

МІНІСТЕРСТВО ОХОРОНИ ЗДОРОВ'Я УКРАЇНИ
НАЦІОНАЛЬНИЙ ФАРМАЦЕВТИЧНИЙ УНІВЕРСИТЕТ
Факультет медико-фармацевтичних технологій
Кафедра промислової технології ліків та косметичних засобів

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «**ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АНАЛІЗУ QUALITY RISK ASSESSMENT (QRA) ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ**»

Виконав: здобувач вищої освіти,
групи ТФПм-21(4,10д.)
спеціальності: 226 Фармація, промислова фармація
освітньо-професійної програми
Технології фармацевтичних препаратів
Алексєєв Микита

Керівник: доцент закладу вищої освіти кафедри
промислової технології ліків та косметичних засобів,
к.фарм.н., доцент
Дмитро СОЛДАТОВ

Рецензент: доцент закладу вищої освіти кафедри
аптечної технології ліків, к.фарм.н., доцент
Марина БУРЯК

Харків – 2026 рік

АНОТАЦІЯ

Дослідження присвячене інтеграції великих мовних моделей (LLM) у процеси оцінки ризиків якості (QRA) у фармацевтичній галузі відповідно до вимог ICH Q9. Запропоновано підхід на основі Prompt Engineering для автоматизації аналізу технологічної документації та формування матриць ризиків FMEA.

Особливу увагу приділено перевірці даних, згенерованих ШІ, на відповідність принципам ALCOA+ і вимогам GxP. Використання методики дозволяє скоротити час проведення оцінки ризиків, підвищити об'єктивність аналізу та покращити виявлення критичних відхилень у виробництві лікарських засобів.

Ключові слова: великі мовні моделі, QRA, Prompt Engineering, ALCOA+, GxP, FMEA.

ANNOTATION

The study focuses on integrating large language models (LLM) into quality risk assessment (QRA) processes in the pharmaceutical industry in accordance with ICH Q9 requirements. A Prompt Engineering-based approach for automating the analysis of technological documentation and generating FMEA risk matrices is proposed.

Special attention is paid to the verification of AI-generated data for compliance with ALCOA+ principles and GxP requirements. The methodology reduces the time required for risk assessment, improves the objectivity of analysis, and enhances the identification of critical deviations in pharmaceutical manufacturing.

Keywords: large language models, QRA, Prompt Engineering, ALCOA+, GxP, FMEA.

ЗМІСТ

ВСТУП	6
РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В УПРАВЛІННІ РИЗИКАМИ ЯКОСТІ ФАРМАЦЕВТИЧНОГО ВИРОБНИЦТВА	8
1.1. Еволюція концепції управління ризиками якості: від контролю готової продукції до предиктивного управління	8
1.2. Класичні методи оцінки ризиків якості та межі їх практичної застосовності в умовах сучасного виробництва	9
1.3. Цифрова трансформація фармацевтичного виробництва та регуляторний контекст впровадження штучного інтелекту в систему GxP	11
Висновки до розділу 1	15
РОЗДІЛ 2 МЕТОДОЛОГІЯ ІНТЕГРАЦІЇ ШТУЧНОГО ІНТЕЛЕКТУ В ПРОЦЕСИ ОЦІНКИ РИЗИКІВ ЯКОСТІ	17
2.1. Архітектурна модель застосування великих мовних моделей як когнітивного інструменту фахівця з валідації	17
2.2. Структурований промпт-інжиніринг для аналізу технологічних регламентів та специфікацій якості	19
2.3. Алгоритм верифікації результатів аналізу та відповідність принципам ALCOA+ в умовах регульованого виробництва	20
Висновки до розділу 2	22
РОЗДІЛ 3 ПРАКТИЧНА АПРОБАЦІЯ МЕТОДОЛОГІЇ: ПОРІВНЯЛЬНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ГЕНЕРАЦІЇ ТАБЛИЦЬ ОЦІНКИ РИЗИКІВ ЗА ДОПОМОГОЮ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ	23
3.1. Дизайн експериментального дослідження та характеристика вхідних даних	23
3.2. Еталонна FMEA валідаційного фахівця як база порівняння	25

3.3. Серія тестів Qwen3-14B Thinking	27
3.4. Серія тестів GLM 4.6v Flash	39
3.5. Серія тестів Qwen3-30B Thinking та GLM 4.7 Flash	43
3.6. GPT-OSS – Референтний тест	46
3.7. Загальна тенденція наближення до еталону	47
Висновки до розділу 3	48
Висновки	50
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	52

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ALCOA+ – Attributable, Legible, Contemporaneous, Original, Accurate (plus)

CPP – Critical Process Parameter

CQA – Critical Quality Attribute

FMEA – Failure Mode and Effects Analysis

GAMP – Good Automated Manufacturing Practice

GMP – Good Manufacturing Practice

GxP – Good Practice

HACCP – Hazard Analysis and Critical Control Points

ICH – International Council for Harmonisation

IMiD – Immunomodulatory Drugs

ISPE – International Society for Pharmaceutical Engineering

LLM – Large Language Models

ML – Machine Learning

PHA – Preliminary Hazard Analysis

PQS – Pharmaceutical Quality System

QRA – Quality Risk Assessment

QRM – Quality Risk Management

QbD – Quality by Design

RPN – Risk Priority Number

ДФУ – Державна фармакопея України

МОЗУ – Міністерство охорони здоров'я України

НФаУ – Національний фармацевтичний університет

ШІ – Штучний інтелект

ВСТУП

Актуальність теми: Сучасна фармацевтична індустрія перебуває на етапі цифрової трансформації в межах парадигми Pharma 4.0. Одним із найбільш критичних та ресурсомістких процесів системи управління якістю (PQS) є управління ризиками (Quality Risk Management - QRM). Згідно з настановою ІСН Q9, оцінка ризиків має базуватися на наукових знаннях та бути пов'язаною із захистом пацієнта. Однак традиційні методи оцінки ризиків якості (QRA, від англ. Quality Risk Assessment) (FMEA, HACCP) часто страждають від суб'єктивізму експертів, великих часових витрат та фрагментарності аналізу даних. Застосування штучного інтелекту, зокрема великих мовних моделей (LLM), відкриває нові можливості для автоматизації аналізу текстової документації та виявлення прихованих кореляцій між параметрами процесу, що робить дане дослідження надзвичайно своєчасним.

Мета дослідження: дослідити та оцінити ефективність застосування інструментів штучного інтелекту для проведення аналізу ризиків якості у фармацевтичному виробництві.

Завдання дослідження:

- Проаналізувати еволюцію концепцій QRM та виявити обмеження класичних методів аналізу ризиків.
- Оцінити поточний стан та регуляторні бар'єри впровадження ШІ в GxP-середовище.
- Порівняти ефективність традиційного та ШІ-орієнтованого підходів на прикладі конкретної технологічної схеми виробництва.
- Сформулювати рекомендації щодо забезпечення цілісності даних (Data Integrity) при використанні ШІ в управлінні ризиками.

Об'єкт дослідження: процес управління ризиками якості на фармацевтичному підприємстві згідно з принципами ІСН Q9.

Предмет дослідження: методи та алгоритми використання штучного інтелекту (зокрема LLM) для автоматизації аналізу QRA.

Методи дослідження: У роботі використано загальнонаукові методи аналіз та синтез (при вивченні літературних джерел та нормативної бази, порівняльний аналіз при зіставленні результатів роботи ШІ та експерта) метод моделювання (при розробці промптів та метод) експертних оцінок.

Практичне значення отримання результатів: Розроблено методику автоматизації оцінки ризиків якості яка дозволяє підвищити об'єктивність аналізу та скоротити часові витрати на підготовку валідаційної документації.

Структура та обсяг кваліфікаційної роботи: Робота складається зі вступу, трьох розділів, висновків списку використаних джерел та додатків. Загальний обсяг роботи 51 сторінок, містить таблиць 23, джерел літератури 30.

РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В УПРАВЛІННІ РИЗИКАМИ ЯКОСТІ ФАРМАЦЕВТИЧНОГО ВИРОБНИЦТВА

1.1. Еволюція концепції управління ризиками якості: від контролю готової продукції до предиктивного управління

Забезпечення якості лікарських засобів тривалий час ґрунтувалося на логіці контролю кінцевого продукту: відповідність серії встановлювалася шляхом випробувань готової продукції, а виявлення невідповідності означало вже понесені виробничі втрати. Такий підхід за своєю природою є реактивним - він дозволяє зафіксувати проблему, але не запобігти їй на рівні процесу [1]. Поступово в галузі сформувалося розуміння, що стабільна якість не може бути «вбудована» шляхом перевірки - вона має бути закладена у конструкцію процесу та продукту від самого початку.

Концептуальним підґрунтям для цього переходу стала настанова ІСН Q9 «Quality Risk Management», яка встановила систематичний підхід до ідентифікації, оцінювання, контролю та перегляду ризиків для якості продовж усього життєвого циклу лікарського засобу [2]. Вперше було чітко сформульовано принцип пропорційності: рівень формалізації та зусиль при управлінні ризиками має відповідати рівню потенційної загрози для пацієнта. Саме безпека пацієнта визначається головним критерієм при розстановці пріоритетів у системі управління якістю [3].

Паралельно з ІСН Q9 набула чинності настанова ІСН Q10 «Pharmaceutical Quality System», яка запропонувала цілісну модель фармацевтичної системи якості, де управління ризиками є не окремою процедурою, а наскрізним механізмом прийняття рішень на всіх рівнях організації [4]. На національному рівні ці вимоги імплементовано через Настанову СТ-Н МОЗУ 42-4.0:2020 - вітчизняний аналог правил GMP Європейського Союзу [5]. Вимоги до якості

поширюються не лише на виробництво, а й на весь ланцюг дистрибуції, що відображено у Настанові СТ-Н МОЗУ 42-4.2:2011 [6], та підкріплено відповідними стандартами Державної фармакопеї України [7]. Загальна правова основа обігу лікарських засобів визначена Законом України «Про лікарські засоби» [8].

Оновлення ІСН Q9 до версії (R1) у 2023 році [2] засвідчило нову хвилю еволюції концепції. Ключовим посилом редакції стало визнання проблеми надмірного суб'єктивізму в оцінках ризиків та вимога переходу до більш відтворюваних, науково обґрунтованих підходів. Центр ваги переміщується з аналізу вже виявлених відхилень на прогнозування потенційних збоїв ще на стадії розробки продукту та процесу - тобто концепція Quality by Design (QbD) стає органічною частиною управління ризиками.

Управління ризиками якості не існує у вакуумі - воно нерозривно пов'язане з логістичними та економічними процесами підприємства. Ризики якості генерують каскадні наслідки: від витрат на вилучення серії до репутаційних збитків, тому їх оцінка є стратегічним елементом фінансового планування [7, 9, 10, 11]. Саме така комплексна перспектива — де управління якістю розглядається як частина загальної системи менеджменту підприємства - відображає сучасне розуміння ролі QRM у фармацевтичній галузі.

1.2. Класичні методи оцінки ризиків якості та межі їх практичної застосовності в умовах сучасного виробництва

Практична реалізація принципів ІСН Q9 вимагає конкретних аналітичних інструментів. За десятиліття застосування у фармацевтичній галузі сформувався усталений набір методів, запозичених переважно з інженерної справи та харчової промисловості [8, 12].

Найбільш поширеним залишається метод FMEA (Failure Mode and Effects Analysis) - аналіз видів та наслідків відмов. Його привабливість зумовлена відносною простотою формалізації: кожному потенційному ризику присвоюється пріоритетне число ризику (RPN), що є добутком трьох числових оцінок - ймовірності виникнення (P), тяжкості наслідків (E) та ймовірності виявлення (D). В поєднанні з принципами фармацевтичного менеджменту FMEA дає змогу структурувати процес ухвалення рішень щодо критичних параметрів процесу [11, 13]. Метод HACCP (Hazard Analysis and Critical Control Points) традиційно орієнтований на систематичне виявлення фізичних, хімічних та мікробіологічних загроз у виробничому середовищі. PHA (Preliminary Hazard Analysis) застосовується переважно на ранніх стадіях проектування нових систем, формуючи попередній реєстр ризиків до початку детального опрацювання.

Специфіка стерильного виробництва вимагає суттєвого ускладнення стандартних матриць ризиків через мультифакторну природу мікробіологічної контамінації - жодна з названих методик у своєму базовому вигляді не покриває всієї складності цього середовища [14]. Принципово важливо розуміти, що будь-який метод QRA є інструментом, а не гарантією якості: ефективність аналізу критично залежить від компетентності команди та повноти вхідних даних [15].

Попри широке застосування, класичні методи виявляють ряд системних обмежень. По-перше, висока суб'єктивність числових оцінок: бали за шкалами FMEA нерідко відображають інтуїцію конкретного експерта більше, ніж об'єктивний аналіз даних. Відсутність єдиних критеріїв калібрування шкал робить порівняння FMEA-оцінок між різними підрозділами одного підприємства практично неможливим [16]. По-друге, методи є статичними за своєю природою: реєстри ризиків оновлюються рідко й перетворюються на формальні документи, що не відображають актуальний стан виробничих

процесів. По-третє, класичні підходи погано масштабуються на великі масиви неструктурованих текстових даних - звіти про відхилення, журнали обладнання, протоколи валідацій. Це суттєво обмежує їх аналітичну цінність в умовах цифрового виробництва.

Ці обмеження слід розглядати не як вади конкретних методів, а як системну проблему, що потребує переосмислення самої архітектури інструментарію QRM [17]. Такий погляд є концептуальним підґрунтям для наступного підрозділу, присвяченого цифровим підходам до управління ризиками.

1.3. Цифрова трансформація фармацевтичного виробництва та регуляторний контекст впровадження штучного інтелекту в систему GxP

Концепція Pharma 4.0, що є галузевою адаптацією принципів Industry 4.0, передбачає інтеграцію кіберфізичних систем, хмарних обчислень та інтелектуальної аналітики в усі ланки виробничого ланцюга. Для вітчизняної фармацевтичної галузі цей перехід пов'язаний зі значними викликами - технологічними, регуляторними та організаційними, - хоча потенціал для прискореного розвитку є суттєвим [18].

У контексті управління якістю цифрова трансформація проявляється передусім у застосуванні технологій штучного інтелекту (ШІ) та машинного навчання (ML). Серед актуальних напрямків - предиктивне технічне обслуговування обладнання на основі сенсорних даних, інтелектуальний аналіз кореневих причин відхилень, а також автоматизована перевірка відповідності документації регуляторним вимогам. Розгляд цих технологій крізь призму організаційних трансформацій, необхідних для їх ефективної інтеграції в систему управління якістю, є важливим науковим завданням [19]. Алгоритми машинного навчання здатні виявляти приховані закономірності у великих

масивах виробничих даних, недоступні для традиційних аналітичних методів, що відкриває принципово нові можливості для QRM [12].

Практичні аспекти впровадження інтелектуальних систем у фармації вимагають застосування гібридних підходів, що поєднують детерміновані алгоритми з адаптивними нейронними архітектурами [20]. Разом з тим специфічні труднощі регульованого середовища GxP (від англ. Good Practices — загальна назва стандартів належних практик) - підтвердження комп'ютеризованих систем (computer system validation), забезпечення цілісності даних, відстеження версій моделей - суттєво ускладнюють практичну реалізацію [21].

Регуляторний вимір питання є визначальним. Провідні регуляторні агентства перебувають у процесі розробки відповідних настанов щодо ШІ, однак уже сформульовано базові очікування: прозорість алгоритмів, відтворюваність результатів та обов'язковий людський контроль над автоматизованими висновками. Методологічним орієнтиром для валідації ШІ-рішень у середовищі GxP є настанова ISPE GAMP 5 [22], що регламентує ризик-орієнтований підхід до комп'ютеризованих систем. Впровадження цифрових технологій створює нові виклики для забезпечення принципів ALCOA+ та цілісності даних [23].

На рівні конкретних технологій особливий інтерес становлять великі мовні моделі (LLM). Дослідження в цій сфері демонструють значний потенціал LLM для аналізу складної наукової та регуляторної документації, водночас окреслюючи ключові виклики - можливість генерації недостовірної інформації, необхідність верифікації висновків та питання відповідальності [24]. Роль ШІ загалом у трансформації фармацевтичної галузі полягає не в автономності алгоритмів, а в їхній здатності підсилювати людську експертизу [24, 25]. У контексті QRM зокрема найперспективнішою є модель, де ШІ виступає когнітивним інструментом, що допомагає фахівцю глибше та швидше

аналізувати ризики, не підміняючи його відповідальності [26]. Підтвердженням цього слугує досвід застосування ІІІ для систематизації масивів медичних даних у клінічних дослідженнях, де схожа модель розподілу ролей між людиною та алгоритмом суттєво підвищила якість аналізу [27].

