями, зберігаючи інформацію, таку як персональні дані, налаштування та історію взаємодії з системою.

- **DrugGenerationModel:** Центральна модель для зберігання та обробки даних про генеровані лікарські засоби. Включає інформацію про структуру молекул, потенційні цілі та результати випробувань.

- **DatasetModel:** Використовується для управління наборами даних, які використовуються у різних експериментах, забезпечуючи індексацію та легкий доступ до них.

- **DockingModel:** Містить результати молекулярного докінгу, допомагаючи в аналізі взаємодій між молекулами ліків та їхніми мішенями.

У розробці схеми бази даних акцент робиться на моделюванні таких ключових ентитетів, як "Drug", "Target" та "Experiment", що дозволяє систематизувати інформацію про потенційні лікарські засоби, їхні цілі (таргети) та результати експериментів. Зв'язки між цими ентитетами реалізовані через відносини, які визначають взаємодії та залежності між різними типами даних, забезпечуючи цілісність та консистенцію інформації.

У додатку І наведено скрипт, написаний на мові програмування JavaScript із використанням фреймворку NestJS та TypeORM. Він визначає клас FullAccountInfoSeeder, який використовується для ініціалізації бази даних за допомогою попередньо визначених даних. Цей сідер (seeder) створює в базі даних приклади записів для різних ентитетів, включаючи користувачів, профілі, набори даних та параметри генерації ліків.

Скрипт використовує бібліотеку `typeorm-extension` для зручного управління сідами та `DataSource` з `TypeORM` для взаємодії з базою даних. Об'єкти, які створюються включають:

- `Account`: Створює запис облікового запису з базовими даними, такими як електронна пошта.
- `User`: Створює користувача з посиланням на обліковий запис та хешованим паролем.
- `Profile`: Створює профіль для користувача з іменем та прізвищем.
- `Dataset`: Записує набір даних, асоційований з користувачем.
- `DrugGenerationParameters` і `DrugGeneration`: Створює параметри для генерації ліків та сам процес генерації, пов'язаний з моделлю та набором параметрів.

Цей скрипт є частиною процесу розробки, що дозволяє розробникам швидко розгорнути базу даних з необхідними тестовими даними для локальної розробки та тестування, забезпечуючи їх наявність після ініціалізації проєкту або її відновлення.

Діаграма зв'язків між реляційними таблицями бази даних наведена в додатку Д.

Крім того, система передбачає інтеграцію з великими даними та їх аналіз, що вимагає забезпечення високої продуктивності та оптимізації запитів до бази даних. Для цього використовуються різні методи оптимізації, такі як індексація даних та розподіл навантаження.

Проектування схеми бази даних також передбачає розробку заходів щодо забезпечення безпеки даних, включаючи авторизацію, аутентифікацію доступу та шифрування чутливої інформації, що є критично важливим для збереження конфіденційності медичної та біологічної інформації.

Загалом, проектування схеми бази даних для ВТАІ забезпечує основу для надійного та ефективного зберігання та обробки даних у рамках проєкту, підтримуючи його основні функціональні вимоги та вимоги до продуктивності.

4.6. Планування та проведення експериментів для ШІ частини проєкту

4.6.1. Опис проведення експериментальної частини проєкту

Планування та проведення експериментів у сфері штучного інтелекту (ШІ) в рамках проєкту ВТАІ, займає ключове місце у розробці інноваційних рішень для виявлення нових лікарських засобів. Частина проєкту присвячено експериментам з ШІ розробляється з застосуванням методології Data-Driven Science (DDS), що дозволяє максимально ефективно використовувати наявні дані для генерації нових знань та відкриттів.

Структурування експериментів: На початковому етапі планування здійснюється детальний аналіз наявних даних і визначення основних завдань, які повинні бути вирішені за допомогою алгоритмів машинного навчання. Це включає визначення ключових метрик успішності, вибір алгоритмів і моделей для тестування, а також підготовку даних для тренування моделей.

Використання методології DDS: DDS передбачає інтеграцію різних джерел даних, таких як хімічні властивості сполук, геномні дані, та клінічні дані, для створення комплексних моделей, які можуть передбачати властивості нових молекул. Цей підхід сприяє глибокому розумінню механізмів дії лікарських засобів та сприяє виявленню нових таргетних терапій.

Аналіз результатів та ітерації: Після проведення експериментів важливо аналізувати отримані результати для оцінки ефективності вибраних моделей. Аналіз включає оцінку точності, чутливості моделей, а також визначення потенційних помилок. На основі аналізу результатів плануються подальші ітерації експериментів для покращення результатів.

