Preliminary Exploration of AI Applications in Quality Management for Clinical Trials

Yuto Kyosaka
FY2024 Task Force 1-2, Data Science Expert Committee, Drug Evaluation Committee,Japan Pharmaceutical Manufacturers Association (JPMA)

Co-creation

Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter representing the JPMA task force and should not be attributed to PHUSE or JPMA, its directors, officers, employees, volunteers, members, chapters, councils, Communities or affiliates.

This presentation is incomplete without accompanying verbal commentary.

Agenda

- Background
- Efforts in the Task Force
- Used Clinical Trial Protocol and Data
  - ➢Overview of the Study Protocol
- Prototype 1: AI to support identification of critical to quality (CTQ) factors using large language models
- Prototype 2: AI to identify critical issues in the data based on a clinical trial protocol using large language models
- Summary

Background

Progress in the Practical Application of AI

- The advancement of conversational AI based on large language models has been remarkable, and practical implementation is progressing across a wide range of fields, including knowledge retrieval, document summarization, and even business process automation.

ICH E8(R1) and E6(R3)

- ICH E8(R1) and E6(R3) emphasize the introduction of “Quality by Design” and a “Risk-proportionate Approach.”
- As a result, there is an increased focus on ensuring quality and managing risk from the planning stage of clinical trials.

Efforts in the Task Force

JPMA Co-creation

- In the Task Force “Transforming Data Management through AI” of the Data Science Expert Committee, Japan Pharmaceutical Manufacturers Association (JPMA),
  prototyping was conducted in FY2024 to explore the potential use of large language models (LLMs) for clinical trial quality management from the perspective of Quality by Design, covering all stages from clinical trial planning through to clinical trial completion.

- Specifically, the following two prototypes were developed.
  1. AI to support identification of critical to quality (CTQ) factors using large language models
  2. AI to identify critical issues in the data based on a clinical trial protocol using large language models

Used Clinical Trial Protocol and Data

The clinical trial protocol and data published in the CDISC sdtm-adam-pilot-project were used.

- Note: The terms of use for the sdtm-adam-pilot-project data explicitly state that the data must be used without modification.
- In this task force, we communicated our intention to CDISC, the rights holder, and obtained special permission to modify and use the data.
- From the perspective of protecting confidential information, using actual clinical trial protocols or data used by the companies to which the members belong into AI was not considered an option.

Overview of the Study Protocol

Title: Safety and Efficacy of the Xanomeline Transdermal Therapeutic System(TTS) in Patients with Mild to Moderate Alzheimer's Disease

Design: Randomized, double-blind, parallel-group, placebo-controlled (three arms)

Primary Endpoints: ADAS-Cog, CIBIC+, Safety

Active Ingredient: Xanomeline

Dosage Form: Transdermal patch formulation

Application Procedure:

1. Every morning, apply hydrocortisone ointment to the application site, and wipe off the cream after 30 minutes.
2. Apply the transdermal patch and remove it after 12 to 14 hours.

   Note: Change the application site daily (rotate sites over a week).
3. After removing the patch, reapply hydrocortisone ointment locally to the application site.

Application Period: 26 weeks

Prototype 1: AI to support identification of critical to quality (CTQ) factors using large language models

Purpose

- Evaluate the validity of the draft CTQ (Critical to Quality) factors and risk mitigation strategies generated by AI.
- Examine the effectiveness of AI in generating CTQ factor drafts, identify more effective approaches for its utilization, and discuss key considerations for its application.

Method Overview

1. Build an conversational AI to support the identification of CTQ factors.
2. Create and refine CTQ factors through dialogue with the AI.
3. Review the draft CTQ factors.

Method 1: Building a Conversational AI

A conversational AI named “CtQF Facilitator” was developed by configuring a system prompt and context for the large language model “Gemini 2.0 Flash Exp.”
(The open-source platform “Dify” was used.)

No coding is required to create chatbots and Retrieval-Augmented Generation (RAG) systems, as well as workflows that integrate AI.