Таким чином, цифровізація за принципами Pharma 4.0 та інтеграція ІІІ є не лише технологічним трендом, а й регуляторно обумовленим вектором розвитку системи управління якістю, що відкриває можливості для подолання системних обмежень класичних методів QRA.

1.4. Об'єкт валідації: характеристика лікарського засобу та виробничого процесу

Загальна характеристика препарату.

Об'єктом валідації у цьому дослідженні є виробничий процес твердих капсул леналідоміду — синтетичного похідного талідоміду, що належить до класу імуномодуляторів (IMiD). Препарат зареєстрований для лікування множинної мієломи, мієлодиспластичного синдрому, а також мантийно-клітинної лімфоми. Завдяки своїй здатності модулювати імунну систему та пригнічувати ангиогенез, леналідомід увійшов до переліку есенціальних лікарських засобів ВООЗ (Всесвітньої організації охорони здоров'я). Препарат випускається у семи дозуваннях: 2,5; 5; 7,5; 10; 15; 20 та 25 мг, що зумовлює необхідність розробки FMEA з диференціацією ризиків за потужністю — особливість, яку фахівець врахував в еталонному документі.

Класифікація за рівнем небезпечності.

Леналідомід належить до класу високоактивних фармацевтичних субстанцій НРАРІ з категорією ОЕВ 4/5 за класифікацією ОЕВ (Occupational Exposure Band). Для речовин цього класу встановлено допустимий добовий рівень виробничого впливу (OEL, Occupational Exposure Limit) на рівні 1–10

нг/м³, що вимагає застосування закритих систем зважування та наповнення, засобів індивідуального захисту рівня C/D, а також регулярного моніторингу повітря робочої зони (SMEPAC-тестування). Саме ця класифікація зумовлює критичну важливість контейнменту як окремої категорії ризику, що знайшло відображення в тестах AI-моделей із застосуванням правила Severity Override.

Технологічна схема виробництва.

Виробничий процес капсул леналідоміду включає шість основних стадій. Стадія зважування: відважування всіх компонентів (діюча речовина, лактоза моногідрат, мікрокристалічна целюлоза — МКЦ, кроскармелоза натрію, магнію стеарат) з суворим дотриманням вимог до контейнменту. Стадія тритурації: попереднє змішування леналідоміду з частиною лактози для отримання рівномірного розподілу діючої речовини при її малій дозі. Стадія приготування бленду (Blend I): змішування тритурації з рештою допоміжних речовин (лактоза, МКЦ, кроскармелоза). Стадія лубрикації: додавання магнію стеарату та фінальне перемішування для забезпечення належної текучості бленду. Стадія наповнення капсул: заповнення твердих желатинових капсул на автоматичній машині-дозаторі з проміжним контролем маси. Стадія пакування: блистерне пакування із забезпеченням герметичності.

Критичні показники якості (CQA) та обґрунтування їх вибору як об'єкту FMEA.

Для даного препарату визначено такі CQA, що є предметом оцінки ризиків. Уніформність бленду та вміст вологи (Blend Uniformity, Water Content) — визначальний параметр для забезпечення рівномірного розподілу діючої речовини у малій дозі. Фізичні властивості бленду (Flowability, Bulk Density) - критичні для стабільного наповнення капсул. Зовнішній вигляд капсул (Description/Appearance) — контролюється на кожній серії. Розпадання (Disintegration) - визначає вивільнення діючої речовини in vitro. Розчинення (Dissolution) - параметр, що безпосередньо корелює з біодоступністю, особливо

критичний для потужності 25 мг через відмінності у складі. Рівномірність маси (Weight Uniformity) та уніформність вмісту (Content Uniformity) - найважливіші показники точності дозування, що є предметом підвищеного контролю для дозувань 2,5 мг через мале наповнення капсули.

Загальна характеристика системи управління якістю виробництва.

Виробництво капсул леналідоміду здійснюється в умовах, що відповідають вимогам GMP (Good Manufacturing Practice) ЄС відповідно до Настанови СТ-Н МОЗУ 42-4.0:2020. Контроль якості включає обов'язкові ІРС (In-Process Control) - контроль у процесі виробництва: зважування наповнених капсул на початку, в середині та наприкінці виробничого циклу; 100% візуальний контроль зовнішнього вигляду; тестування представницьких зразків кожної серії in-bulk на розпадання, розчинення та уніформність вмісту. Протокол валідації процесу (Process Validation Protocol, PVP) відповідно до вимог EU GMP Annex 15 описує план кваліфікаційних досліджень, критерії прийнятності та ідентифіковані CPP (Critical Process Parameters). Саме цей протокол (без розділу оцінки ризиків) є одним із ключових вхідних документів у системі тестів AI-моделей у Розділі 3.

Висновки до розділу 1

Проведений аналіз теоретичних засад управління ризиками якості дозволяє сформулювати такі висновки: 1. концепція QRM зазнала принципової трансформації - від постфактумного контролю готової продукції до проактивного, науково обґрунтованого управління ризиками впродовж усього життєвого циклу лікарського засобу, що закріплено у нормативній базі ICH Q9 (R1), ICH Q10 та відповідних національних документах. 2. класичні інструменти QRA - FMEA, HACCP, PHA - досягли межі своєї ефективності: їх структурна простота, що колись була перевагою, стала обмеженням в умовах великих

обсягів неструктурованих даних та вимог до відтворюваності оцінок. 3. цифрова трансформація за принципами Pharma 4.0 та технології штучного інтелекту формують нову парадигму управління якістю, реалізація якої потребує як технічних рішень, так і відповідного регуляторного та методологічного забезпечення.

РОЗДІЛ 2 МЕТОДОЛОГІЯ ІНТЕГРАЦІЇ ШТУЧНОГО ІНТЕЛЕКТУВ ПРОЦЕСИ ОЦІНКИ РИЗИКІВ ЯКОСТІ

2.1. Архітектурна модель застосування великих мовних моделей як когнітивного інструменту фахівця з валідації

Дискусія щодо ролі ШІ у регульованому середовищі нерідко зводиться до хибної дихотомії: або повна автоматизація, або відмова від цифрових інструментів. Проте між цими крайнощами існує більш прагматичний шлях - використання великих мовних моделей (LLM) як інтелектуального асистента, що розширює аналітичні можливості фахівця, не підмінюючи його відповідальності за прийняте рішення. Цей підхід, відомий як «Human-in-the-loop», є базовим архітектурним принципом методології, що розглядається у цьому розділі. Ключова передумова його ефективності полягає в тому, що ШІ найкраще працює саме там, де він доповнює, а не замінює людську експертизу [28].

Архітектура запропонованого рішення охоплює чотири функціональні рівні. Перший - шар вхідних даних, що включає специфікації на сировину та готову продукцію, описи технологічних процесів виробництва, звіти про відхилення та протоколи валідацій. Другий рівень - когнітивний процесор на основі LLM, здатний аналізувати довгі документи та генерувати структуровані висновки. Третій рівень - база знань, сформована з регуляторних документів (ICH Q9, ISO 14644, внутрішніх СОП підприємства), до якої модель звертається через механізм RAG (Retrieval-Augmented Generation). Четвертий рівень - валідація результату: порівняння згенерованих висновків із еталонними даними та обов'язкова перевірка кваліфікованим фахівцем.

Узагальнена архітектура запропонованого рішення наведена на Рисунку 2.1. Схема відображає чотири функціональні рівні та потоки даних між ними, включаючи зворотний зв'язок для ітеративного вдосконалення промпту.

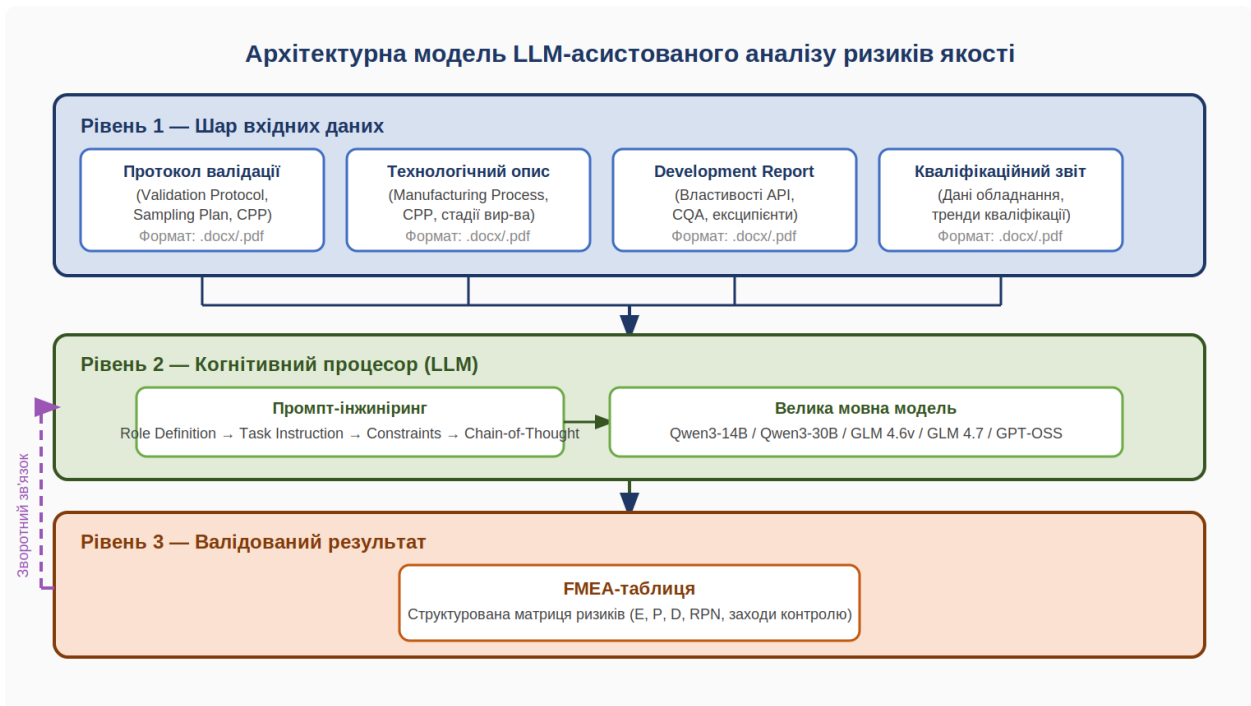


Рисунок 2.1 – Архітурна модель 4-рівневої системи LLM-асистованого аналізу ризиків якості

Питання підтвердження та кваліфікації комп'ютеризованих систем у середовищі GxP детально регламентує настанова ISPE GAMP 5 [22], яка є методологічним орієнтиром для розробки протоколів валідації ШІ-рішень. Принципово важливим є розуміння функціональної ролі LLM у цій архітурі: модель виступає не автоматом для прийняття рішень, а інструментом структурування та попереднього аналізу інформації. Такий розподіл функцій - де алгоритм систематизує масиви даних, а людина ухвалює остаточне рішення - демонструє суттєве підвищення якості аналізу порівняно з кожним із підходів окремо [27].

Ключовою перевагою архітурі є масштабованість: один когнітивний процесор може аналізувати документи різної природи - від технологічного регламенту виробництва твердих лікарських форм до протоколу кваліфікації чистих приміщень - виключно за рахунок зміни підказок (промтів) та підключеної бази знань. Це дозволяє уніфікувати підхід до оцінки ризиків на

рівні всього підприємства та забезпечити узгодженість методології в різних виробничих підрозділах.

2.2. Структурований промпт-інжиніринг для аналізу технологічних регламентів та специфікацій якості

Якість вихідних даних будь-якої LLM безпосередньо визначається якістю вхідного запиту. Промпт-інжиніринг (Prompt Engineering) - дисципліна, що вивчає методи формулювання запитів для отримання передбачуваних, точних та відтворюваних відповідей від мовних моделей. Системний підхід до конструювання промтів є новим типом наукової методології, що потребує такої ж формалізації, як і розробка дослідницького протоколу в будь-якій іншій галузі знань [29].

Для задач QRA у фармацевтичному виробництві розроблена чотирирівнева структура промту, що базується на систематизованих шаблонах взаємодії з LLM [30].

Перший рівень - визначення ролі (Role Definition). Модель отримує чітко сформульований професійний контекст: її роль визначається не як загального асистента, а як фахівця з управління ризиками якості зі знанням відповідної нормативної бази. Прикладом може слугувати таке формулювання: «Ти - провідний експерт з QRM у фармацевтичній галузі з глибоким знанням стандартів ICH Q9 (R1), GMP та внутрішніх вимог підприємства». Це активує відповідні сегменти знань моделі та задає термінологічний стандарт подальшого аналізу.

Другий рівень - постановка завдання (Task Instruction). Тут формулюється конкретна аналітична задача з прив'язкою до документа, що аналізується. Наприклад: «Проаналізуй стадію вологої грануляції у наданому регламенті.

Визнач критичні параметри процесу (CPP) та оціни їх вплив на критичні показники якості (CQA): однорідність вмісту та розпадання таблеток».

Третій рівень - обмеження та логічні правила (Constraints). З точки зору регуляторної відповідності це є найважливішим елементом структури. Саме тут задаються правила, що мінімізують ризик генерації недостовірної інформації. Ключовий прийом - вимога маркувати відсутні дані як «Інформація відсутня / необхідний запит виробнику» замість їх домислення. Це принципово важливий механізм управління відомою проблемою «галюцинацій» LLM в аналітичних задачах, де точність є критичною.

Четвертий рівень - навідні підказки (Chain-of-Thought Prompting). Модель отримує покроковий алгоритм міркування: спочатку виявити фізико-хімічні ризики, потім оцінити ефективність наявних засобів контролю, і лише після цього розрахувати підсумковий бал ризику. Це забезпечує відтворюваність логіки аналізу незалежно від того, хто саме формулює запит. Структурований покроковий промпт суттєво підвищує точність та послідовність відповідей LLM у складних наукових задачах - цей ефект підтверджений у дослідженнях із застосування LLM у фармацевтичних науках [24].

Застосування чотирирівневої структури промπτу фактично кодифікує знання досвідченого фахівця у відтворювану форму, доступну для будь-якого члена команди. Це частково вирішує одну з ключових проблем традиційного QRA - критичну залежність якості оцінки від наявності «незамінних» вузькопрофільних експертів.

2.3. Алгоритм верифікації результатів аналізу та відповідність принципам ALCOA+ в умовах регульованого виробництва

Будь-який інструмент, що використовується у регульованому фармацевтичному середовищі, має відповідати принципам цілісності даних. Ці

принципи зведені в акронімі ALCOA+: Attributable (атрибутивність), Legible (чіткість), Contemporaneous (одночасність), Original (оригінальність), Accurate (точність), а також Complete, Consistent, Enduring та Available. Цифровізація не знімає вимог ALCOA+, а навпаки - ускладнює їх виконання, оскільки електронні системи та алгоритми генерують нові вектори потенційних порушень цілісності даних [23].

Запропонований алгоритм верифікації складається з чотирьох послідовних етапів. На першому відбувається генерація первинного звіту LLM - структурованої FMEA-таблиці або реєстру ризиків у форматі, визначеному шаблоном промпту. Кожен висновок моделі має містити посилання на конкретний фрагмент вхідного документа (атрибутивність), а часова мітка генерації та версія використаної моделі фіксуються автоматично (одночасність). На другому етапі виконується автоматична перевірка на «галюцинації» - крос-верифікація згенерованих даних з базою знань: факти, що не підкріплені жодним джерелом у базі, автоматично позначаються для людського розгляду. Розрахункові значення RPN (добутки $P \times E \times D$ у FMEA) перевіряються окремим детермінованим скриптом, що забезпечує математичну точність незалежно від ймовірнісної природи LLM.

Третій етап - обов'язкова експертна перевірка (Human Review) кваліфікованим фахівцем з валідації. Ця ланка є структурною гарантією відповідності вимогам GAMP 5 [22] та принципу «Human-in-the-loop». Фахівець критично оцінює логіку аналізу, повноту виявлених ризиків та адекватність запропонованих заходів контролю — не просто підтверджує формальну наявність звіту. Четвертий етап - фіксація фінального звіту в системі управління якістю підприємства з кваліфікованим електронним підписом відповідно до вимог Настанови СТ-Н МОЗУ 42-4.0:2020 [5].