Документація експериментів: Кожен етап експерименту документується для забезпечення прозорості та можливості повторення результатів. Документація включає технічні звіти, код експериментів, результати тестування та рекомендації для подальших досліджень. Приклад документації зображений на рисунку 4.7.

Model name	Model purpose	Input data	Expected input data size	Output data	Expected output data size	Used for (Feature)	Execution time (for Expected input data size)
BiAAE	SMILES generation using GES, time, dose	array[GES (L1000 level 3) + time (hours) + dose]; file = .npy	-	SMILES: file = .parquet	100k (based on requested parameter)	Drug generation	5m
Voxel VAE + ShapeCaption	SMILES generation using molecule voxel representation	molecule: file = .sdf	-	SMILES: file = .parquet	100k (based on requested parameter)	Drug generation	15m
DeNoiseGAN	SMILES generation using molecule voxel representation, GES, time, dose	molecule: file = .sdf, array[GES (L1000 level 3) + time (hours) + dose]; file = .npy	-	SMILES: file = .parquet	100k (based on requested parameter)	Drug generation	20m
SMILES VAE [Generation] (Not implemented)	SMILES generation using SMILES	-	-	SMILES: file = .parquet	100k (based on requested parameter)	Drug generation	5m
SMILES VAE [Embeddings]	Create SMILES embeddings	SMILES: file = .csv	-	SMILES, embeddings: file = .parquet	-	SMILES embeddings	5m
Drug likeliness	Predicts metrics of generated drugs	SMILES	100k	CSV with numeric values	100k x 120 columns	Drug likeliness	2.5h (EXAMPLE)
H2O binding class	Predicts affinity of drug-target	SMILES, FASTA + molecule and target props: file=.parquet + used model as CLI argument		CSV with predicted binding or activity class and probability	INPUT_LEN * 3 columns	Predict binding class of drug-target pair	Molecule + target featurization +2s per 1000 records
H2O activity class	Predicts affinity of drug-target	SMILES, FASTA + molecule and target props: file=.parquet + used model as CLI argument		CSV with predicted binding or activity class and probability	INPUT_LEN * 3 columns	Predict activity class of drug-target pair	Molecule + target featurization +2s per 1000 records
Metabolic stability	Predicts metabolic stability class of a compound.	SMILES (CSV)	-	CSV with predicted class and corresponding prediction probability	INPUT_LEN * 2 columns	Assesing generated compounds, specially their metabolic stability during intake	15s for ~400 SMILES (depends on the input data size)
DeepPurpose	Drug-Protein affinity	protein: file = [.pdbqt, .fasta], smiles: file = .parquet	-	SMILES, affinity: file = .parquet	INPUT_SMILES-LEN * 2 columns	Drug-Protein affinity	-
Pafnucy	Drug-Protein affinity	protein: file = .pdbqt, molecules: file = .sdf	-	SMILES, affinity: file = .parquet	INPUT_SMILES-LEN * 2 columns	Drug-Protein affinity	-

Рис. 4.7. Приклад документації моделей III

Серед моделей на рисунку 4.7 можна визначити кілька перспективних підходів для генерації та аналізу хімічних сполук:

- Комбінація геномних та структурних даних: моделі BiAAE (Bidirectional Adversarial Autoencoder) [28], DeNoiseGAN (нова напівконтрольована генеративна змагальна мережа для придушення зображень сітківки адаптивної оптики) [29] та Voxel VAE + Shape Caption (генеративна мережа описів для 3D фігур) [30] довели свою ефективність у генерації нових сполук на основі різних вхідних даних.

- Створення ембедінгів для молекул: використання моделей VAE та RNN для створення векторних представлень відкриває нові можливості для глибинного аналізу хімічних структур.

4.6.2. Розробка моделі H2O для прогнозування метаболічної стабільності сполуки

Під час одного з спринтів було розроблено модуль «Metabolic Stability Estimator». Модуль базується на навченій моделі H2O AutoML[31] для

прогнозування метаболічної стабільності сполуки та використовує відбитки Моргана для перетворення введених SMILES у функції.

Вхідні дані

Очікується, що вхідні дані будуть у форматі DataFrame. Модуль автоматично застосовує правильну модель до рядків з даними біотрансформатора та без них.

Очікуваний вхідний DataFrame повинен містити принаймні стовпець smiles. Для даних біотрансформатора він також повинен містити metabolites_n , metabolites_total_n , responses_n , responses , enzymes_n , enzymes.