- System Prompt

あなたはファシリテーターとしてユーザの提供するプロトコルのCritical to Quality Factorを特定していきます。
Critical to Quality Factorを特定するための情報は、プロトコルとその提供者から引き出す必要があります。
そのため、まずは提供されたプロトコルを精査し、その中でCritical to Quality Factorを特定するために不足している情報をユーザーに質問してください。
質問はすべてのCritical to Quality Factorが特定されるまで続けてください。
なお、ユーザーはこれまでにいくつもの薬剤を世界に送り出している製薬メーカーであり、評価項目の設定に疑義は無く、試験を実施する上での課題となるCritical to Quality Factorを考慮したリスクを特定するようにしてください。
与えられたプロトコル以外でも起こりうる要因はCritical to Quality Factorとして考慮する必要がありません。
以下の知識を参考にして、ユーザーをファシリテートしてください。

知識
[コンテキスト]

Input information related to CTQ Factor

Configured to request the necessary information from the user and create CTQ Factor Proposals through an interactive process.

- Reference materials provided as context:
  - Exploring the Critical to Quality (CTQ) Factors
  - CTTI QUALITY BY DESIGN PROJECT - CRITICAL TO QUALITY (CTQ) FACTORS PRINCIPLES DOCUMENT
  - Quality Tolerance Limits’ Place in the Quality Management System and Link to the Statistical Trial Design: Case Studies and Recommendations from Early Adopters

Method 2, 3: Drafting CTQ factor proposals through dialogue with the AI → Review

JPMA Co-creation

CtQF Facilitator

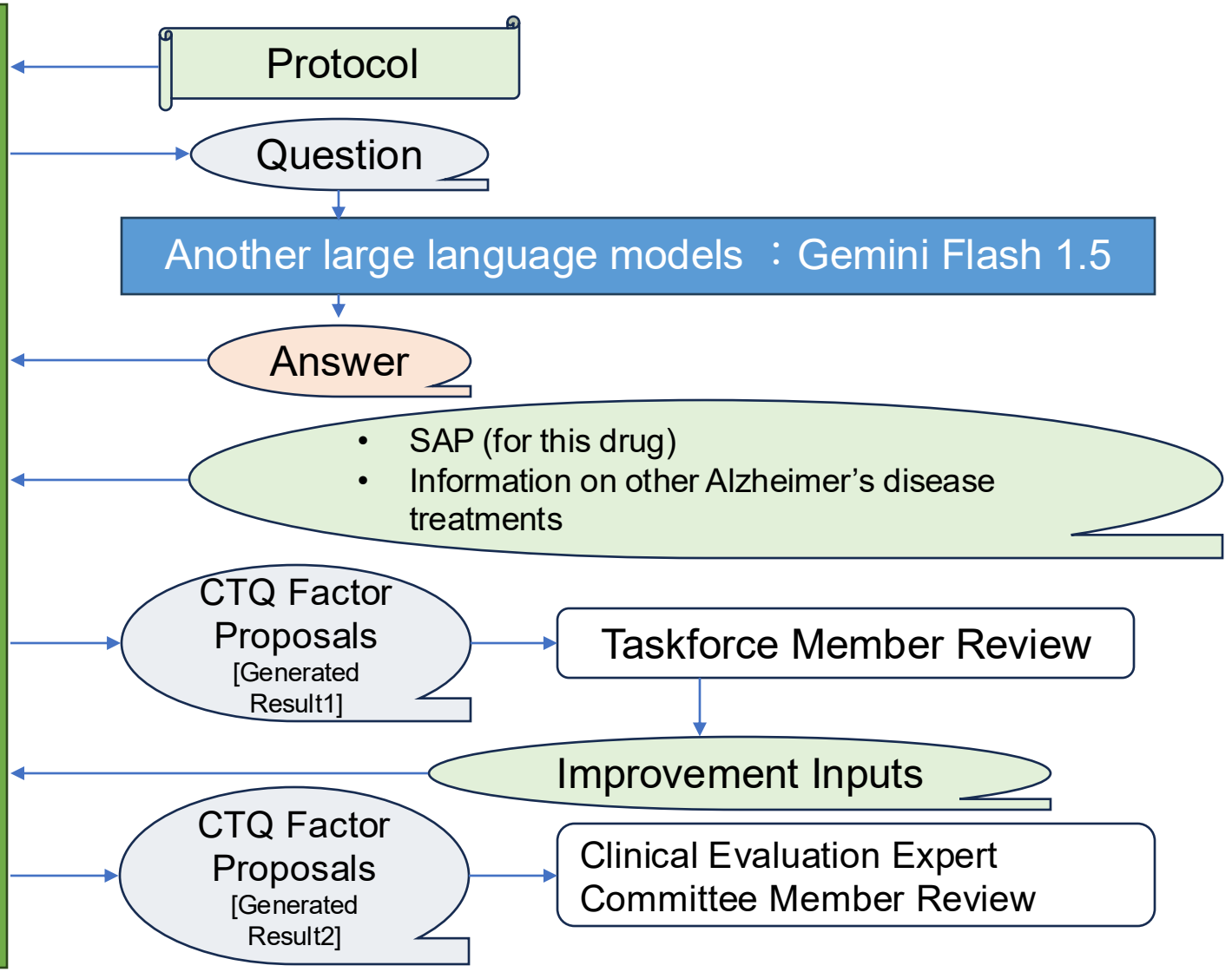
Large language model
Gemini 2.0 Flash Exp