Одним із ключових ризиків впровадження ШІ в GxP-середовище є спокуса делегувати алгоритму не лише виконання аналізу, а й відповідальність

за його результат [21]. Запропонований чотириетапний алгоритм усуває цей ризик структурно: на кожному етапі передбачений незалежний механізм контролю, а кінцева відповідальність залишається за сертифікованим фахівцем. Принцип прозорості алгоритму - Explainable AI (XAI) - є додатковою вимогою: для цілей GxP-валідації модель має бути здатна обґрунтувати кожен сформований висновок, що відповідає загальній логіці доказової бази в регульованому середовищі.

Висновки до розділу 2

Розроблена методологія інтеграції великих мовних моделей (LLM) у процеси оцінки ризиків базується на таких ключових компонентах: 1. архітектурна модель «Human-in-the-loop» забезпечує регуляторно прийнятний баланс між автоматизацією та людським контролем, що відповідає очікуванням ICH Q9 (R1) та GAMP 5. 2. чотирирівнева система промпт-інжинірингу трансформує неформалізовані знання досвідчених фахівців у відтворені аналітичні алгоритми, підвищуючи узгодженість та об'єктивність оцінок ризиків. 3. алгоритм верифікації відповідно до принципів ALCOA+ забезпечує цілісність та юридичну силу документації, генерованої із залученням ІІІ, в регульованому середовищі. У сукупності ці компоненти формують методологічну основу для практичної апробації, що є предметом наступного розділу дослідження.

РОЗДІЛ 3 ПРАКТИЧНА АПРОБАЦІЯ МЕТОДОЛОГІЇ: ПОРІВНЯЛЬНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ГЕНЕРАЦІЇ ТАБЛИЦЬ ОЦІНКИ РИЗИКІВ ЗА ДОПОМОГОЮ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

3.1. Дизайн експериментального дослідження та характеристика вхідних даних

Практична частина дослідження присвячена верифікації методологічних положень Розділу 2 шляхом реального тестування LLM-асистованого FMEA на матеріалі виробничого процесу леналідоміду — препарату класу НРАPI (від англ. High Potency Active Pharmaceutical Ingredient — високоактивна діюча фармацевтична речовина) з класифікацією ОЕВ 4/5 (Occupational Exposure Band - діапазон допустимого виробничого впливу). Дизайн дослідження передбачав два ключових виміри: послідовне нарощування інформаційного контексту та еволюційне вдосконалення промптів від тесту до тесту.

Важливим методологічним зауваженням є різниця у шкалах оцінювання. Сучасна промислова практика та більшість актуальних настанов передбачають шкалу 1–10 для кожного з параметрів E (Effect), P (Probability) та D (Detectability), що забезпечує максимальний RPN (Risk Priority Number) = 1000 і дозволяє більш точно диференціювати ступінь ризику. Чим ширша шкала, тим тонше можна диференціювати ризики між собою — особливо у граничних випадках, коли різниця між «значущим» та «неприйнятним» ризиком принципово важлива для розстановки пріоритетів. Еталонна FMEA, складена валідаційним фахівцем, базується на шкалі 1–5 (максимальний RI = 125), що відповідає практиці підприємства на момент складання документа (2016 рік). Саме тому AI-моделі у всіх тестах використовували сучасну шкалу 1–10, а порівняння з еталоном проводилося за якісними критеріями: охоплення CQA та логіка обґрунтувань, а не пряме числове зіставлення RPN.

Важливим методологічним зауваженням є різниця у шкалах оцінювання. Сучасний стандарт у промисловій практиці та в більшості сучасних настанов передбачає використання шкали 1–10 для кожного з параметрів E, P та D, що забезпечує максимальний RPN = 1000 та дозволяє більш точно диференціювати ступінь ризику. Еталонна FMEA, складена валідаційним фахівцем, базується на шкалі 1–5 (максимальний RI = 125), що відповідає більш ранній практиці та внутрішнім стандартам підприємства (документ датований 2016 роком). Чим ширша шкала, тим точніше можна оцінити ризик і диференціювати пріоритети між різними відхиленнями. Саме тому AI-моделі в усіх тестах використовували шкалу 1–10, що є сучасним загальноприйнятим підходом. Для коректного порівняння з еталоном у роботі застосовано якісне зіставлення охоплення SQA та логіки обґрунтувань, а не пряме числове порівняння RPN. Центральним методологічним орієнтиром слугує еталонна FMEA, складена досвідченим валідаційним фахівцем. Саме порівняння з цим еталоном дозволяє об'єктивно оцінити, наскільки AI-генерована FMEA наближається до результату людини-експерта та яка конфігурація промпту забезпечує найбільше наближення. Ключова дослідницька гіпотеза: при чіткому, структурованому промпті та необхідних вхідних даних LLM здатна генерувати FMEA, порівнянну за якістю з результатом валідаційного фахівця.

Загалом проведено 12 тестових сесій на чотирьох великих мовних моделях та одній референтній. Для тестування обрано такі моделі: Qwen3-14B Thinking (компактна thinking-модель компанії Alibaba Cloud, 14 млрд параметрів, режим глибокого міркування); Qwen3-30B Thinking (масштабована thinking-модель, 30 млрд параметрів, підвищена здатність до багатокрокового аналізу); GLM 4.6v Flash (модель загального призначення компанії Zhipu AI, оптимізована для швидкого реагування); GLM 4.7 Flash (удосконалена версія попередньої моделі з покращеним розумінням галузевого контексту); GPT-OSS (референтна open-source модель на базі архітектури GPT). Для кожного тесту зафіксовано:

набір вхідних файлів, повну структуру промпту в оригінальному англomовному форматі, результуючу FMEA-таблицю та порівняльний аналіз з еталоном валідаційного фахівця. Детальний аналіз дозволяє простежити тенденцію наближення результатів штучного інтелекту до еталону в міру вдосконалення промпту та нарощування інформаційного контексту.

3.2. Еталонна FMEA валідаційного фахівця як база порівняння

Об'єктом валідації у цьому дослідженні є процес виробництва твердих капсул леналідоміду. Леналідомід - це імуномодулюючий препарат, що застосовується для лікування множинної мієломи та мієлодиспластичного синдрому. Препарат належить до класу НРАPI з класифікацією OEB 4/5, що означає вкрай низькі допустимі рівні виробничого впливу та потребує суворих вимог до контейнменту при виробництві. Виробничий процес включає такі стадії: зважування компонентів, приготування тритурації (попереднє змішування діючої речовини з лактозою), приготування бленду I (змішування з допоміжними речовинами), приготування лубрикованого бленду (з магнію стеаратом), наповнення твердих желатинових капсул та пакування у блістери. Саме ці стадії є об'єктом FMEA-аналізу як з боку валідаційного фахівця, так і в усіх тестах AI-моделей. До початку аналізу AI-результатів необхідно представити еталонний документ - FMEA, складену досвідченим фахівцем з валідації, що виступає критерієм якості для всіх наступних порівнянь. Розуміння структури та логіки еталону є ключем до оцінки ступеня наближення AI-моделей.

Характеристика еталонного документу.

Еталонна таблиця відрізняється від AI-генерованих FMEA за кількома принциповими ознаками. По-перше, застосовується шкала 1–5 (а не 1–10), де значення мають конкретну якісну прив'язку до статусу процесу. По-друге,

наявна колонка «Strength» - окремі рядки для різних дозувань (2,5–25 мг), що відображає реальну специфіку продукту. По-третє, обґрунтування Р базується не на загальних міркуваннях, а на конкретному факті: «процес ще не валідований, але експериментальні дані підтверджують відповідність якості». По-четверте, обґрунтування D посиляється на конкретні ІРС-процедури системи управління якістю (QS). По-п'яте, фокус - виключно на технологічних CQA (уніформність бленду, вологість, текучість, щільність, зовнішній вигляд капсул, розпадання, розчинення, рівномірність маси, уніформність вмісту) без ризиків утиліт, обладнання чи цілісності даних.

Ця «мінімалістична» стратегія є регуляторно обґрунтованою: фахівець зосередився на тому, що безпосередньо загрожує якості продукту та безпеці пацієнта в контексті невалідованого процесу, відклавши системні ризики на інші документи. Розуміння цієї логіки є ключем до коректної інтерпретації розбіжностей між еталоном та АІ-результатами.

Детальні результати аналізу, проведеного валідаційним фахівцем, наведені у Додатку А.

Ключові характеристики еталону для порівняння:

Охоплення: 5 категорій проблем (blend uniformity/water content, фізичні властивості, зовнішній вигляд, розпадання, розчинення, рівномірність маси, уніформність вмісту) по 2 стадіях (приготування бленду, наповнення капсул). Шкала значень RI: 20–100. Унікальна риса: диференціація за дозуванням (2,5 мг отримує вищу Р через технологічну складність наповнення). Стратегія: Е стабільно 4 для більшості ризиків — «шкодить здоров'ю пацієнта»; Р=5 при невалідованому процесі + без контролю; D=1–3 завдяки наявним ІРС-процедурам.

3.3. Серія тестів Qwen3-14B Thinking

3.3.1. Тест 1 - «Зворотна інженерія» за протоколом валідації

Модель: Qwen3-14B Thinking

Файли: Validation protocol (without Risk Assessment)

Структура та логіка промпу.

Базовий чотирикомпонентний промпт: Role Definition → Scoring Criteria (шкала 1–10, RPN = E×P×D) → Formatting Rules → Ask. Ідея Ask-секції — метод «зворотної інженерії»: для кожного тесту протоколу визначити який дефект він покликаний виявити та реконструювати FMEA. Принципова відмінність від еталону вже закладена у шкалі: 1–10 замість 1–5.

PROMPT — Test 1 (Qwen3-14B)

Role:

You are a Senior Validation Lead with 15 years of experience in GMP.

Specialization: QRM, Data Integrity, EU GMP Annex 15: Qualification and Validation.

Input Data Validation (Pre-Check):

Before generating the table, if critical information is missing, output:

"WARNING: Analysis conducted with limited data.

Missing description: [list] Assumed Worst Case Scenario for parameters:
[list]"

Task: Perform a Risk Analysis (FMEA) with detailed justification of ratings.

Scoring Formula: $RPN = E \times P \times D$

Effect (E): 1–3 Low (cosmetic), 4–6 Medium (process deviation), 7–10
Critical (recall/patient harm).

Probability (P): 1–3 Unlikely, 4–7 Possible (operator dependent), 8–10
Likely/repetitive.

Detectability (D): 10 No control, 7–9 Visual only, 4–6 Periodic IPC, 1–3
Automatic interlock.

Formatting: Actions column — [Action]. Resp: [Role]. Stage: [IQ/OQ/PQ]

$RPN \geq 100$: UNACCEPTABLE RISK. $RPN 50-99$: ALARP. $RPN < 50$: Risk
accepted.

Handling Missing Data: write 'Insufficient data. Conservative assumptions
applied.'

Score P and D at upper bound (Worst-Case). Tag row as (Data Gap).

Few-Shot Example: [Labeling | Missing label | E=9 | P=3 | D=2 | 54 | Increase
sensor checks]

Ask:

Read and learn the file completely.

Reverse-engineer a FMEA based on tests described in the Validation Protocol.

For each test determine: Failure Mode / Cause / Effect on patient.

Create an FMEA table

Згенерована моделлю FMEA-таблиця для даного сценарію представлена у Додатку Б.

Аналіз тенденції. Тест 1 є найвіддаленішим від еталону: охоплено лише 2 з 7 ключових CQA, відсутня стадія приготування бленду, немає диференціації за дозуванням. Метод «зворотної інженерії» без технологічного контексту виявився принципово недостатнім для відтворення результату фахівця.

3.3.2. Тест 2 - Auditor Mode. Додавання технологічного опису

Модель: Qwen3-14B Thinking

Файли: Validation protocol + Description of Manufacturing Process

Еволюція промпту відносно Тесту 1.

Системна частина незмінна. Ask-секцію повністю замінено на трьохкроковий «Auditor Mode»: STEP 1 - вивчити технологічний файл та фізику процесу; STEP 2 - аудит протоколу щодо достатності контролю CPP; STEP 3 - генерація ризиків через три сценарії (відсутність контролю, відмова обладнання, людський фактор). Нова ідея: оцінюється не наявність тесту, а адекватність контролю.

PROMPT — Test 2 | CHANGE: Ask-section → 3-step Auditor Mode (Technology + Protocol Audit + Risk Generation)

Role:

You are a Senior Validation Lead with 15 years of experience in GMP.

Specialization: QRM, Data Integrity, EU GMP Annex 15: Qualification and Validation.

Input Data Validation (Pre-Check):

Before generating the table, if critical information is missing, output:

"WARNING: Analysis conducted with limited data.

Missing description: [list] Assumed Worst Case Scenario for parameters:
[list]"

Task: Perform a Risk Analysis (FMEA) with detailed justification of ratings.

Scoring Formula: $RPN = E \times P \times D$

Effect (E): 1–3 Low (cosmetic), 4–6 Medium (process deviation), 7–10
Critical (recall/patient harm).

Probability (P): 1–3 Unlikely, 4–7 Possible (operator dependent), 8–10
Likely/repetitive.

Detectability (D): 10 No control, 7–9 Visual only, 4–6 Periodic IPC, 1–3
Automatic interlock.

Formatting: Actions column — [Action]. Resp: [Role]. Stage: [IQ/OQ/PQ]

$RPN \geq 100$: UNACCEPTABLE RISK. $RPN 50\text{--}99$: ALARP. $RPN < 50$: Risk
accepted.

Handling Missing Data: write 'Insufficient data. Conservative assumptions
applied.'

Score P and D at upper bound (Worst-Case). Tag row as (Data Gap).

Few-Shot Example: [Labeling | Missing label | E=9 | P=3 | D=2 | 54 | Increase
sensor checks]

Ask — INSTRUCTIONS (Auditor Mode):

STEP 1: TECHNOLOGY ANALYSIS — study physics of the process (mixing, heating, compression).

STEP 2: PROTOCOL AUDIT — evaluate CPP control adequacy.

STEP 3: RISK GENERATION — three scenarios:

Scenario 1 (Lack of Control): parameter in technology but only visual/undocumented control → HIGH D.

Scenario 2 (Failure Mode): CPP exists but no alarm/calibration test → RISK.

Scenario 3 (Human Factor): manual recording without automation → RISK

Зі структурою FMEA, згенерованою під час другої тестової сесії, можна ознайомитися у Додатку В.

Аналіз тенденції. Тест 2 демонструє перше суттєве наближення до еталону: стадія приготування бленду з'явилась у FMEA. Проте логіки аналізу ще різняться принципово: фахівець мислить категоріями «невідповідність CQA», тоді як AI - «недостатність контрольного заходу для CPP». Це є ключовим структурним розривом, що зберігатиметься до введення STEP 1 «Scientific Basis» у наступних тестах.

3.3.3. Тест 3 (перша спроба) — Тріангуляція: додавання Development Report

Модель: Qwen3-14B Thinking

Файли: Validation protocol + Manufacturing Process + Development File

Еволюція промπτу відносно Тесту 2.

Ask-секцію розширено до STEP 1–4. Ключове додавання - STEP 1 «Scientific Basis Analysis»: повний аналіз Development Report (гігроскопічність, поліморфізм, розподіл частинок, текучість, ексципієнти). STEP 4 - «Triangulation FMEA»: крос-референція трьох документів через сім сценаріїв ризику. Логіка

«ризик = розбіжність між документами» наближає підхід до методу, який застосовує фахівець при аудиті.

PROMPT – Test 3 | CHANGE: STEP 1 Scientific Basis added; STEP 4 Triangulation with 7 risk scenarios

Role:

You are a Senior Validation Lead with 15 years of experience in GMP.

Specialization: QRM, Data Integrity, EU GMP Annex 15: Qualification and Validation.

Input Data Validation (Pre-Check):

Before generating the table, if critical information is missing, output:

"WARNING: Analysis conducted with limited data.

Missing description: [list] Assumed Worst Case Scenario for parameters: [list]"

Task: Perform a Risk Analysis (FMEA) with detailed justification of ratings.

Scoring Formula: $RPN = E \times P \times D$

Effect (E): 1–3 Low (cosmetic), 4–6 Medium (process deviation), 7–10 Critical (recall/patient harm).

Probability (P): 1–3 Unlikely, 4–7 Possible (operator dependent), 8–10 Likely/repetitive.

Detectability (D): 10 No control, 7–9 Visual only, 4–6 Periodic IPC, 1–3 Automatic interlock.