Результати передбачень

у_pred,у_proba

1,0.9284701444314636

1,0.8303635733369047

0,0.1282162827781566

1,0.5079910725245599

0,0.1184409998810173

Модуль повертає результати передбачення як Pandas DataFrame з двома стовпцями:

- у_pred – передбачений клас.
 - 0 означає метаболічну **нестабільність** (період напіввиведення <30 м)
 - 1 означає метаболічну **стабільність** (період напіврозпаду > 30 м)
- у_proba – необроблена безпорогова ймовірність прогнозування.

Приклад використання

1. Створення прогнозів

```
import pandas as pd
```

```

from metstab import MetStabClassifier

# Load input data
data = pd.read_csv('input.csv')

# inference model
# use class as a context manager to automatically release used resources
with MetStabClassifier() as model:
    pred = model.predict(data)
# visualize predictions
print(pred)

```

input.csv:

smiles

```

Br.Cn1c(=N)n(CCOc2ccc(Cl)cc2)c2ccc(Cl)cc21
COC12CC3CC(C1)C(CC(=O)O)(c1ccc(F)cc1)C(C3)C2
CCOC(=O)[C@@H](CC(C)C)NP(=O)(OC[C@@H]1OC(OC(C)=O)[C@@H](NC(
C)=O)[C@@H](OC(C)=O)[C@@H]1OC(C)=O)Oc1cccc1
COC(=O)c1cccc(NC(=O)Nc2ccc(-c3ccnc4[nH]cnc34)cc2)c1
CNC(=O)[C@@H](NC(=O)c1ccc(-
c2ccc(CSc3nc4sccc4c(=O)[nH]3)c(F)c2)o1)C(C)C

```

2. Генерація прогнозів для вхідного файлу з додатковими даними біотрансформатора:

```

import pandas as pd

from metstab import MetStabClassifier

# Load input data
data = pd.read_csv('example_data.csv')

# inference model

```

```

# use class as a context manager to automatically release used resources
with MetStabClassifier() as model:
    pred = model.predict(data)

# visualize predictions
print(pred)

```

Модуль автоматично виявляє, де біотрансформатор не зміг згенерувати додаткові дані, використовував спрощену модель для цих сполук і об'єднував їх, враховуючи вихідне індексування.

3. Створення прогнозів для вхідного файлу з іншим конфігураційним файлом:

```

import pandas as pd

from metstab import MetStabClassifier

# Load input data
data = pd.read_csv('example_data.csv')

# inference model
# use class as a context manager to automatically release used resources
with MetStabClassifier(config_path='config_new.json') as model:
    pred = model.predict(data)

# visualize predictions
print(pred)

```

Показники для базової моделі (5-кратний CV)

F1	Accuracy	AUC	Precision	Recall	MCC
0.7333	0.6978	0.7872	0.6395	0.8646	0.4302

Набір даних

Набір даних для навчання та оцінки було вилучено та зібрано зі сполук **ChEMBL32** із визначеним періодом напіввиведення та оцінкою на мікросомах людини, а для стовпця *standard_relation* взято лише рядки з «= \Rightarrow ». Його додатково попередньо обробили та автоматично збільшили дискретизацію в складках за допомогою H₂O.

Перетворення SMILES на фічі

SMILES були перетворені за допомогою **відбитків пальців Моргана**, зокрема така конфігурація використовувалася під час навчання та оцінювання: *radius=2*, *nBits=2048*, *useChirality=True* і *useBondTypes=True*.

Ця конфігурація також використовується як стандартна та доповнює стандартну модель.

Вибір функцій було зроблено за допомогою методу важливості перестановки у випадковому лісі.

Попередня обробка даних біотрансформатора

Додаткові дані біотрансформатора попередньо обробляються за допомогою користувальницького одноразового кодера для *реакцій колонок і ферментів*. Для кодування використовуються лише ті ідентифікатори, які містяться у словниках (*chembl32_enriched_vocab_reactions.npy*, *chembl32_enriched_vocab_enzymes.npy*) (більше 10 входжень у навчальний спліт).

ВИСНОВКИ

З огляду на проведені дослідження та аналіз у рамках магістерської роботи, можна зробити висновок, що розвиток платформ з інтеграцією штучного інтелекту є важливим кроком у подоланні глобальних викликів, зокрема у сфері охорони здоров'я. Окремим викликом для менеджерів є управління такими інноваційними проєктами, адже в них високий рівень невизначеності.

Робота виявила актуальність теми, досліджуючи проблему обмеженого доступу до ефективного лікування та високу вартість розробки нових медичних препаратів. Проєкт розробки інформаційної платформи з штучним інтелектом спрямований на підвищення ефективності та скорочення часу необхідного для відкриття нових лікарських засобів.