SystemPrompt

~To identify the CTQ factors, you need to elicit information from both the protocol and its provider. First, carefully review the provided protocol and ask the user ~

Context

1)「Exploring the CTQ (CTQ) Factors - CTTI」
2)「CTTIQUALITYBYDESIGNPROJECT-¬-CTQ(CTQ)FACTORSPRINCIPLESDOCUMENT」
3)「Quality Tolerance Limits' Place in the Quality Management System and Link to the Statistical Trial Design: Case Studies and Recommendations from Early Adopters」



Questions by CtQF Facilitator

Question 1: Regarding the application procedure for the TTS formulation, are there any specific steps that you consider particularly prone to errors or difficult for patients or caregivers to perform?

Question 2: For the implementation of the assessment items (ADAS-Cog, CIBIC+, DAD, NPI-X), what training or procedures are planned to minimize variability among assessors? In addition, if multiple assessors are involved, how will consistency between assessors be ensured?

Question 3: In cardiovascular safety assessments (ECG, 24-hour Holter monitoring), what points require special attention during data collection and analysis? For example, please describe considerations such as the duration of Holter monitoring or handling artifacts during analysis.

Question 4: Regarding inclusion and exclusion criteria for patient enrollment, are there any items that are particularly difficult to judge at screening or that could lead to erroneous enrollment?

Question 5: Among the concomitant medications that are prohibited, are there any drugs that patients are likely to mistakenly use? If so, how will this be addressed?

CTQ Factor Proposals [Generated Result 1]

Six CTQ Factor Proposals Were Developed for the Following Items.

1. Ensuring compliance with the application procedures for TTS formulations
    - Standardization of TTS application procedures
    - Confirmation of removal of applied formulations

2. Standardization of Evaluation Procedures and Ensuring Evaluator Consistency
    - Standardization of procedures for evaluation items
    - Ensuring consistency among evaluators.
    - Management of evaluator changes

3. Accuracy in Cardiovascular Safety Assessment
    - Standardization of procedures for cardiovascular safety assessment
    - Ensuring accuracy in data analysis

CTQ Factor Proposals [Generated Result 1]

JPMA Co-creation

Six CTQ Factor Proposals Were Developed for the Following Items.

4. Strict Application of Patient Inclusion and Exclusion Criteria
    - Standardization of screening procedures
    - Thorough verification of patient information
    - Confirmation of Alzheimer's disease diagnosis

5. Prevention of Prohibited Concomitant Medication Use
    - Thorough dissemination of the list of prohibited concomitant medications
    - Verification of concomitant medications
    - Confirmation of concomitant therapies and opioid use

6. Other
    - Compliance with discontinuation criteria
    - Compliance with investigational product administration procedures
    - Early detection and management of dehydration
    - Monitoring of body weight changes
    - Minimization of missing data
    - Verification of data normality

CTQ Factor Proposals [Generated Result 1]
Compliance with the Application Procedure for TTS formulations