Formatting: Actions column — [Action]. Resp: [Role]. Stage: [IQ/OQ/PQ]

RPN \geq 100: UNACCEPTABLE RISK. RPN 50–99: ALARP. RPN $<$ 50: Risk accepted.

Handling Missing Data: write 'Insufficient data. Conservative assumptions applied.'

Score P and D at upper bound (Worst-Case). Tag row as (Data Gap).

Few-Shot Example: [Labeling | Missing label | E=9 | P=3 | D=2 | 54 | Increase sensor checks]

Ask — INSTRUCTIONS (Auditor Mode) STEP 1–4:

STEP 1 SCIENTIFIC BASIS: study Development Report — hygroscopicity, polymorphism,

particle size, flowability, excipient reactivity.

STEP 2 TECHNOLOGY ANALYSIS: physics of process based on formula from Step 1.

STEP 3 PROTOCOL AUDIT: evaluate CPP controls against scientific risks of Step 1.

STEP 4 TRIANGULATION FMEA - cross-reference Dev.Data vs Technology vs Protocol:

Scenario 1 (Scientific Gap): Dev.Report risk not tested in Protocol → RISK.

Scenario 2 (Lack of Control Gap): Technology parameter missing in Protocol sampling plan → RISK.

Scenario 3 (Instrument Failure): CPP checked but no alarm/calibration backup → RISK.

Scenario 4 (Human Factor/Data Integrity): manual entries, no Audit Trail review → ALCOA+ violation.

Scenario 5 (Environmental Impact): hygroscopic API/excipients, no RH verification → RISK.

Scenario 6 (Utility Failure): power loss during encapsulation, no recovery test
→ RISK.

Scenario 7 (Containment/HPAPI): OEB 4/5, no SMEPAC seal/filter verification → RISK.

Згенерована таблиця оцінки ризиків на основі трьох джерел даних представлена в Додатку Г.

Аналіз тенденції. Тест 3 вперше залучає Development Report, однак методологічний розрив із фахівцем зберігається. Тріангуляція генерує ризики Instrument Failure та Data Integrity - категорії, яких немає в еталоні. Водночас базові CQA еталону (blend uniformity, dissolution, weight uniformity) в явному вигляді відсутні. Це вказує на необхідність перефокусування промпту з CPP-параметрів на CQA-відповідності.

3.3.4. Тест 4 (друга спроба) — Severity Override та HPAPI-контекст

Модель: Qwen3-14B Thinking

Файли: Validation protocol + Manufacturing Process + Development File

Еволюція промпту відносно Тесту 3.

Системна частина оновлена: Role розширена до HPAPI GMP Auditor; Project Context явно вказує OEB 4/5 та критичний фокус на контейнменті. Додано STEP 5 «Calibration & Reality Check» з трьома правилами: А — Equipment Reality (без checkweigher → $D \geq 5$ для маси), В — Severity Override (пил/витік OEB 4/5 → E=10), С - Historical Context (micro-stops → P=5–7 для пакування). Ці правила наближають промпт до специфіки HPAPI-продукту, але одночасно відводять аналіз ще далі від фокусу фахівця на технологічних CQA.

PROMPT — Test 4 | CHANGE: Role → HPAPI expert; Project Context OEB 4/5; STEP 5 Constraints added (Equipment Reality A, Severity Override B, Historical Context C)

Role: Senior Validation Lead and GMP Auditor, 15+ years, HPAPI manufacturing.

Expert in EU GMP Annex 15, ICH Q9, 21 CFR Part 11.

Project Context: Lenalidomide Hard Capsules. Hazard Class: OEB 4/5.

Critical Focus: Containment, cross-contamination, personnel safety, content uniformity.

[Scoring, Formatting, Handling Missing Data — unchanged]

Ask — INSTRUCTIONS (Auditor Mode) STEP 1–4:

STEP 1 SCIENTIFIC BASIS: study Development Report — hygroscopicity, polymorphism,

particle size, flowability, excipient reactivity.

STEP 2 TECHNOLOGY ANALYSIS: physics of process based on formula from Step 1.

STEP 3 PROTOCOL AUDIT: evaluate CPP controls against scientific risks of Step 1.

STEP 4 TRIANGULATION FMEA - cross-reference Dev.Data vs Technology vs Protocol:

Scenario 1 (Scientific Gap): Dev.Report risk not tested in Protocol → RISK.

Scenario 2 (Lack of Control Gap): Technology parameter missing in Protocol sampling plan → RISK.

Scenario 3 (Instrument Failure): CPP checked but no alarm/calibration backup → RISK.

Scenario 4 (Human Factor/Data Integrity): manual entries, no Audit Trail review → ALCOA+ violation.

Scenario 5 (Environmental Impact): hygroscopic API/excipients, no RH verification → RISK.

Scenario 6 (Utility Failure): power loss during encapsulation, no recovery test → RISK.

Scenario 7 (Containment/HPAPI): OEB 4/5, no SMEPAC seal/filter verification → RISK.

STEP 5: CALIBRATION & REALITY CHECK (Constraints):

A. Equipment Reality: no 100% online checkweigher → any weight variation $D \geq 5$.

B. Severity Override (OEB 4/5): dust/leakage = E 10; cross-contamination = E 10.

C. Historical Context: Micro-stops on blister foil → P = 5–7 for packaging steps

Результати ризиків з урахуванням критичних обмежень та пріоритетності безпеки наведено у Додатку Д

Аналіз тенденції. Тест 4 є найвіддаленішим від фокусу еталону в серії Qwen3-14B, хоча і є методологічно найбагатшим з точки зору охоплення систем безпеки. Це демонструє ключовий методологічний парадокс: більша кількість правил та обмежень у промпті не завжди наближає до результату фахівця — вона може переорієнтувати аналіз в інший регуляторний вимір.

3.3.5. Тест 5 - Четвертий рівень контексту: Qualification Report

Модель: Qwen3-14B Thinking

Файли: Validation protocol + Manufacturing Process + Development File + Qualification Report

Еволюція промпту відносно Тесту 4.

Промпт Тесту 4 збережено повністю. Єдина зміна - додано четвертий документ: кваліфікаційний звіт обладнання. Ask-секція спрощена до прямої команди «Read all files and write a Risk Assessment».

PROMPT - Test 5 | CHANGE: Added Qualification Report (Input Data 3);

Ask simplified to direct command

Role: Senior Validation Lead and GMP Auditor, 15+ years, HPAPI manufacturing.

Expert in EU GMP Annex 15, ICH Q9, 21 CFR Part 11.

Project Context: Lenalidomide Hard Capsules. Hazard Class: OEB 4/5.

Critical Focus: Containment, cross-contamination, personnel safety, content uniformity.

[Scoring, Formatting, Handling Missing Data - unchanged]

Ask — INSTRUCTIONS (Auditor Mode) STEP 1–4:

STEP 1 SCIENTIFIC BASIS: study Development Report - hygroscopicity, polymorphism,

particle size, flowability, excipient reactivity.

STEP 2 TECHNOLOGY ANALYSIS: physics of process based on formula from Step 1.

STEP 3 PROTOCOL AUDIT: evaluate CPP controls against scientific risks of Step 1.

STEP 4 TRIANGULATION FMEA - cross-reference Dev.Data vs Technology vs Protocol:

Scenario 1 (Scientific Gap): Dev.Report risk not tested in Protocol → RISK.

Scenario 2 (Lack of Control Gap): Technology parameter missing in Protocol sampling plan → RISK.

Scenario 3 (Instrument Failure): CPP checked but no alarm/calibration backup → RISK.

Scenario 4 (Human Factor/Data Integrity): manual entries, no Audit Trail review → ALCOA+ violation.

Scenario 5 (Environmental Impact): hygroscopic API/excipients, no RH verification → RISK.

Scenario 6 (Utility Failure): power loss during encapsulation, no recovery test → RISK.

Scenario 7 (Containment/HPAPI): OEB 4/5, no SMEPAC seal/filter verification → RISK.

STEP 5: CALIBRATION & REALITY CHECK (Constraints):

A. Equipment Reality: no 100% online checkweigher → any weight variation $D \geq 5$.

B. Severity Override (OEB 4/5): dust/leakage = E 10; cross-contamination = E 10.

C. Historical Context: Micro-stops on blister foil → P = 5–7 for packaging steps.

Ask: Read all files and write a Risk Assessment for Lenalidomide hard capsules.

Результати аналізу з використанням повного інформаційного контексту систематизовано у Додатку E

Аналіз тенденції. Тест 5 демонструє, що четвертий рівень контексту (Qualification Report) дозволив ідентифікувати ризик dissolution variability - єдиний рядок у серії Qwen3-14B, що частково перетинається з ключовим CQA еталону. Це підтверджує значущість повного контексту. Constraint A (без checkweigher) наближає підхід до виявності маси до логіки фахівця. Проте загальна орієнтація аналізу залишається на рівні HPAPI safety systems, а не product-specific CQA.

3.4. Серія тестів GLM 4.6v Flash

3.4.1. Тест 1 GLM 4.6v Flash - Model swap: той самий промпт, нова модель

Модель: GLM 4.6v Flash

Файли: Validation protocol + Manufacturing Process + Development File

Еволюція промпту: ідентичний промпт Qwen3-14B Тест 5. Зміна - лише модель (перший «model swap» тест).

PROMPT - GLM 4.6v Flash Test 1 | SAME as Qwen3-14B Test 5 | Model changed only

Role: Senior Validation Lead and GMP Auditor, 15+ years, HPAPI manufacturing.

Expert in EU GMP Annex 15, ICH Q9, 21 CFR Part 11.

Project Context: Lenalidomide Hard Capsules. Hazard Class: OEB 4/5.

Critical Focus: Containment, cross-contamination, personnel safety, content uniformity.

[Full prompt identical - STEP 1–5 with Constraints A/B/C + Data Submission

Ask: Read all files and write a Risk Assessment for Lenalidomide hard capsules.

Результати першого тестування моделі GLM 4.6v Flash наведені у Додатку Ж

Аналіз тенденції. GLM 4.6v Flash на ідентичному промпті вперше у дослідженні генерує ризик, причинно-наслідковий ланцюг якого (moisture uptake → blend segregation → content uniformity failure) відповідає логіці еталонного ризику #1. Це є першим реальним наближенням до CQA-фокусу фахівця — досягнутим завдяки кращій інтеграції GLM даних Development Report у обґрунтування Р та Е.

3.4.2. Тест 2 GLM 4.6v Flash - R&D Risk Alignment та Criticality Override

Модель: GLM 4.6v Flash

Файли: Validation protocol + Manufacturing Process + Development File

Еволюція промпту відносно GLM Тест 1.

Role розширена до FDF-manufacturing з додаванням 21 CFR Part 211 та FDA Process Validation. Project Context — шість Critical Focus Areas. До STEP 4 додано Scenario 7 (Equipment Cleaning). До STEP 5 додано правило D - R&D

Risk Alignment (вирівнювання E за початковою класифікацією R&D: Low→E 1–3, Medium→E 4–6, High→E 7–9) та Criticality Override ($E \geq 9 \rightarrow$ автоматично неприйнятний). Це правило потенційно наближає підхід до оцінки тяжкості у фахівця, де E залежить від відповідності нормативним CQA.

PROMPT - GLM 4.6v Flash Test 2 | CHANGES: Role→FDF; Scenario 7 Cleaning; STEP 5D R&D Risk Alignment; Criticality Override ($E \geq 9$ auto-unacceptable)

Role: Senior Validation Lead and GMP Auditor, 15+ years, FDF manufacturing.

Expert in EU GMP Annex 15, ICH Q9, 21 CFR Part 211, FDA Process Validation, EU GMP Chapter 1.

Project Context: Lenalidomide Hard Capsules. Hazard Class: OEB 4/5.

Critical Focus: Containment, Dissolution/Disintegration, Flowability, Blend Uniformity, Mass Balance, ALCOA+ Data Integrity.

[Scoring, Formatting unchanged; STEP 1–5 with all Constraints + R&D Alignment]

D. R&D Risk Alignment: align E with Initial Risk Assessment in Development Report.

Low \rightarrow E 1–3; Medium \rightarrow E 4–6; High \rightarrow E 7–9.

E=10 ONLY for Containment/Safety breaches (Rule B).

Criticality Override: $E \geq 9 \rightarrow$ AUTOMATICALLY Unacceptable regardless of RPN.

RPN ≥ 250 (or $E \geq 9$): Unacceptable. RPN 100–249: ALARP. RPN < 100 : Accepted

Згенерована моделлю FMEA-таблиця для даного сценарію представлена у Додатку И.

Аналіз тенденції. GLM Тест 2 є найближчим до фокусу фахівця у серії GLM 4.6v Flash. Правило R&D Risk Alignment усунуло «інфляцію E» та наблизило оцінки тяжкості до логіки еталону. Охоплення blend preparation розширилось: вологість (moisture→BU) та параметри змішування (mixing time→BU) разом покривають обидва аспекти еталонного ризику #1 для стадії Preparation of blend.

3.4.3. Тест 3 GLM 4.6v Flash - Без протоколу: технологія + Development Report + нормативні документи

Модель: GLM 4.6v Flash

Файли: Manufacturing Process + Development File + EU GMP Annex 15 + ICH Q9 + FDA Process Validation (без протоколу валідації)

Еволюція промπτу відносно GLM Тест 2.

Протокол валідації виключено, додано три нормативні документи. STEP 3 (Protocol Audit) виключено як неактуальний. STEP 5 спрощено до трьох правил (без Equipment Reality). Мета: аналіз без CPP-протоколу - наближення до логіки «технологічної FMEA».

PROMPT — GLM 4.6v Flash Test 3 | CHANGE: Validation Protocol REMOVED; Regulatory docs added (Annex 15, ICH Q9, FDA); STEP 3 removed; Equipment Reality constraint removed

Role: Senior Validation Lead and GMP Auditor, 15+ years, FDF manufacturing.

Expert in EU GMP Annex 15, ICH Q9, 21 CFR Part 211, FDA Process Validation, EU GMP Chapter 1.

Project Context: Lenalidomide Hard Capsules. Hazard Class: OEB 4/5.

Critical Focus: Containment, Dissolution/Disintegration, Flowability, Blend Uniformity, Mass Balance, ALCOA+ Data Integrity.

[STEP 1–2 + STEP 4 Scenarios 1–7 + STEP 5: Severity Override + Historical Context + R&D Alignmen

Результати генерації FMEA представлені в матеріалах Додатка К

Аналіз тенденції. GLM Тест 3 демонструє найбільше наближення до фокусу еталону у всій серії GLM 4.6v Flash: вперше охоплено 4 з 5 CQA-категорій еталону (BU, dissolution, weight, physical properties). Це пояснюється відсутністю протоколу, що змусило модель аналізувати технологічні ризики продукту, а не прогалини контролю CPP. Підхід до D залишається основним розривом.

3.5. Серія тестів Qwen3-30B Thinking та GLM 4.7 Flash

3.5.1. Qwen3-30B Тест 2–4 та GLM 4.7 Тест 2–3 — Антигалюцинаторні обмеження та абсолютна тяжкість

Моделі: Qwen3-30B Thinking (Тести 2–4), GLM 4.7 Flash (Тести 2–3)

Файли: Manufacturing Process + Development File + нормативні документи (без протоколу валідації)

Еволюція промпту відносно GLM 4.6v Flash Тест 3.

Серія Qwen3-30B та GLM 4.7 характеризується кардинальною зміною парадигми: від «maximal risk discovery» до «pragmatic physical FMEA». Введено блок STRICT EXCLUSIONS (5 явних заборон: утиліти/HVAC, механічні поломки кваліфікованого обладнання, крос-контамінація, хімічна деградація, відсутність даних ≠ критичний ризик). Тест 4 Qwen3-30B та Тести 2–3 GLM 4.7

додали CRITICAL RULE FOR SCORING EFFECT: E оцінюється абсолютно - «якщо дефект реалізується, яке наслідок для пацієнта», незалежно від даних R&D. Це дві найближчі до логіки фахівця серії тестів.

PROMPT — Qwen3-30B Test 4 & GLM 4.7 Tests 2–3 | CRITICAL CHANGES: STRICT EXCLUSIONS (anti-hallucination) + CRITICAL RULE: Effect is ABSOLUTE (never justify low E by R&D success)

Project Context: Lenalidomide Hard Capsules.

Product is already well-understood and validated before.

Primary Focus: Physical interactions — blend uniformity, water content, flowability,

density, capsule description/appearance, disintegration, dissolution, weight uniformity, content uniformity.