В рамках дипломної роботи:

- Було проведено дослідження підходів управління проєктами з розробки штучного інтелекту.
- Проведено аналіз предметної області, який дозволив детально вивчити сучасні підходи, технології та виклики, що виникають під час розробки платформи з штучним інтелектом для відкриття нових ліків.
- Було здійснено аналіз конкурентів та інвестиційних досліджень, що включав аналіз ринку, потенційних інвестиційних можливостей, ризиків та вимог до фінансування проєкту.
- Створено паспорт проєкту, що містить опис продукту, його завдань та цілей, а також детальний опис інформаційної платформи з її ключовими функціями та
- Розроблено життєвий цикл проєкту та здійснена декомпозиція робіт проєкту, що включала планування етапів розробки платформи та поділ на окремі завдання та підзавдання.
- Розраховано модель та симульовано розклад робіт проєкту, створено часовий план виконання завдань.

- Здійснено планування ресурсів та формування бюджету проєкту, визначено необхідні ресурси, включаючи людські, технічні та фінансові.
- Було проведено розрахунок вартості проєкту та оцінено загальні витрати на його реалізацію.
- Побудовано організаційну структуру проєкту, сформовано команду, визначено ролі та відповідальності учасників.
- Розроблено план взаємодії з зацікавленими сторонами, встановлено комунікацію та забезпечено залученість усіх стейкхолдерів у проєкт.
- Визначено ризики та стратегії їх управління, розроблено план мінімізації або усунення потенційних ризиків.
- Розроблено інформаційне та програмне забезпечення проєкту.
- Здійснено моніторинг виконання проєкту, використовуючи показники методу освоєного обсягу для оцінки прогресу та своєчасного реагування на зміни.

Аналіз основних аспектів управління проєктом, зокрема вартості, часу, ризиків, і якості, підкреслив важливість стратегічного підходу до планування та виконання проєктних завдань. Використання інноваційних підходів та сучасних технологій, таких як штучний інтелект, забезпечує не тільки прискорення дослідницького процесу, але й знижує загальні витрати на розробку.

Впровадження розробленої платформи в підприємстві blackthorn.ai і подальше її використання демонструє практичну значимість та економічну ефективність вибраної стратегії управління проєктом. Використання теорії ігор у процесах управління стейкхолдерами дозволило оптимізувати взаємодії всередині та ззовні компанії, що сприяло підвищенню загальної продуктивності проєкту.

Наукова новизна роботи полягає у розробці моделі управління проєктами з розробки штучного інтелекту, що може слугувати основою для подальших досліджень у цій галузі та в впровадження теорії ігор для управління стейкхолдерами. Результати дослідження були високо оцінені на міжнародних конференціях.

ПЕРЕЛІК ВИКОРИСТАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ

1. Information Technology and Implementation (IT&Is-2023)
2. Biewald, L. Why are Machine Learning Projects so Hard to Manage? URL: <https://medium.com/@l2k/why-are-machine-learning-projects-so-hard-to-manage-8e9b9cf49641> (дата звернення: 01.03.2024)
3. Scrum Methodology: Benefits for AI Projects URL: <https://www.tooli.qa/insights/scrum-methodology-benefits-for-ai-projects> (дата звернення: 01.03.2024)
4. How do you manage AI projects with Scrum? URL: <https://www.linkedin.com/advice/0/how-do-you-manage-ai-projects-scrum-skills-artificial-intelligence> (дата звернення: 01.03.2024)
5. Hotz, N. What is CRISP DM? URL: <https://www.datascience-pm.com/crisp-dm-2/> (дата звернення: 01.03.2024)
6. Saltz, J., Sutherland, A., Hotz, N. Achieving Lean Data Science Agility Via Data Driven Scrum 2022
7. Lim, J., Ryu, S., Kim, J.W., et al. Molecular generative model based on conditional variational autoencoder for de novo molecular design 2018, C. 31
8. Shayakhmetov, R., Kuznetsov, M., Zhebrak, A., et al. Molecular Generation for Desired Transcriptome Changes With Adversarial Autoencoders 2020
9. Skalic, M., Jiménez, J., Sabbadin, D., et al. Shape-Based Generative Modeling for de Novo Drug Design 2019, C. 1205–1214
10. Drug discovery URL: https://en.wikipedia.org/w/index.php?title=Drug_discovery&oldid=1196196676 (дата звернення: 05.04.2024)
11. Blanco-González, A., Cabezón, A., Seco-González, A., et al. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies Multidisciplinary Digital Publishing Institute Multidisciplinary Digital Publishing Institute, 2023, C. 891
12. Druchok, M., Gurbuch, O., Yarish, D., et al. Toward efficient generation, correction, and properties control of unique drug-like structures - Druchok - 2021 - Journal of Computational Chemistry - Wiley Online Library C. 746–760
13. Yarish, D., Garkot, S., Grygorenko, O.O., et al. Advancing molecular graphs with descriptors for the prediction of chemical reaction yields 2023, C. 76–92
14. Pharma.ai URL: <https://insilico.com/platform> (дата звернення: 03.04.2024)

15. Pun, F.W., Liu, B.H.M., Long, X., et al. Identification of Therapeutic Targets for Amyotrophic Lateral Sclerosis Using PandaOmics – An AI-Enabled Biological Target Discovery Platform *Frontiers*, 2022
16. AI-accelerated Drug Discovery for Precision Medicine URL: <https://www.receptor.ai/> (дата звернення: 03.04.2024)
17. Opening new frontiers in drug discovery URL: <https://cyclicarx.com/science/> (дата звернення: 03.04.2024)
18. Hidary, J. AI and quantum technologies can also help deliver more than drug development URL: <https://www.weforum.org/agenda/2024/01/ai-and-quantum-revolution-transform-drug-discovery/> (дата звернення: 06.04.2024)
19. How AI reduces the cost and time of drug discovery and development – NaturalAntibody URL: <https://www.naturalantibody.com/use-case/how-ai-reduces-the-cost-and-time-of-drug-discovery-and-development/> (дата звернення: 06.04.2024)
20. AI's potential to accelerate drug discovery needs a reality check URL: <https://www.nature.com/articles/d41586-023-03172-6> (дата звернення: 06.04.2024)
21. Настанова до Зводу знань з управління проєктами (Настанова РМВОК) - Сьоме видання
22. Online Gantt chart maker for project management URL: <https://ganttpro.com/#> (дата звернення: 30.04.2024)
23. Scrum of Scrums URL: <https://www.agilealliance.org/glossary/scrum-of-scrums/> (дата звернення: 19.04.2024)
24. Web Interface URL: <https://chembl.gitbook.io/chembl-interface-documentation/> (дата звернення: 19.04.2024)
25. Bank, R.P.D. RCSB PDB: Homepage URL: <https://www.rcsb.org/> (дата звернення: 19.04.2024)
26. ICD-11 Reference Guide URL: <https://icdcdn.who.int/icd11referenceguide/en/html/index.html#general-features-of-icd11> (дата звернення: 19.04.2024)
27. The NHGRI-EBI Catalog of human genome-wide association studies URL: <https://www.ebi.ac.uk/gwas/> (дата звернення: 19.04.2024)
28. BiAAE - Molecular Generation for Desired Transcriptome Changes with Adversarial Autoencoders *Insilico Medicine*, 2024

29. Wang, S., Li, K., Yin, Q., et al. Semi-supervised generative adversarial learning for denoising adaptive optics retinal images Optica Publishing Group Optica Publishing Group, 2024, C. 1437–1452
30. Shasya, S., Sharma, S., Bhimalapuram, P. Generative schemes for drug design with shape captioning 2023, C. 83
31. H2O AutoML: Automatic Machine Learning — H2O 3.46.0.1 documentation URL: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html> (дата звернення: 10.05.2024)

ДОДАТОК А

Довідка про впровадження результатів магістерської роботи

Довідка

Про впровадження результатів магістерської роботи

Заремби Вікторії Валеріївни

“Управління проектом з розробки інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб і таргетів”

Заремба В.В. на підприємстві BLACKTHORN AI LTD виконувала роль проектною менеджерки та успішно виконала свої обов'язки в рамках проекту з розробки інформаційної платформи з штучним інтелектом для дослідження ліків, хвороб і таргетів

Опис виконаних робіт:

- Ініціація та планування проекту: Забезпечення визначення цілей, завдань, бюджету та ресурсів проекту.
- Розробка та впровадження MVP (мінімально життєздатного продукту): Координація команди розробників для створення базової версії платформи, забезпечення її функціональності та стабільності.
- Моніторинг та контроль ходу проекту: Постійний аналіз прогресу робіт і вжиття коригувальних заходів для розв'язання проблем та затримок.
- Закриття проекту: Підготовка звітності про виконані роботи, оцінка ефективності проекту та забезпечення передачі проектних результатів стейкхолдерам.