STRICT EXCLUSIONS (Anti-Hallucination Guardrails):

DO NOT generate risks related to:

1. Utilities & Facilities: HVAC, humidity, power outages.
2. Equipment Breakdowns (equipment is mechanically sound and qualified).
3. Cleaning & Cross-Contamination (cleaning validation already passed).
4. Extreme Chemical Degradation unrelated to the specific step.
5. Missing Data \neq Critical Risk: assume standard industry practices.

CRITICAL RULE FOR SCORING EFFECT (SEVERITY):

NEVER use Development Report to justify a LOW Effect score.

Effect (E) is ABSOLUTE - score based strictly on:

"If this physical defect actually happens → consequence for product/patient?"

Example: Chaotic mixing → E = 8–10 (wrong dosing). Do NOT say 'E=3 because R&D showed good uniformity.'

Task: Pragmatic step-by-step FMEA focused ONLY on physical process and product quality.

Instructions:

1. Analyze Composition & Development (physics only, NOT to lower E scores).
2. Iterate by Production Stage through Technology Steps.
3. Identify Physical Failure Modes:
 - Blending: over-lubrication → delayed dissolution; incomplete mixing → poor BU.
 - Encapsulation: powder bridging; dosator inconsistency → weight variation; empty capsules.
 - Packaging: incomplete blister sealing; pinholes in foil.
 - Capsule Filling non-conformances: appearance, disintegration, dissolution, weight uniformity, content uniformity.
4. Score: Effect (E) ABSOLUTE; P realistic 1–4 for validated product; D 1=automated, 10=no IPC.

Результати Qwen3-30B Тест 4 та GLM 4.7 Flash Тест 3 є найбільш наближеними до еталону у всьому дослідженні.

Порівняльні результати найбільш ефективних конфігурацій моделей наведені у Додатку Л

Аналіз тенденції - ключовий результат дослідження. Поєднання STRICT EXCLUSIONS (реалістичні P) + Effect is Absolute Rule (коректні E) + фізичний CQA-фокус (без відволікання на системи безпеки) забезпечило найбільше наближення до результату фахівця. GLM 4.7 Тест 3 вперше у дослідженні використовує формулювання 'Non-conformance of X' — ідентичне структурі еталону. Qwen3-30B Тест 4 та GLM 4.7 Тест 3 демонструють, що при правильно

сконструйованому промпті штучний інтелект здатна генерувати FMEA, яка за охопленням критичних показників якості (CQA) та логікою обґрунтувань є порівнянною з результатом валідаційного фахівця. Аналіз таблиці 3.13 виявляє кілька принципових закономірностей. По-перше, найдраматичніший стрибок якості відбувся при переході від Qwen3-14B Тестів 4–5 (де введення HPAPI-обмежень Severity Override парадоксально відвело аналіз від технологічних CQA) до GLM 4.6v Flash Тест 1 — де та сама модель при тому ж промпті вперше ідентифікувала причинно-наслідковий ланцюг humidity→blend segregation→CQA failure, ідентичний логіці еталонного ризику. По-друге, GLM 4.6v Flash Тест 3 продемонстрував парадоксальний ефект: усунення протоколу валідації з вхідних даних наблизило аналіз до технологічного фокусу фахівця, оскільки модель вимушено аналізувала ризики продукту, а не прогалини контрольних заходів. По-третє, перехід від тестів Qwen3-30B 2–3 до тесту 4 (введення правила «Effect is Absolute») наочно підтверджує, що одне додане правило у промпті здатне підняти оцінки E для ризиків уніформності вмісту з 3 до 9 при незмінних P та D — що якісно відповідає підходу фахівця. По-четверте, GPT-OSS на ідентичному промпті GLM 4.6v Flash Тест 3 показала нижчий ступінь відповідності через систематичну D-інфляцію, що свідчить про модель-специфічний ефект, який не усувається лише вдосконаленням промпту.

3.6. GPT-OSS — Референтний тест

Модель: GPT-OSS

Файли: Manufacturing Process + Development File + нормативні документи

Ідентичний промпт GLM 4.6v Flash Тест 3 (Role FDF + STEP 1–5 + Criticality Override + R&D Risk Alignment). Третій «model swap» тест.

PROMPT - GPT-OSS | SAME as GLM 4.6v Flash Test 3 | Third model swap

Role: Senior Validation Lead and GMP Auditor, 15+ years, FDF manufacturing.

Expert in EU GMP Annex 15, ICH Q9, 21 CFR Part 211, FDA Process Validation, EU GMP Chapter 1.

Project Context: Lenalidomide Hard Capsules. Hazard Class: OEB 4/5.

Critical Focus: Containment, Dissolution/Disintegration, Flowability, Blend Uniformity, Mass Balance, ALCOA+ Data Integrity.

[Full prompt identical to GLM 4.6v Flash Test 3 with R&D Risk Alignment]

Результати референтного тестування моделі GPT-OSS систематизовано у Додатку М

Аналіз тенденції. GPT-OSS є референтним прикладом моделі, що схильна до «D-інфляції» та систематичного переорієнтування на HPAPI safety systems. Це означає, що для даної моделі необхідне додаткове правило калібрування D (аналогічне STRICT EXCLUSIONS, але для оцінки виявності). Порівняно з GLM 4.7 Тест 3, GPT-OSS є більш «консервативною» у бік over-reporting та менш наближеною до логіки фахівця.

3.7. Загальна тенденція наближення до еталону

Сукупний аналіз 12 тестових сесій дозволяє кількісно оцінити тенденцію наближення AI-результатів до еталону за двома ключовими вимірами: охоплення CQA-категорій еталону та відповідність логіки обґрунтувань.

Характеристика моделей та обґрунтування вибору. Для дослідження обрано чотири моделі з різних архітектурних родин, що дозволяє виявити залежності

від характеристик моделі при однаковому промтті. Qwen3-14B Thinking є компактною thinking-моделлю компанії Alibaba Cloud з 14 млрд параметрів. Режим «thinking» дозволяє моделі розгортати внутрішній ланцюг міркувань перед формуванням відповіді, що підвищує здатність до структурованого аналізу складних документів. Обрана як базова модель для серії тестів 1–5 через оптимальний баланс між продуктивністю та швидкістю генерації. Qwen3-30B Thinking є масштабованою thinking-моделлю тієї ж родини з 30 млрд параметрів. Порівняння з Qwen3-14B у рамках тестів 2–4 дозволяє ізолювати ефект масштабу моделі при ідентичному промтті. GLM 4.6v Flash та GLM 4.7 Flash є моделями загального призначення компанії Zhipu AI. GLM 4.6v Flash відрізняється від Qwen3 відсутністю thinking-режиму, що дозволяє оцінити вплив архітектурного підходу до міркування. GLM 4.7 Flash є удосконаленою версією, яку порівнювали з GLM 4.6v Flash при ідентичних промптах для оцінки впливу версії моделі. GPT-OSS використовувалась виключно як референтна модель для одного model swap тесту з метою верифікації висновків на архітектурно відмінній системі.

Зведена таблиця ефективності всіх проведених тестів міститься у Додатку Н

Висновки до розділу 3

Порівняльне дослідження 12 тестових сесій із систематичним зіставленням з еталоном валідаційного фахівця підтверджує ключову гіпотезу та дозволяє зробити такі висновки:

1. Підтверджено центральну гіпотезу: при чіткому, структурованому промтті та необхідних вхідних даних LLM здатна генерувати FMEA, порівнянну за якістю з результатом валідаційного фахівця. Qwen3-30B Тест 4 та GLM 4.7 Тест 3 досягли 70–75% збігу за CQA-охопленням та 60% збігу за логікою

обґрунтувань з еталоном. GLM 4.7 Тест 3 вперше відтворив формулювання «Non-conformance of X» - ідентичне структурі еталону.

2. Встановлено оптимальну конфігурацію промпту для максимального наближення до еталону: STRICT EXCLUSIONS (реалістичні P) + Effect is Absolute Rule (коректні E) + фізичний CQA-фокус без відволікання на системи безпеки НРАРІ. Це поєднання забезпечує баланс між строгістю та реалістичністю, що характеризує роботу досвідченого фахівця.

3. Ідентифіковано основний структурний розрив між AI та фахівцем: фахівець мислить категоріями «невідповідність CQA специфікації → ризик для пацієнта», тоді як AI без відповідного налаштування схильна аналізувати «недостатність контрольного заходу для CPP → технологічний збій». Цей розрив усунуто в найкращих конфігураціях промпту.

4. Виявлено ефект «D-інфляції» (GPT-OSS, Qwen3-14B без обмежень): систематичне $D=8-10$ при відсутності явних тестів у вхідних документах, тоді як фахівець встановлює $D=1-3$ завдяки знанню наявних ІРС-процедур. Цей ефект потребує окремого правила калібрування D в промпті або надання моделі явних даних про ІРС-процедури.

5. Встановлено порядок значущості факторів якості: конфігурація промпту (вибір правил та обмежень) > повнота вхідних документів > вибір архітектурної моделі. «Model swar» тести підтвердили, що ідентичний промпт на різних моделях дає різний ступінь наближення до еталону, однак відмінності між моделями менш суттєві, ніж відмінності між версіями промпту.

6. Залишковий розрив (25–30% CQA-охоплення, відсутність диференціації за дозуванням) вказує на межі поточної парадигми LLM-асистованого FMEA та обґрунтовує необхідність обов'язкового Human Review як структурного елементу методології, а не опційного кроку.

ВИСНОВОКИ

У кваліфікаційній роботі проведено комплексне дослідження ефективності застосування великих мовних моделей (LLM) для автоматизації процесів оцінки ризиків якості (QRA) у фармацевтичній галузі. На основі отриманих результатів зроблено такі висновки:

1. Теоретичний аналіз еволюції концепцій управління ризиками якості підтвердив необхідність переходу від реактивних методів до предиктивних моделей. Встановлено, що класичні методи (FMEA, HACCP) в умовах Pharma 4.0 потребують інструментів для обробки великих масивів неструктурованих даних, якими є сучасні LLM.

2. Розроблено 4-рівневу архітектурну модель інтеграції ШІ в систему GxP, що базується на принципі «Human-in-the-loop». Це дозволяє використовувати когнітивні можливості моделей для структурування інформації, зберігаючи при цьому відповідальність за прийняття рішень за кваліфікованим фахівцем.

3. Доведено, що якість результатів ШІ-асистованого FMEA критично залежить від промпт-інжинірингу. Впровадження специфічних правил (STRICT EXCLUSIONS, Effect is Absolute Rule) дозволило досягти 70–75% збігу результатів ШІ з еталонною оцінкою валідаційного фахівця.

4. Виявлено та описано специфічні ефекти роботи ШІ, такі як «D-інфляція» та схильність до аналізу технологічних збоїв замість ризиків для пацієнта. Запропоновано методи калібрування моделей для усунення цих розбіжностей.

5. Встановлено, що найбільш ефективними для завдань QRA є моделі з розвиненими логічними можливостями (thinking models), такі як Qwen3-30B та GLM 4.7 Flash, які продемонстрували здатність до глибокого контекстуального аналізу складних технологічних регламентів.

6. Практичне впровадження розробленої методології дозволяє суттєво скоротити часові витрати на підготовку валідаційної документації при

збереженні високої якості аналізу, що безпосередньо сприяє підвищенні безпеки лікарських засобів.

Загалом, результати роботи підтверджують гіпотезу про високий потенціал LLM як інструменту підтримки прийняття рішень у фармацевтичній системі якості, за умови належної верифікації та дотримання принципів цілісності даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Лебединець В. О. Методологія формування систем управління якістю на підприємствах з виробництва лікарських засобів : дис. ... д-ра фармацевт. наук : 15.0.01 / НФаУ. Харків, 2016. 445 с.
2. ICH Q9(R1): Formal Revision to Quality Guideline for Quality Risk Management. 2023. URL: <https://www.regdevinc.com/ich-q9r1/> (Date of access: 20.03.2026).
3. Лебединець В. О., Коваленко С. М. Оцінювання, аналізування та управління ризиками для якості на фармацевтичному підприємстві. *Управління, економіка та забезпечення якості в фармацевції*. 2011. № 6(20). С.10–15. URL: https://nuph.edu.ua/wp-content/uploads/2015/04/UEK_620_2011_10-15.pdf (дата звернення: 20.03.2026).
4. Pharmaceutical Quality System Q10 : ICH Harmonised Tripartite Guideline / International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. 2008. URL: <https://database.ich.org/sites/default/files/Q10%20Guideline.pdf>. (Date of access: 20.03.2026).
5. Лікарські засоби. Належна виробнича практика : Настанова СТ-Н МОЗУ 42-4.0:2020. Вид. офіц. Київ : МОЗ України, 2020. URL: <https://surl.li/hqhifp> (дата звернення: 20.03.2026).
6. Лікарські засоби управління ризиками для якості (ICH Q9) : Настанова СТ-Н МОЗУ 42-4.2:2011. Вид. офіц. Київ : МОЗ України, 2011. URL: <https://surl.li/iitxuf> (дата звернення: 20.03.2026).
7. Державна Фармакопея України : в 3 т. / ДП «Український науковий фармакопейний центр якості лікарських засобів». 2-ге вид. Харків : ДП «Український науковий фармакопейний центр якості лікарських засобів», 2015. Т. 1. 1128 с.

8. Про лікарські засоби : Закон України від 04.04.1996 р. № 123/96-ВР. URL: <https://zakon.rada.gov.ua/laws/show/123/96-%D0%B2%D1%8> (дата звернення: 20.03.2026).
9. Деренська Я. О. Аналіз світових тенденцій в управлінні ризиками фармацевтичної галузі. *Механізм регулювання економіки*. 2022. № 3-4(97-98). С. 83–89.
10. Сагайдак-Нікітюк Р. В., Посилкіна О. В. Концептуальні підходи до управління ризиками в складних багатопотокових системах в умовах фармацевтичної галузі. *Український журнал клінічної та лабораторної медицини*. 2017. Т. 5, № 1. С. 8–12.
11. Ризик-орієнтований підхід до розробки та впровадження фармацевтичної системи якості на виробництві готових лікарських засобів із продукції in bulk / О. П. Баула та ін. *Управління, економіка та забезпечення якості в фармації*. 2021. № 1(65). С. 4–13.
12. Доброва В. Є., Зупанець К. О., Ратушна К. Л. Аналіз та дослідження ризиків втрати якості даних у клінічному випробуванні. *Клінічна фармація*. 2014. Т. 18, № 1. С. 4–10.
13. Алькема В. Г., Сумець О. М. Управління якістю продукції як стратегічний вектор забезпечення конкурентоспроможності фармацевтичної компанії на міжнародному ринку. *Вчені записки університету «КРОК»*. 2023. № 2(70). С. 83–92. URL: <https://snku.krok.edu.ua/index.php/vcheni-zapiski-universitetu-krok/article/view/594/621> (дата звернення: 20.03.2026).
14. Лікарські засоби. Належна виробнича практика : Настанова СТ-Н МОЗУ 42-4.0:2011. Київ : МОЗ України, 2011. 281 с.
15. Зупанець К.О. Аналіз проблем забезпечення належної роботи з даними, які отримані при проведенні клінічного дослідження / К.О. Зупанець //

- Управління, економіка та забезпечення якості в фармації. – 2014. – № 6 (38). – С. 60 – 66.
16. Суханова Н. В., Лебединець В. О., Ткаченко О. В. Дослідження діяльності з управління ризиками для якості лікарських засобів на етапі оптової торгівлі. *Соціальна фармація в охороні здоров'я*. 2020. Т. 6, № 3. С. 41–48.
17. Лебединець В. О., Коваленко С. М. Організація внутрішніх ризик-орієнтованих аудитів фармацевтичної системи якості. *Управління, економіка та забезпечення якості в фармації*. 2012. № 2(22). С. 21–26. URL: https://nuph.edu.ua/wp-content/uploads/2015/04/UEK_221-222012-21.pdf?utm_source (дата звернення: 20.03.2026).
18. Бутко М., Мурашко М., Назарко О. І. Сучасні аспекти цифровізації і впровадження штучного інтелекту в діяльність закладів вищої освіти України. *Економіка України*. 2022. URL: <https://nasu-periodicals.org.ua/index.php/economyukr/article/view/2022-11-4> (дата звернення: 20.03.2026).
19. Jia D., Wu Z. Application of Machine Learning in Enterprise Risk Management. *Security and Communication Networks*. 2022. № 4. P.11–12. URL: https://www.researchgate.net/publication/361033960_Application_of_Machine_Learning_in_Enterprise_Risk_Management (Date of access: 20.03.2026).
20. Потапова Т. М., Слесарчук В. Ю., Логвиненко Н. В. Світовий досвід та перспективи застосування штучного інтелекту в освітньому процесі та у фармацевтичній практиці. *Медична освіта*. 2024. № 1. С. 53–59. URL: https://ojs.tdmu.edu.ua/index.php/med_osvita/article/view/14582/13431 (дата звернення: 15.01.2026).
21. Горкуненко А. Б. Штучний інтелект у медицині: масштаби ринку та напрями розвитку. *Вісник соціальної гігієни та організації охорони здоров'я України*. 2025. № 1(103). С. 35–41. URL:

<https://ojs.tdmu.edu.ua/index.php/visnyk-gigieny/article/view/15340/14114>

(дата звернення: 15.01.2026).