BLACKTHORN AI LTD

Kemp House, 128 City Road, United Kingdom, London, EC1V 2NX

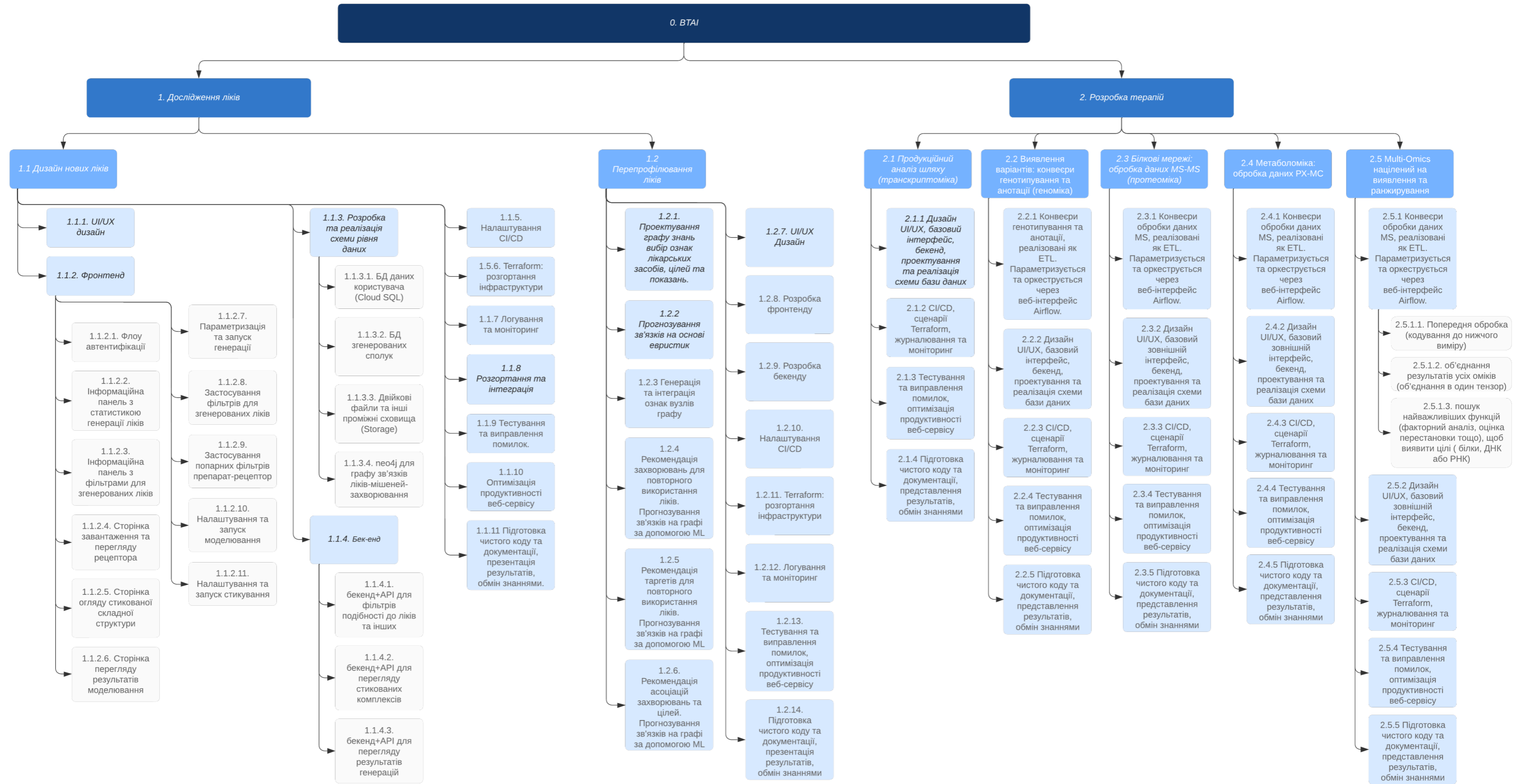
Директор
Гурбич Олександр

«3» травня 2024 р.



ДОДАТОК Б

WBS проєкту



ДОДАТОК В

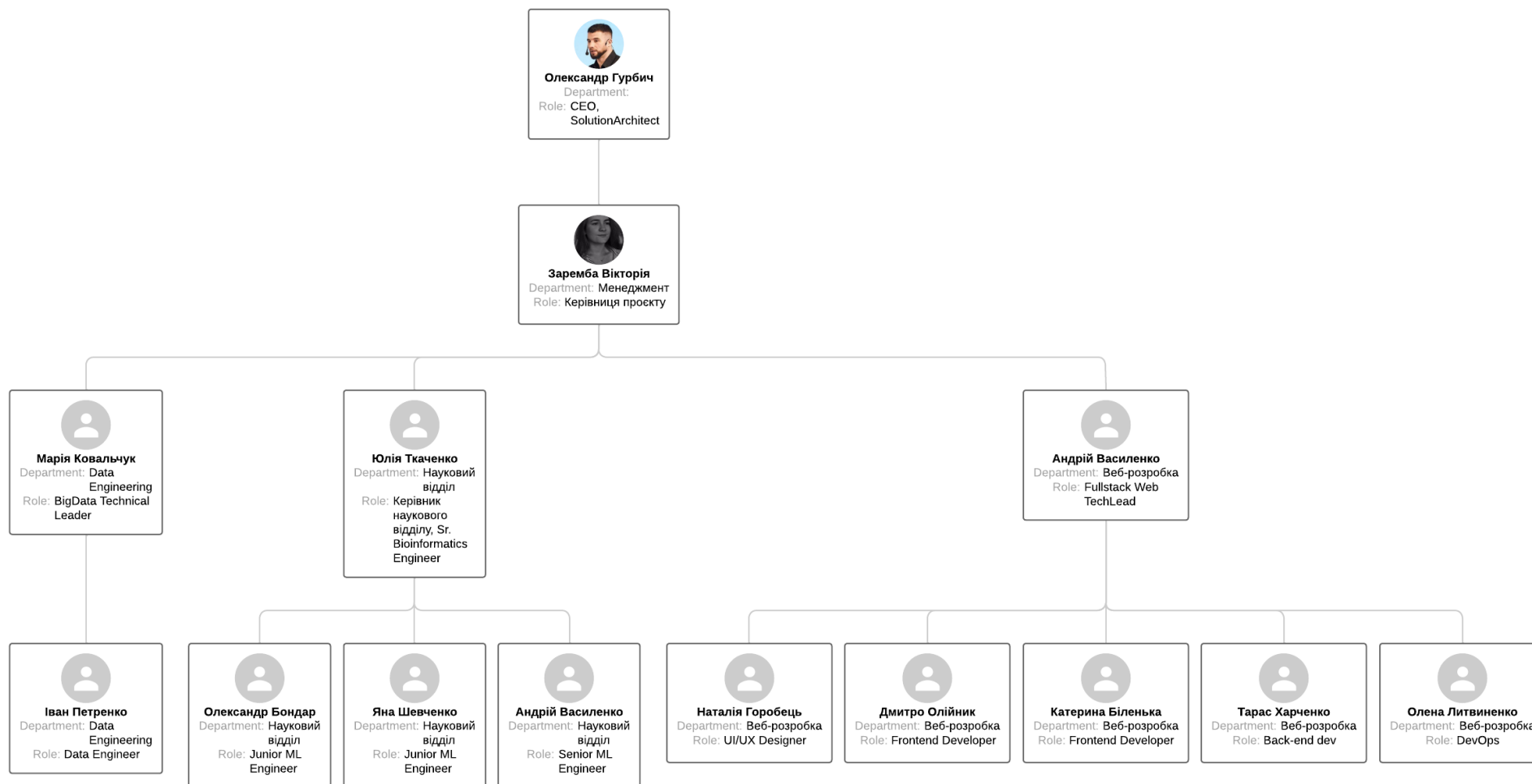
Оцінки часу та призначення ресурсів для MMP проєкту

WBS	Задачі	Оцінка, днів	Ресурси
2	Розробка терапій		
2.1	Продукційний аналіз шляху (транскриптоміка)		
2.1.1	Дизайн UI/UX, базовий інтерфейс, бекенд, проєктування та реалізація схеми бази даних	30	UI/UI Designer, Frondend Dev, Backend Dev
2.1.2	CI/CD, сценарії Terraform, журналювання та моніторинг	20	DevOps
2.1.3	Тестування та виправлення помилок, оптимізація продуктивності веб-сервісу	15	ML Engineer, full-stack
2.1.4	Підготовка чистого коду та документації, представлення результатів, обмін знаннями	5	ML Engineer, full-stack
2.2	Виявлення варіантів: конвеєри генотипування та анотації (геноміка)		
2.2.1	Виявлення варіантів: конвеєри генотипування та анотації (геноміка)	160	
2.2.2	Дизайн UI/UX, базовий інтерфейс, бекенд, проєктування та реалізація схеми бази даних	30	UI/UI Designer, Frondend Dev, Backend Dev
2.2.3	CI/CD, сценарії Terraform, журналювання та моніторинг	20	DevOps
2.2.4	Тестування та виправлення помилок, оптимізація продуктивності веб-сервісу	5	ML Engineer, full-stack
2.2.5	Підготовка чистого коду та документації, представлення результатів, обмін знаннями	5	ML Engineer, full-stack
2.3	Білкові мережі: обробка даних MS-MS (протеоміка)		
2.3.1	Конвеєри обробки даних MS, реалізовані як ETL Параметризується та оркеструється через веб-інтерфейс Airflow	180	Sr. Bioinformatician, ML Engineer (1)
2.3.2	Дизайн UI/UX, базовий зовнішній інтерфейс, бекенд, проєктування та	30	UI/UI Designer, Frondend