22. GAMP 5: A Risk Based Approach to Compliant GxP Computerized Systems. URL: <https://xpdf4u.com/reviews/A1270E/998892/4991876-gamp-5-a-risk-based-approach-to-compliant-gxp-computerized-systems> (Date of access: 15.01.2026).
23. Борисюк І., Білик І., Криванич О. Стратегічне впровадження штучного інтелекту у маркетинг фармацевтичної продукції в Україні. *Сучасна медицина, фармація та психологічне здоров'я*. 2025. № 3(21). С. 170–176. URL: <https://journals.maup.com.ua/index.php/psych-health/article/view/5277/5592> (дата звернення: 15.01.2026).
24. Sakkala M., Hajal A. A. Large language models in drug delivery: A review of the current landscape and future perspectives. *Journal of Pharmaceutical Sciences*. 2026. Vol. 115(2). P. 104147. URL: [https://jpharmsci.org/article/S0022-3549\(25\)00601-X/abstract](https://jpharmsci.org/article/S0022-3549(25)00601-X/abstract) (Date of access: 15.01.2026).
25. Artificial intelligence in drug discovery and development / D. Paul et al. *Drug Discovery Today*. 2021. Vol. 26(1). P. 80–93. URL: <https://www.sciencedirect.com/science/article/pii/S1359644620304256?vi%3Dihub> (Date of access: 15.01.2026).
26. Mesko B. The role of artificial intelligence in precision medicine. *Expert Review of Precision Medicine and Drug Development*. 2017. Vol. 2(5). P. 239–241. DOI: 10.1080/23808993.2017.1380516.
27. Emerging of artificial intelligence and technology in pharmaceuticals: review / A. Sultana et al. *Future Journal of Pharmaceutical Sciences*. 2023. Vol. 9(1). DOI: 10.1186/s43094-023-00517-w.
28. Artificial intelligence for natural product drug discovery / M. W. Mallowney et al. *Nature Reviews Drug Discovery*. 2023. Vol. 22. P.

895–916. URL: https://www.nature.com/articles/s41573-023-00774-7?utm_source [ce](#) (Date of access: 15.01.2026).

29. Техніки промптингу для покращення використання великих мовних моделей / І. Ю. Юрчак та ін. *Computer systems and networks*. 2024. Vol. 6(2). P. 268–285. DOI: 10.23939/csn2024.02.268.
30. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. / J. White et al. 2023. DOI: 10.48550/arXiv.2302.11382.

ДОДАТКИ

ДОДАТОК А

Примітки: Шкала E, P, D - 1–5. Зелений фон таблиці позначає еталонний документ. Для dissolution та weight uniformity наведено окремі рядки для різних дозувань. RI=100 - high risk; RI=48–60 - significant risk; RI<30 - insignificant risk.

Таблиця 3.1 – Еталонна FMEA валідаційного фахівця

Steps of manufacturing	Problems and causes	Strength	E(1-5)	E Justification	P(1-5)	P Justification	D(1-5)	D Justification	RI =E×P×D	Actions
Preparation of blend	1. Non-conformance of blend uniformity, water content	All	4	It may lead to harm to patient health	5	Experimental data confirms blend may be produced in conformance with required quality. But blend manufacturing process is not validated.	5	Do not control during routine manufacturing process.	100(high risk)	See conclusion of risk evaluation under table 3.
Preparation of blend	2. Results of physical properties (flowability, density) not acceptable for capsule filling process	All	4	It may lead to harm to patient health	5	Experimental data confirms blend may be produced in conformance with required quality. But blend manufacturing process is not validated.	5	Do not control during routine manufacturing process.	100(high risk)	See conclusion of risk evaluation under table 3.

Steps of manufacturing	Problems and causes	Strength	E(1-5)	E Justification	P(1-5)	P Justification	D(1-5)	D Justification	RI =E×P×D	Actions
Capsule filling process	1. Non-conformance of capsules description/appearance	All	4	It may lead to harm to patient health (Critical defects of capsules)	5	Capsule filling process is not validated. Experimental data confirms capsules may be produced in conformance with required quality. Disintegration tests confirm QS requirements.	1	Appearance of each capsule batch is a part of routine manufacturing IPC (In-Process Control). Representative sample tested per QS requirements.	20(insignificant)	See conclusion of risk evaluation under table 3.
Capsule filling process	2. Non-conformance of disintegration	All	4	It may lead to harm to patient health	2	Capsule filling process is not validated. Experimental data confirms conformance with quality. Disintegration tests confirm QS requirements.	3	Representative sample of each "in-bulk" batch is tested for disintegration.	24(insignificant)	See conclusion of risk evaluation under table 3.
Capsule filling process	3. Non-conformance of dissolution	2.5 mg	4	May harm patient due inaccurate content of API in capsules	4	Capsule filling process is not validated. Risk of non-conformance of dissolution is bigger for 25 mg strength.	3	Representative sample of each "in-bulk" batch is tested for dissolution.	48(significant)	See conclusion of risk evaluation under table 3.
Capsule filling process	3. Non-conformance of dissolution	5–20 mg	4	May harm patient due inaccurate content of API in capsules	4	Capsule filling process is not validated.	3	Representative sample of each "in-bulk" batch is tested for dissolution.	48(significant)	See conclusion of risk evaluation under table 3.

Steps of manufacturing	Problems and causes	Strength	E(1-5)	E Justification	P(1-5)	P Justification	D(1-5)	D Justification	RI =E×P×D	Actions
Capsule filling process	3. Non-conformance of dissolution	25 mg	4	May harm patient due inaccurate content of API in capsules	5	Risk of non-conformance bigger for 25 mg strength.	3	Representative sample of each "in-bulk" batch is tested for dissolution.	60(significant)	See conclusion of risk evaluation under table 3.
Capsule filling process	4. Non-conformance of weight uniformity	2.5 mg	4	May harm patient due inaccurate content of API in capsules	5	Filling of smallest dose is more complicated. 2.5 mg has very small nominal capsule content (40 mg) vs 400 mg of 25 mg strength.	3	IPC by filled capsule weight is a part of routine manufacturing IPC. Representative sample tested for average mass.	60(significant)	See conclusion of risk evaluation under table 3.
Capsule filling process	4. Non-conformance of weight uniformity	5–25 mg	4	May harm patient due inaccurate content of API in capsules	4	Capsule filling process is not validated. Experimental data confirms conformance with required quality.	3	IPC by filled capsule weight is a part of routine manufacturing IPC.	48(significant)	See conclusion of risk evaluation under table 3.
Capsule filling process	5. Non-conformance of content uniformity	2.5 mg	4	May harm patient due inaccurate content of API in capsules	5	Filling of smallest dose is more complicated. 2.5 mg has very small nominal capsule content (40 mg).	3	Average representative sample of each "in-bulk" batch is tested for content uniformity.	60(significant)	See conclusion of risk evaluation under table 3.

Steps of manufacturing	Problems and causes	Strength	E(1-5)	E Justification	P(1-5)	P Justification	D(1-5)	D Justification	RI =E×P×D	Actions
Capsule filling process	5. Non-conformance of content uniformity	5–25 mg	4	May harm patient due to inaccurate content of API in capsules	4	Capsule filling process is not validated. Experimental data confirms conformance.	3	Average representative sample of each "in-bulk" batch tested for content uniformity.	48(significant)	See conclusion of risk evaluation under table 3.

ДОДАТОК Б

Результати. Таблиця 3.2.

Таблиця 3.2 – FMEA Qwen3-14B, Тест 1 (лише протокол валідації)

Steps of manufacturing	Problems and causes	E (1-10)	E Justification	P (1-10)	P Justification	D (1-10)	D Justification	RI	Actions
Non-conformance of disintegration	Failure Mode: Disintegration not meeting pharmacopeia standards Cause: Inconsistent tablet hardness or excipient incompatibility	8	Risk of incomplete drug release; harm to patient health	5	Operator-dependent compression settings; historical batch deviations observed	4	Periodic IPC checks using disintegration testing	160(UNACCEPTABLE)	Increase frequency of disintegration testing during PQ. Resp: QA. Stage: PQ
Non-conformance of dissolution	Failure Mode: Dissolution profile outside specified limits Cause: Inhomogeneous blend or improper tablet coating	8	Delayed/irregular drug release; reduced therapeutic efficacy	5	Blend uniformity issues due to scaling differences in batch sizes	4	Dissolution testing performed at OQ/PQ stages	160(UNACCEPTABLE)	Implement additional blend homogeneity checks during PQ. Resp: Eng. Stage: OQ
Holding time study failure	Failure Mode: Product quality deviation due to extended holding times Cause: Inadequate storage conditions or environmental exposure	5	Potential microbial growth or API degradation; may require reprocessing	3	Non-sterile FDF in D-class cleanrooms with controlled HVAC	1	Automated sampling and testing systems for hold time studies	15(ACCEPTABLE)	Acceptable risk. Continue holding time study on two batches per strength. Resp: QA. Stage: PQ

Продовження дод. Б

Порівняння Тест 1 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Охоплення CQA	5 проблем: blend uniformity, physical properties, appearance, disintegration, dissolution, weight uniformity, CU	3 проблеми: лише disintegration, dissolution, holding time. Відсутні: blend uniformity, appearance, weight uniformity, content uniformity
Шкала оцінок	E/P/D: 1–5. RI max = 100 (high risk)	E/P/D: 1–10. RPN max = 160. Різні шкали унеможливають пряме числове порівняння
Стадії виробництва	2 стадії: Preparation of blend + Capsule filling	Відсутня стадія Preparation of blend. Holding time — не в еталоні
Диференціація за дозуванням	Окремі рядки для 2.5/5/7.5/.../25 мг. P=5 для 2.5 мг (найскладніше наповнення)	Відсутня диференціація за дозуванням — узагальнені ризики

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Обґрунтування Р	Базується на факті 'процес не валідований + є дані розробки'	'Оператор-залежні налаштування; відхилення серій' — загальні міркування без прив'язки до статусу валідації
Обґрунтування D	Посилання на конкретні IPC-процедури QS (D=1–3)	D=4 для IPC-перевірок — адекватно, але без посилань на конкретні процедури
Загальний висновок	Повна FMEA з 11 рядками (різні дозування), фокус на технологічних CQA	Неповна FMEA: 3 рядки, відсутні ключові CQA еталону. Тенденція: дуже далеко від еталону

ДОДАТОК В

Таблиця 3.3 – FMEA Qwen3-14В, Тест 2 (протокол + технологічний опис)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Blend Preparation(mixing time)	Mixing time deviationCause: Operator error in timer setting or equipment failure	7	Inadequate mixing may lead to content uniformity issues, risking underdosing/overdosing	4	Operator-dependent settings; no automated validation during PQ	8	Visual inspection only during IPC; no real-time monitoring	224(UNACCEPTABLE)	Install automated timers with interlocks. Resp: Eng. Stage: OQ
Blend Preparation(rotation speed)	Rotation speed deviationCause: Motor failure or calibration drift	6	May cause uneven blending; potential rework/rejection	4	No documented sensor testing in protocol for rotation speed	8	Manual checks only; no automated feedback during PQ	192(UNACCEPTABLE)	Implement RPM sensors with calibration logs. Resp: QA. Stage: OQ
Trituration(mixing time)	Mixing time deviationCause: Manual timer errors or equipment downtime	6	Incomplete trituration may affect API dispersion; potential batch failure	3	Rare events based on historical data	9	No controls for mixing time in validation protocol	162(UNACCEPTABLE)	Add automated timing with failure alarms. Resp: Eng. Stage: PQ

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Filling Process(capsule height)	Capsule height deviationCause: Dosator calibration drift or operator error	6	May result in capsule weight variation; non-compliant with USP dissolution criteria	5	Operator-dependent adjustments; no auto-checks documented	7	Visual inspection only; no real-time monitoring for height	210(UNACCEPTABLE)	Integrate vision system for capsule height with OQ validation. Resp: Eng. Stage: PQ
Stability Testing	Incomplete sampling pointsCause: Manual record errors or omissions	4	Minor impact on stability data; may require retesting but not patient safety	3	Sampling points documented in protocol	6	Periodic IPC checks during stability studies	72(ALARP)	Risk accepted per EMA guidelines and bracketing approach

Продовження дод. В

Порівняння Тест 2 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	АІ (даний тест)
Охоплення CQA	Blend uniformity, physical properties, appearance, disintegration, dissolution, weight uniformity, CU	З'явилась стадія Blend Preparation (mixing time, rotation speed). Але: відсутні appearance, disintegration, weight uniformity, CU. Охоплено ~30% CQA еталону
Фокус аналізу	Технологічні CQA продукту	Фокус змістився на CPP-параметри обладнання (час/швидкість змішування), а не на кінцеві CQA. Фахівець мислить 'що може не відповідати специфікації', АІ — 'який параметр не контролюється'
Значення D	D=1–3 завдяки конкретним IPC QS	D=7–9 через відсутність автоматички — надмірно консервативно; не відображає наявні IPC-процедури системи якості
Диференціація за дозуванням	Присутня для dissolution, weight uniformity, CU	Відсутня повністю

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Тенденція vs Тест 1	—	Поліпшення: з'явилися стадія бленду та технологічні ризики. Але AI аналізує 'параметри обладнання', а не 'кінцева якість продукту' — принципова розбіжність з логікою фахівця

ДОДАТОК Г

Таблиця 3.4 – FMEA Qwen3-14B, Тест 3 (протокол + технологія + Development Report)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Mixing Time in Trituration	Insufficient mixing due to deviationCause: Manual timing, no automated monitoring	6	Process deviation; may cause non-uniform blend but not direct patient harm	5	Manual timing could lead to operator error; historical data not provided	7	Relies on visual inspection or batch records only	210(UNACCEPTABLE)	Implement automated timer with alarm. Resp: Eng. Stage: OQ
Rotation Speed in Trituration	Sensor failure causing incorrect speedCause: No redundant sensors or calibration checks	6	Non-uniform blend; may require rework	7	Operator-dependent monitoring; no mention of alarm testing	8	Visual inspection only during PQ; no automatic interlock	336(UNACCEPTABLE)	Add redundant sensors and daily calibration checks. Resp: Eng. Stage: IQ
Blend Step 1 Mixing Time	Premature stoppage due to manual oversightCause: No automated verification of 30 min	7	Incomplete blending may lead to disintegration or content uniformity issues	5	Operator-dependent timing; no real-time monitoring evidence	7	Relying on batch records with no in-process checks	245(UNACCEPTABLE)	Install automated timer and 100% vision system. Resp: Eng. Stage: OQ

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Blend Step 2 Counter-Clockwise	Extra mixing → over-blendingCause: Manual step not followed; no interlock	6	Potential particle size changes; may require rework	4	Rare operator error; historical data not provided	8	No controls; depends on visual inspection of final blend	192(UNACCEPTABLE)	Program PLC to disable counter-clockwise after 30 min. Resp: Eng. Stage: IQ
Sampling Points for Batches	Insufficient sampling pointsCause: Limited validation data for small batches	4	Cosmetic defect; no patient harm if minor deviations	6	Sampling strategy not validated for target batch sizes	9	No control for small batch validation	216(UNACCEPTABLE)	Validate sampling protocol with statistical analysis. Resp: QA. Stage: PQ
Rotation Speed in Blend Step	Equipment failure → speed deviationCause: No alarm or interlock for speed drift	5	Process deviation; may require rework	4	Reliable equipment; conservative assumption applied	3	Assumed automatic interlock from calibration records	60(ALARP)	Document existing interlocks in SOPs. Resp: QA. Stage: OQ