	реалізація схеми бази даних		Dev, Backend Dev
2.3.3	CI/CD, сценарії Terraform, журналювання та моніторинг	20	DevOps
2.3.4	Тестування та виправлення помилок, оптимізація продуктивності веб-сервісу	10	ML Engineer, full-stack
2.3.5	Підготовка чистого коду та документації, представлення результатів, обмін знаннями	5	ML Engineer, full-stack
2.4	Метаболоміка: обробка даних РХ-МС		
2.4.1	Метаболоміка: обробка даних РХ-МС	180	Sr. Bioinformatician, ML Engineer (1)
2.4.2	Дизайн UI/UX, базовий зовнішній інтерфейс, бекенд, проектування та реалізація схеми бази даних	30	UI/UI Designer, Frondend Dev, Backend Dev
2.4.3	CI/CD, сценарії Terraform, журналювання та моніторинг	20	DevOps
2.4.4	Тестування та виправлення помилок, оптимізація продуктивності веб-сервісу	10	ML Engineer, full-stack
2.4.5	Підготовка чистого коду та документації, представлення результатів, обмін знаннями	5	ML Engineer, full-stack
2.5	Multi-Omics націлений на виявлення та ранжирування		
2.5.1	Конвеєри обробки даних MS, реалізовані як ETL Параметризується та оркеструється через веб-інтерфейс Airflow	40	Sr. Bioinformatician, ML Engineer (1)
2.5.2	Дизайн UI/UX, базовий зовнішній інтерфейс, бекенд, проектування та реалізація схеми бази даних	40	UI/UI Designer, Frondend Dev, Backend Dev
2.5.3	CI/CD, сценарії Terraform, журналювання та моніторинг	20	DevOps
2.5.4	Тестування та виправлення помилок, оптимізація продуктивності веб-сервісу	10	ML Engineer, full-stack
2.5.5	Підготовка чистого коду та документації, представлення результатів, обмін знаннями	5	ML Engineer, full-stack

ДОДАТОК Г

Організаційна структура



ДОДАТОК Г

Скрипт для ініціалізації бази даних

```
import { Seeder, SeederFactoryManager } from 'typeorm-extension';
import { DataSource } from 'typeorm';
import { Account } from '../entities/account.entity';
import { ACCOUNT_UUID, USER_UUID, UNIVERSAL_UUID } from
'../constants/account-seed.constant';
import { Profile } from '../entities/profile.entity';
import { User } from '../entities/user.entity';
import { hashPassword } from '../utils/hash-password.util';
import { DrugGeneration } from '../entities/drug-generation.entity';
import { Dataset } from '../entities/datasets.entity';
import { DrugGenerationParameters } from '../entities/drug-generation-
parameters.entity';
import { DrugGenerationModel } from '../enum/drug-generation-model';

export default class FullAccountInfoSeeder implements Seeder {
  public async run(dataSource: DataSource, factoryManager:
SeederFactoryManager): Promise<any> {
    const accountFactory = await factoryManager.get(Account);
    await accountFactory.save({
      id: ACCOUNT_UUID,
      email: 'seed.mail@htgmolucular.com',
    });

    await factoryManager.get(User).save({
      id: USER_UUID,
      accountId: ACCOUNT_UUID,
      email: 'seed.mail@htgmolucular.com',
      password: hashPassword('password123'),
```

```
});
```

```
await factoryManager.get(Dataset).save({  
  id: UNIVERSAL_UUID,  
  name: 'Dataset Name',  
  userId: USER_UUID,  
});
```

```
await factoryManager.get(DrugGenerationParameters).save({  
  id: UNIVERSAL_UUID,  
  moleculeFilePath: 'path',  
  numberOfSamples: 1,  
});
```

```
await factoryManager.get(DrugGeneration).save({  
  id: UNIVERSAL_UUID,  
  name: 'Generation Name',  
  generatedDatasetId: UNIVERSAL_UUID,  
  modelType: DrugGenerationModel.BIAAE,  
  parametersId: UNIVERSAL_UUID,  
  userId: USER_UUID,  
});
```

```
await factoryManager.get(Profile).save({  
  userId: USER_UUID,  
  firstName: 'Alex',  
  lastName: 'Smith',  
});  
}}
```