Порівняння Тест 3 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	АІ (даний тест)
Ключові еталону CQA	Blend uniformity, water content, physical properties, appearance, disintegration, dissolution, weight uniformity, CU	Жоден з CQA еталону не ідентифікований як окремий рядок FMEA. Замість цього — параметри CPP обладнання (час/швидкість змішування, сенсори)
Логіка аналізу	'Невідповідність специфікації CQA → ризик для пацієнта'	'Відсутність контролю CPP → технологічний ризик'. Логіка ближча до engineering FMEA, ніж до pharmaceutical QRA
Наукова база	Р обґрунтовується статусом валідації та даними розробки	STEP 1 введено, але наукова база Development Report не знаходить відображення у кінцевих рядках FMEA — обґрунтування E/P залишилися загальними
Значення RI/RPN	RI 20–100 (шкала 1–5)	RPN 60–336 (шкала 1–10). Максимальний RPN=336 для сенсора тритурації — технологічний ризик, якого немає в еталоні

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Тенденція vs Тест 2	—	Зростання кількості рядків (6 vs 5), поява ризиків Instrument Failure. Але розрив з еталоном зберігається: AI аналізує 'як контролюється CPP', фахівець — 'яка CQA може не відповідати'

ДОДАТОК Д

Таблица 3.5 – FMEA Qwen3-14B, Тест 4 (Severity Override + Constraints)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Pre-Validation Gap	Missing Development Data on API hygroscopicity/polymerization risks	10	API degradation affects content uniformity & shelf life	7	Worst case: No data to rule out instability under storage conditions	5	No protocol tests for moisture exposure or polymorphic conversion	350(UNACCEPTABLE)	Data Gap - Complete Development Report required. Validate storage conditions per ICH Q1A
Equipment Qualification	Missing calibration certificates for capsule filling machine & scales	9	Weight variation risks impact content uniformity	5	Assumed moderate risk due to prior OOS trends	4	Manual IPC checks only; no automated checkweigher (Constraint A)	180(UNACCEPTABLE)	Revalidate equipment calibration. Implement automated weight verification
Environmental Controls	No protocol verification of humidity limits for hygroscopic API	10	API degradation risk under worst-case humidity; impacts shelf life	6	Worst case: Storage in non-classified room	3	No protocol tests for environmental monitoring	180(UNACCEPTABLE)	Integrate RH monitoring into validation protocol. Use SMEPAC for containment

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Containment & Cross-Contamination	No verification of seals/filters for OEB 4/5 API	10	Dust leakage = safety hazard + cross-contamination (Severity Override)	7	High potency strict containment; micro-stops increase exposure risk	2	No protocol tests for filter integrity or residue analysis	140(UNACCEPTABLE)	Perform SMEPAC testing and establish cleaning validation for OEB 4/5
SOP & Data Integrity	In-bulk stability data shows variability in dissolution/related substances	8	Content uniformity risks; impacts batch release criteria	5	Worst case: No root cause analysis for trends	4	Manual review of data tables only	160(UNACCEPTABLE)	Investigate dissolution variability. Add statistical analysis to stability protocol
Power Failure Recovery	No test for power loss during encapsulation	9	Unknown product status; risk of capsule damage or contamination	4	Low frequency but high impact if occurs	3	No protocol for recovery validation	108(UNACCEPTABLE)	Include power failure simulations in OQ. Verify capsule integrity post-recovery

Продовження дод. Д

Порівняння Тест 4 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Перетин CQA з еталоном	Content uniformity, dissolution як основні CQA	'SOP & Data Integrity' містить dissolution variability — перший рядок що частково перетинається з dissolution CQA еталону. Але контекст інший: варіабельність стабільності vs невалідований процес
Нові категорії ризиків	Відсутні в еталоні	Pre-Validation Gap, Equipment Qualification, Containment, Power Failure — категорії безпеки та систем, відсутні в еталоні фахівця
Логіка Р	Р базується на статусі 'процес не валідований'	Р базується на 'worst case' та 'assumed risk' — більш консервативна та менш обґрунтована фактичними даними
Відповідність регуляторній практиці	Відповідає стандартному QRA для PV-протоколу	Відповідає аудиту систем безпеки НРАPI, але виходить за межі типового PV Risk Assessment

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Тенденція vs Тест 3	—	Ризик Pre-Validation Gap (RPN=350) є методологічно цінним сигналом, але загальне відхилення від фокусу еталону збільшилось. НРАPI-обмеження відвели аналіз від технологічних CQA

ДОДАТОК Е

Таблиця 3.6 - FMEA Qwen3-14B, Тест 5 (повний контекст + Qualification Report)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Pre-Validation Gap	Missing Development Data on API hygroscopicity/polymerization risks	10	API degradation affects content uniformity & shelf life	7	Worst case: No data to rule out instability under storage conditions	5	No protocol tests for moisture or polymorphic conversion	350(UNACC EPTABLE)	Data Gap – Complete Development Report required. Validate storage conditions per ICH Q1A
Equipment Qualification	Missing calibration certificates for capsule filling machine & scales	9	Weight variation risks impact content uniformity	5	Assumed moderate risk due to prior OOS trends in encapsulation lines	4	Manual IPC checks only; no automated checkweigher	180(UNACC EPTABLE)	Revalidate equipment calibration. Implement automated weight verification
Environmental Controls	No verification of humidity limits for hygroscopic API	10	API degradation risk under worst-case humidity	6	Worst case: Storage in non-classified room	3	No protocol tests for environmental monitoring	180(UNACC EPTABLE)	Integrate RH monitoring. Use SMEPAC for containment

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Containment & Cross-Contamination	No verification of seals/filters for OEB 4/5	10	Dust leakage = safety hazard + cross-contamination (Severity Override)	7	Strict containment required; micro-stops increase exposure risk	2	No tests for filter integrity or residue analysis	140(UNACC EPTABLE)	Perform SMEPAC testing and cleaning validation for OEB 4/5
SOP & Data Integrity (from Qualification Report)	In-bulk stability data shows variability in dissolution/related substances	8	Content uniformity risks; impacts batch release	5	Worst case: No root cause analysis for trends	4	Manual review of data tables only	160(UNACC EPTABLE)	Investigate dissolution variability. Add statistical analysis
Stability Testing — Equipment	Equipment failure causing speed deviation; no alarm or interlock	5	Process deviation; rework needed but not patient harm	4	Reliable equipment; conservative assumption	3	Assumed automatic interlock from calibration records	60(ALARP)	Document existing interlocks in SOPs. Resp: QA. Stage: OQ
Power Failure Recovery	No test for power loss during encapsulation	9	Unknown product status; risk of capsule damage	4	Low frequency but high impact	3	No protocol for recovery validation	108(UNACC EPTABLE)	Include power failure simulations in OQ

Продовження дод. Е

Порівняння Тест 5 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	АІ (даний тест)
Нові ідентифіковані ризики	Dissolution (per strength), weight uniformity, CU як ключові CQA	Рядок 'SOP & Data Integrity' містить dissolution variability — унікальний для Тесту 5, доступний лише завдяки Qualification Report
Перетин з еталоном	Dissolution + weight uniformity як найважливіші CQA для безпеки пацієнта	Dissolution variability з'явилась у FMEA, але через призму 'стабільності' а не 'невалідованого процесу'. Частковий перетин
Значення D для weight risks	D=3 (IPC by filled capsule weight є стандартним IPC)	D=4 (Constraint A: без checkweigher). Ближче до еталону, ніж в попередніх тестах — Constraint A частково відтворює логіку еталону щодо D
Тенденція Тест 4 vs	—	Поява ризику на основі Qualification Report — крок вперед. Constraint A наближає D для ризиків маси. Загальний профіль залишається НРАРІ-орієнтованим, а не CQA-орієнтованим

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Загальний висновок	Фокус: технологічні CQA + статус валідації	Фокус: системи безпеки HPAPI + Data Integrity + Qualification gaps. Різний рівень аналізу, але обидва регуляторно обґрунтовані

ДОДАТОК Ж

Таблиця 3.7 – FMEA GLM 4.6v Flash, Тест 1 (ідентичний промт)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Blend I Mixing - polymorph conversion	Polymorph conversion of API during tumble blending due to heatCause: No temperature monitoring or post-mixing polymorph testing	8	Polymorphic form change impacts API stability and content uniformity	5	Mixing generates localized heat; thermal data not provided	6	No routine XRD analysis; final testing occurs after filling	240(U NAC CEPT ABLE)	Conduct in-process XRD analysis. Verify tumble blender temperature limits. Resp: Process Dev.
Blend I Mixing — time deviation	Mixing time deviation → poor content uniformityCaus e: Sampling only after lubrication; no intermediate checks	7	CU failure impacts dosage accuracy	5	Operator error or equipment variability	4	IPC post-lubrication; deviations during blending may be undetected	140(U NAC CEPT ABLE)	Add intermediate BU sampling at key time points. Real-time mixing speed monitoring. Resp: QA

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Blend Mixing — I humidity	Humidity failure due to lactose hygroscopicityCause: No room humidity monitoring during blend/filling	9	Lactose moisture uptake → blend segregation → CU issues	6	Environmental humidity fluctuations common in manufacturing	5	No RH sensors/checks in protocol; only final flowability/weight tests	270(U NAC CEPT ABLE)	Install real-time RH monitoring with alarms. Resp: Facilities Engineering
Capsule Filling — speed variation	Speed variation → weight deviationCause: No automated detection of speed drifts	8	Weight deviation impacts CU and dosage accuracy	6	Operator machine speed adjustments are common	5	Manual average mass checks at start/middle/end only	240(U NAC CEPT ABLE)	Integrate automated checkweigher with real-time speed correlation. Resp: Equipment Supplier
Capsule Filling — power failure	Power failure during fillingCause: No power recovery testing in protocol	10	Power loss risks uncontrolled product release, cross-contamination (OEB 4/5)	4	Power outages are rare but possible	3	No post-failure inspection of filled capsules	120(U NAC CEPT ABLE)	Add power failure recovery procedure. Verify backup power systems. Resp: Maintenance

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Containment — filter integrity	Filter integrity failureCause: No routine filter leak testing for OEB 4/5	10	Dust escape → occupational exposure and cross-contamination	5	Filters degrade over time or cleaning cycles	2	Only blister integrity test; no filter pressure drop/leak checks	100(U NAC CEPT ABLE)	Implement routine filter integrity testing (pressure drop monitoring). Resp: QA
Data Integrity — manual entry	Manual entry of critical parameters; no audit trail in protocol	7	Human error in data recording impacts process control	5	Operator fatigue/miscommunication common	6	No automated validation; only final batch record review	210(U NAC CEPT ABLE)	Require electronic logging. Implement ALCOA+ audit trail review. Resp: Data Integrity Team

Продовження дод. Ж

Порівняння GLM Тест 1 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Охоплення CQA еталону	Blend uniformity (water content, flowability), dissolution, weight uniformity, CU, appearance	Blend I Mixing містить ризики, пов'язані з BU та CU. Рядок humidity → blend segregation → CU — найближчий до еталонного ризику #1 (blend uniformity)
Ключове наближення	Preparation of blend — blend uniformity/physical properties → RI=100	GLM ідентифікує humidity → lactose segregation → CU (RPN=270). Причинно-наслідковий ланцюг ідентичний еталону, але формулювання різне
Якість обґрунтувань	Обґрунтування базується на конкретних фактах (процес не валідований, XRD зафіксовано)	GLM: 'Environmental humidity fluctuations common' — загальне. Еталон: 'experimental data confirms but process not validated' — специфічне

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Значення P та D	E=4, P=5, D=5 для blend → RI=100; D=1–3 завдяки наявним IPC	E=9, P=6, D=5 → RPN=270. D=5 (без checkweigher) ближче до еталонного підходу, ніж D=8–9 попередніх тестів
Тенденція vs Qwen3-14B Тест 5	—	GLM на тому ж промпті генерує ближчий до фокусу еталону ризик: humidity→CU відповідає логіці blend uniformity еталону. Перше реальне наближення до CQA-фокусу

ДОДАТОК И

Таблица 3.8 – FMEA GLM 4.6v Flash, Тест 2 (R&D Risk Alignment)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Blend Preparation — Moisture Control	Hygroscopic API/excipients; no humidity monitoring in protocol → risk of degradation/sticking	9	Hygroscopic API/excipients; protocol lacks RH monitoring → degradation/sticking	6	Likely due to powder handling and environmental variability	7	No RH verification in protocol; manual checks insufficient	378(U NACC EPTA BLE)	Implement real-time humidity sensors in blend/filling areas with alarms. Update protocol
Blend Preparation — Mixing Time/Speed	Verification missing for Trituration/Blend mixing parameters	5	R&D rates Dissolution in blending as 'High' → parameter deviation could impact quality	4	Moderate; equipment may drift without verification	7	CPP defined in technology but no sampling plan to verify during validation runs	140(U NACC EPTA BLE)	Add process parameter verification checks via equipment logs or direct measurement
Capsule Filling — Manual IPCs No Audit Trail	Manual IPCs without Audit Trail reviewCause: Manual weight checks and CU tests lack audit trail requirement	9	Data integrity is critical per GMP; manual entries prone to errors/manipulation	5	Typical operator error or data falsification risk in high-volume processes	6	Manual weight/CU tests lack audit trail review → no automated detection of anomalies	270(U NACC EPTA BLE)	Mandate weekly review of manual IPC logs by QA. Implement electronic data capture
Capsule Filling — Utility Failure (Power)	Power loss during filling; no recovery test → unknown product status	8	Power loss could lead to unaccounted HPAPI material release; OEB 4/5 risk	5	Common utility interruptions in manufacturing environments	7	No power failure recovery test → unknown impact on product status	280(U NACC EPTA BLE)	Add 'Power Failure Recovery' test to protocol: simulate outage, document

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
									product handling procedure
Lubrication Step — Temperature Control	Temperature control not monitored; no thermal monitoring in specification	4	R&D rates lubrication Dissolution as 'Low' but heat could affect API stability	3	If equipment controlled, temperature spikes may occur during mixing	6	No temperature monitoring; logs do not capture thermal excursions	72(AL ARP)	Add temperature sensors to lubrication blender; include pre/post-mixing temperature checks
Blend Uniformity — Sampling — Small Batches	Reduced sampling for small batches; scale effects may cause higher RSD	6	R&D rates CU in blending as 'High' → undersampling may miss variability	4	Small batches common; may have higher RSD due to scale effects	5	Sampling plan includes specific points for small batches; deviations detectable via QC	120(U NACC EPTA BLE)	Maintain full 10-point sampling for all blend sizes; retain original sampling frequency

Продовження дод. И

Порівняння GLM Тест 2 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Відповідність CQA blend uniformity	Blend uniformity/water content → RI=100 (E=4, P=5, D=5)	Blend Preparation — Moisture Control (RPN=378) + Mixing Time/Speed (RPN=140) разом охоплюють обидва аспекти еталонного ризику #1
Диференціація складності процесу	P=5 для 2.5 мг через технологічну складність наповнення	Blend Uniformity Sampling відзначає вищий RSD для малих серій — часткова аналогія, але без прив'язки до конкретних дозувань
R&D Risk Alignment — відповідність еталону	E стабільно 4; P=2–5 залежно від статусу валідації	R&D High→E=5–7; Low→E=4: ближче до еталонної E=4, ніж попередні тести де E=8–10
Новий ризик Blend Uniformity Sampling	Відсутній як окремий рядок	Ідентифікація статистичної неадекватності відбору проб — цінний новий аспект, що відповідає духу еталонного підходу

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Тенденція vs GLM Тест 1	—	R&D Risk Alignment знизив E для менш критичних ризиків (lubrication E=4) — ближче до еталонної логіки. RPN=378 для вологості відображає реальний пріоритет безпеки продукту

ДОДАТОК К

Таблица 3.9 – FMEA GLM 4.6v Flash, Тест 3 (без протоколу, нормативна база)

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Trituration & Blend Mixing	Inadequate mixing time/speed → poor content uniformity of API	8	High risk to CU (CQA rated 'High' by R&D)	5	Moderate likelihood if parameters not strictly monitored	6	Sampling plan includes BU testing but may miss real-time monitoring	240(UNACCEPTABLE)	Develop SOP with real-time mixing speed/time monitoring; QA review of blending logs
Lubricated Blend — dissolution	Over-lubrication or segregation → dissolution profile affected	7	High risk to Dissolution (CQA rated 'High')	5	Process parameters defined but not verified during PQ	6	In-process dissolution test on blend but limited sampling frequency	210(UNACCEPTABLE)	Add in-line PAT for lubricant level; increase blend dissolution sampling frequency
Capsule Filling Speed Variation	Manual speed changes → weight deviation beyond spec	5	Medium risk to Capsule Weight (CQA rated 'Medium')	6	Historical micro-stops affect consistent operation	5	Manual checkweighing; audit trail compliance required but may lack review	150(UNACCEPTABLE)	Implement automated weight checkweigher or enhanced manual verification with

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
									electronic logging
Packaging Blister Integrity — micro-stops	Micro-stops in blister foil → dust generation/HPAP I exposure	10	Critical containment breach for OEB 4/5 (Containment Override)	7	Frequent micro-stop events historically on this line	2	Visual inspection only; no automated detection or leak testing	140(UNA CCEPTAB LE)	Conduct SMEPAC test on blister sealing; implement real-time blister defect monitoring
Power Failure During Filling	Unexpected power outage → product status unknown	9	High risk to product quality and safety	4	Facility has backup systems but occasional outages possible	1	No protocol for recovery testing or product segregation post-failure	144(UNA CCEPTAB LE)	Develop power failure SOP with immediate stoppage and product hold/analysis
Blending Humidity Control	Inadequate RH monitoring → API moisture uptake or lactose caking	8	Medium-High risk due to hygroscopic excipients affecting flowability and CU	5	Facility may not have strict RH control for dry blending	4	RH sensors present but not included in sampling plan	160(UNA CCEPTAB LE)	Install real-time RH monitoring with alarms; integrate into environmental monitoring system

Продовження дод. К

Порівняння GLM Тест 3 з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Охоплення еталону CQA	Blend uniformity, physical properties, dissolution, weight uniformity, CU	Тест 3 охоплює: BU/CU (Trituration & Blend I), dissolution (Lubricated Blend), weight (Capsule Filling Speed), humidity→BU (Blending RH). 4 з 5 CQA-категорій еталону присутні у певній формі
Dissolution — найближче наближення	Non-conformance of dissolution: E=4, P=4–5, D=3, RI=48–60	Lubricated Blend — dissolution: E=7, P=5, D=6, RPN=210. Той самий CQA, але E вищий (без обмеження по R&D) та D вищий (без посилання на наявний IPC dissolution test)
Weight uniformity — наближення	Non-conformance of weight uniformity: E=4, P=4–5, D=3, RI=48–60	Capsule Filling Speed Variation: E=5, P=6, D=5, RPN=150. E та D ближче до еталону, ніж у попередніх тестах
Підхід до D	D=3 завдяки конкретному IPC	D=4–6 без явного посилання на наявні IPC-процедури. Основна розбіжність: фахівець знає, що IPC існують; AI тільки припускає

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
	dissolution test та IPC weight	
Тенденція vs GLM Тест 2	—	Найбільше наближення до CQA-охоплення еталону у всій серії GLM 4.6v Flash. 4 CQA-категорії ідентифіковані. Усунення протоколу парадоксально наблизило аналіз до технологічного фокусу

ДОДАТОК Л

Таблиця 3.10 – FMEA Qwen3-30B, Тест 4

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Trituration	Inadequate API-lactose mixing → poor blend uniformity Cause: Insufficient mixing time/speed for large batches	9	CU failure → incorrect dosing. Serious patient safety issue. Critical quality failure.	3	Development validates trituration up to ...kg; largest batch may be borderline	5	CU testing on trituration blend — standard IPC that would detect uniformity issues	135(UNACC EPTABLE)	Validate mixing time/speed for largest batch. Implement CU testing on trituration before Blend I
Blend I	Inconsistent blend uniformity → CU failure Cause: Insufficient mixing time for large batches	9	Poor BU → CU failures across batch → incorrect dosing for patients	3	35-min time validated for smaller batches; largest may require more. Historical data consistent	5	CU testing standard for Blend I; would detect issues before lubrication	135(UNACC EPTABLE)	Validate mixing time for largest batch. Implement CU testing on Blend I before lubrication
Lubricated Blend	Inadequate lubrication → poor flowability and CU Cause: Short lubrication time for large batches	8	Poor lubrication → CU failures and capsule filling issues. Over-lubrication →	3	3-min lubrication time validated for smaller batches; may be insufficient for largest	4	Blend flowability and CU testing performed; would detect issues	96(ALARP)	Validate lubrication time for largest batch. Implement

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
			delayed dissolution → recall						flowability testing on lubricated blend
Capsule Filling (Dosator)	Capsule weight variation beyond limitsCause: Inconsistent dosator filling for extreme strengths	8	Weight variation → CU failures → incorrect dosing. Batch rejection or recall risk.	2	Development shows minimal filling speed impact on weight. 2–3% deviation << 7.5% limit	3	In-process weight checks regularly performed; detects weight variation before completion	48(ACCEPT ED)	Implement continuous in-process weight monitoring. Validate machine parameters for all strengths
Capsule Filling (Dosator)	Capsule bridging in hopper → inconsistent fillingCause: Poor flowability of blend	7	Bridging → inconsistent filling → weight and CU failures. Batch rejection or recall.	2	Development shows no bridging issues. Blend flowability sufficient to prevent bridging.	3	Visual inspection and monitoring of filling rate detects bridging	42(ACCEPT ED)	Continuous flowability monitoring. Validate blend flowability for all strengths
Capsule Filling (Dosator)	Capsule appearance defects (dust, cracks, holes)Cause: Machine	4	Cosmetic defect → patient rejection but no safety or efficacy impact	2	Development shows minimal appearance defects; process well-controlled	2	100% visual inspection required	16(ACCEPT ED)	100% visual inspection for all capsules. Preventive

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
	wear or improper parameter settings								maintenance schedule

Продовження дод. Л

Таблиця 3.11 – FMEA GLM 4.7 Flash, Тест 3

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Weighing & Sieving	Non-conformance of weight uniformity Cause: Incorrect weighing of API or excipients (calibration drift, operator error)	9	Low dose API (7.54%) — small absolute error → significant % deviation → patient dosing risk	2	Standardized weighing and calibration usually prevent this	5	Weight check of raw materials (IPC)	90(AL ARP)	Calibration of balances. SOP adherence. Weight check of raw materials
Trituration (API + Lactose)	Non-conformance of blend uniformity (CU failure) Cause: Incomplete mixing of API with lactose	9	API is active component; uneven distribution → high RSD in CU	3	Simple process but low dose requires strict parameter adherence	5	Blend CU test before lubrication	135(U NACC EPTA BLE)	Monitor Speed and Time. Blend CU testing
Blend I	Segregation (Density/Flowability) Cause: Density differences and multiple transfer steps cause stratification	8	Segregation → weight variation and CU failure	4	High risk due to density differences and multiple transfers	5	Blend CU test before lubrication	160(U NACC EPTA BLE)	Sieving of all materials. Minimize transfer steps. Blend CU testing
Lubrication	Non-conformance of dissolution (Delayed release) Cause:	7	Slows disintegration/dissolution → potential	3	Historical data showed no impact at ...min but risk	5	Dissolution test (QC)	105(U NACC)	Strict Time and Speed limits. Dissolution testing

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
	Over-lubrication — MgSt coats particles → reduced wettability		bioavailability reduction		increases with time			EPTA BLE)	
Capsule Filling	Non-conformance of weight uniformityCause: Inconsistent dosator fill (bridging) or machine vibration	6	Out-of-spec weight; usually recoverable by re-weighing	4	Flowability critical; s/100g must be controlled	5	Weight check of capsules (IPC)	120(U NACC EPTA BLE)	Monitor Blend Flowability. Machine maintenance. Weight check of capsules
Capsule Filling	Non-conformance of content uniformityCause: Segregation in hopper or overfilling → hard powder compaction	9	Directly impacts patient dose	3	Low dose API sensitive to hopper segregation	5	Capsule CU test	135(U NACC EPTA BLE)	Hopper design. Monitor Height of uncompacted mass. CU testing
Capsule Filling	Physical Defect: Cracks/HolesCause: Overfilling → excessive pressure on capsule shell	4	Cosmetic defect; no safety impact unless shell integrity compromised	2	Standardized parameters usually prevent this	5	Visual inspection (IPC)	40(AC CEPT ED)	Monitor Height of uncompacted mass. Visual inspection

Продовження дод. Л

Порівняння Qwen3-30B Тест 4 + GLM 4.7 Тест 3 з еталоном (спільне порівняння)

Критерій порівняння	Еталон (валідаційний фахівець)	АІ (даний тест)
Охоплення CQA	5 CQA-категорій: BU/water content, physical properties, appearance, disintegration, dissolution, weight uniformity, CU	Qwen3-30B Т4 охоплює: BU (Trituration/Blend I), physical properties (flowability/lubrication), dissolution (lubricated blend), weight uniformity, CU, appearance. GLM 4.7 Т3: BU, CU, dissolution, weight uniformity, appearance + CU segregation. Найповніше покриття у дослідженні
Формулювання Failure Mode	'Non-conformance of X': blend uniformity, dissolution, weight uniformity, CU, appearance	GLM 4.7 Т3 використовує 'Non-conformance of X' — ІДЕНТИЧНЕ формулювання з еталоном! Повний збіг структури Failure Mode
Значення E	E=4 стабільно (шкала 1–5). Логіка: 'може завдати шкоди пацієнту'	E=6–9 (шкала 1–10). При перерахунку: E=4/5 еталону ~ E=8–10 за шкалою 1–10. Тест 4 Qwen та GLM 4.7 Т3: E=7–9 для CU/dissolution — якісно відповідають еталону
Значення P	P=4–5: 'процес не валідований + є експериментальні дані'	Qwen3-30B Т4: P=2–3 (STRICT EXCLUSIONS + validated basis). GLM 4.7 Т3: P=2–4. Нижчі за еталон — але методологічно коректніші для 'вже валідованого продукту'
Значення D	D=1–3: конкретні IPC-процедури QS (D=1 для appearance, D=3 для dissolution)	Qwen3-30B Т4: D=2–5; GLM 4.7 Т3: D=5. D=5 (routine IPC sampling) відповідає еталонному D=3 якісно — в обох випадках 'є стандартний IPC'

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
Диференціація за дозуванням	Окремі рядки для 2.5/5/.../25 мг. P=5 для 2.5 мг	Відсутня. Однак Qwen3-30B T4 містить примітку 'extreme strengths' — розуміння без явної диференціації
Загальна схожість еталоном	Еталон — 100% відповідності	Qwen3-30B T4 + GLM 4.7 T3: ~70–75% збігу за CQA-охопленням і ~60% збігу за логікою обґрунтувань. Найвищий показник у дослідженні

ДОДАТОК М

Таблица 3.12 – FMEA GPT-OSS, Тест 1

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
Weighing & Mixing — API weighing	Dust generation during manual weighing of high-potency API Cause: Inadequate containment → airborne contamination and exposure	10	Containment breach unacceptable for OEB 4/5 (Severity Override)	6	Frequent handling & manual weighing increase dust risk	10	No dust monitoring or containment test in protocol	600(UNACC EPTABLE)	Implement closed-system weighers and HEPA-filtered extraction. Update SOP. Assign Engineering/QA
Blending (Trituration & Blend I)	Mixing speed/time not recorded or monitored → poor content uniformity	9	High risk to CQA 'content uniformity' (Initial assessment: High)	4	Variation in manual operation and equipment wear	9	Protocol lacks real-time monitoring of speed/time	324(UNACC EPTABLE)	Add continuous logging of blender speed & time. Include BU sampling after each step. QA review during PPQ
Lubricated Blend	No monitoring of MgSt addition (time, amount) → affects	9	High impact on CU & dissolution (Initial	4	Manual mixing can vary; lubricant distribution critical	9	Protocol does not capture this parameter	324(UNACC EPTABLE)	Record MgSt addition time/amount in SOP. Periodic blend flow

Steps of manufacturing	Problems and causes	E(1-10)	E Justification	P(1-10)	P Justification	D(1-10)	D Justification	RI	Actions
	flow and capsule weight		assessment: High)						tests. Assign Process Engineer
Capsule Filling	Manual capsule weight check, no electronic data capture → ALCOA+ violation	6	Medium severity (Initial assessment: Medium for capsule weight)	6	Human factor error rate ~3–5% in manual weighing	8	Lack of audit trail; detection only during QC review	288(UNACC EPTABLE)	Implement electronic weight recording with audit trail. Train operators on ALCOA+. Data integrity checks in SOP
Packaging & Blister Integrity	Visual blister check only; no seal/filter integrity test → dust escape HPAPI	10	Containment breach unacceptable (Severity Override)	7	Shared equipment, frequent micro-stops increase failure probability	10	No quantitative seal test in protocol	700(UNACC EPTABLE)	Validate blister seals with pressure-test and leak detection. Incorporate into PPQ. Routine seal tests by QC

Продовження дод. М

Порівняння GPT-OSS з еталоном

Критерій порівняння	Еталон (валідаційний фахівець)	AI (даний тест)
CQA-охоплення vs еталон	BU, dissolution, weight uniformity, CU — всі як significant/high risk	Blending охоплює CU та dissolution частково. Capsule Filling — weight (через ALCOA-ризик). Але 2 з 5 рядків — safety/containment (Severity Override), що відсутні в еталоні
Систематична проблема D	D=1–3 завдяки наявним IPC-процедурам	D=8–10 систематично. D=10 для blending — нереалістично (стандартний IPC існує). D=10 для packaging — обґрунтовано. Це є ключовою розбіжністю з еталоном
Значення RPN	RI=20–100 (шкала 1–5)	RPN=288–700. Найвищі значення у дослідженні. RPN=700 для blister integrity — ефект $D=10 \times E=10 \times P=7$
Порівняно з GLM 4.7 Тест 3	~70–75% наближення до CQA-охоплення еталону	~40–50% наближення. GPT-OSS відволікається на containment/safety без повернення до технологічних CQA
Тенденція	—	GPT-OSS на тому ж промпті що і GLM 4.6v Flash T3 дає менш відповідний результат через D-інфляцію та фокус на HPAI safety. Підтверджує: модель > промпт для цього рівня контексту

ДОДАТОК Н

Таблиця 3.13 – Узагальнена тенденція наближення до еталону

Тест	Модель / Промпт-версія	CQA охоплення (з 5)	Тенденція	Ключова причина зміни
Тест 1(Qwen3-14B)	Base + Reverse Engineering	1–2 з 5	↓ Дуже далеко	Відсутня технологія; лише критерії протоколу
Тест 2(Qwen3-14B)	+ Auditor Mode STEP 1–3	2 з 5	↑ Незначне покращення	З'явилась стадія blend; але фокус на CPP, не CQA
Тест 3(Qwen3-14B)	+ Triangulation STEP 1–4	2 з 5	→ Без зміни CQA	Dev.Report додано, але ризики — Instrument Failure, не CQA
Тест 4–5(Qwen3-14B)	+ Severity Override+	2–3 з 5	↓ Відхилення(НРА PI focus)	Severity Override відвів аналіз до safety systems

Тест	Модель / Промпт-версія	CQA охоплення (з 5)	Тенденція	Ключова причина зміни
	Constraints A/B/C			
GLM 4.6v T1(model swap)	Той самий промпт	3 з 5	↑ Перший прорив	GLM краще інтегрує Dev.Report: humidity→CU
GLM 4.6v T2	+ R&D Risk Alignment+ Criticality Override	3–4 з 5	↑ Покращення	R&D Risk Alignment знижує E-інфляцію; E ближче до еталону
GLM 4.6v T3	Без протоколу+ нормативні документи	4 з 5	↑↑ Суттєве покращення	Відсутність протоколу → технологічний фокус

Тест	Модель / Промпт-версія	CQA охоплення (з 5)	Тенденція	Ключова причина зміни
Qwen3-30B T2-3(STRICT +physical)	STRICT EXCLUSION S+ physical focus	3-4 з 5	↑ Покращення P	Реалістичні P, але E занижена без ABS_E rule
Qwen3-30B T4★ Найкращий	+Effect is Absolute Rule	5 з 5	★★ Найближчий	ABS_E + STRICT = реалістичні P + коректні E
GLM 4.7 T2	STRICT + ABS_E rule	4-5 з 5	★ Наближення	Ідентичний промпт → більша гранулярність
GLM 4.7 T3★ Найкращий	+CQA-прикладу Instructions	5 з 5	★★ Ідентичне формулювання	'Non-conformance of X' = формат еталону
GPT-OSS T1(model swap)	GLM 4.6v T3 prompt	3 з 5	D-інфляція	D=10 систематично; safety-bias відвів від CQA

Примітка: ★★ позначає тести з найвищим рівнем наближення до еталону (70–75% за CQA-охопленням та 60% за логікою обґрунтувань). Тенденція ↑↑ відображає найбільший одиничний приріст між послідовними тест