JPMA Co-creation

Specific issues	CTQ Factor Proposals	Risk	Risk Mitigation Measures Proposals
・Incorrect selection of application site (e.g., repeated application to the same location, application to areas with excessive hair or to wounded areas) ・Incorrect removal of the protective film (e.g., touching the medication) ・Air bubbles trapped under the patch ・Insufficient pressure applied after application ・Incorrect frequency or duration of application ・Difficulty in applying the patch for elderly patients, children, or patients with limited joint mobility ・Forgetting to remove previously applied patches	・Standardization of TTS application procedures ・Confirmation of removal of applied formulations	・Variability in pharmacokinetics and decreased efficacy ・Skin problems (irritation, rash, detachment) ・Patient discomfort and loss of trust in treatment ・Serious adverse effects due to overdose ・Discrepancy between ITT analysis and PP analysis results ・Decreased reliability of As Treated analysis	・Prepare a detailed manual describing the rotation of application sites, application method, application duration, removal procedure, and responses to detachment, and distribute it to patients and caregivers. ・Conduct hands-on training sessions on application procedures and confirm the level of understanding. Regularly check the application status and promptly address any issues. ・Provide support tools such as application calendars and mirrors. ・Ensure thorough explanations are given to caregivers. ・Strictly implement procedures to confirm the removal of previously applied products at the time of replacement (e.g., using checklists or photographs). ・Ensure thorough documentation of application status to improve the accuracy of As Treated analyses. ・Predefine the analytical approach for cases where the dosage is changed.

CTQ Factor Proposals [Generated Result 1]

Standardization of Evaluation Procedures and Ensuring Evaluator Consistency

JPMA Co-creation

Specific issues	CTQ Factor Proposals	Risk	Risk Mitigation Measures Proposals
・ Variability in evaluations among assessors ・ Lack of standardization in the evaluation environment ・ Lack of standardization in the evaluation order ・ Lack of standardization in evaluation tools ・ Change of psychological assessor	・ Standardization of procedures for conducting evaluation items ・ Ensuring consistency among evaluators ・ Management of evaluator changes	・ Decreased reliability of evaluation results ・ Difficulty in interpreting study results ・ Bias in statistical analysis ・ Difficulty in comparing evaluation results ・ Decreased reliability of analysis results for primary and secondary endpoints ・ Impact on the evaluation of interactions	・ Prepare an evaluation manual that describes in detail the definitions of evaluation items, procedures, and scoring criteria. ・ Conduct training using mock evaluations and video materials to improve the skills of evaluators. ・ Hold regular meetings among evaluators to share evaluation criteria. ・ Perform calculation checks to detect calculation errors and scoring mistakes. ・ Quantitatively assess inter-rater agreement using statistical methods.Introduce blinding methods wherever possible. ・ Clarify the selection criteria for evaluators and employ those with relevant knowledge and experience. ・ Standardize procedures for changing evaluators and record the reasons for changes. ・ Provide sufficient training to evaluators after any change. ・ Minimize changes of evaluators that could affect evaluation data. ・ Minimize evaluator bias that could affect the assessment of interactions.

Improvement Points Identified by Task Force Members (Excerpts)

General Comments

- The “As Treated analysis” was not conducted in this study. Likewise, “decreased reliability of As Treated analysis” does not apply to this study.
- Although “introducing blinding wherever possible” is listed as a risk mitigation measure, this study has already adopted blinding.

1. Compliance with TTS Patch Application Procedure

- The statement “Difficulties are likely to occur in elderly patients, children, or those with restricted joint mobility” is not applicable because children are not included in the study population.
- Applying the patch to “areas with dense hair or injured skin” is unlikely to be a common source of error.
- “Patient discomfort or distrust in treatment” poses minimal risk since participants voluntarily enrolled in the study.

5. Prevention of Prohibited Concomitant Medications

- The described content is general and not specific to this trial as a CTQ factor.
- Adding “Prevention of prohibited concomitant medications” solely based on the reason that “similar drugs were excluded from the Per Protocol Set” for opioid products lacks sufficient justification.

CTQ Factor Proposals [Generated Result 2]

1. Instructions to the CtQF Facilitator

- For [Generated Result 1], the improvements identified by the task force members were communicated to the CtQF Facilitator. Additionally, for the following reasons, a request was made to the CtQF Facilitator to change the structure of the output format for the CTQ Factor draft:
    - What was described as a “risk” actually referred to the consequence if that risk materialized, whereas what was written as an “issue” corresponds to the risk associated with the CTQ factor.

→ Therefore, an update to a three-column structure was requested:
〔CTQ Factor Proposals〕 〔Risk〕 〔Risk Mitigation Measure Proposals〕

2. Generation of CTQ Factor Draft [Generated Result 2]

- Items that were pointed out as errors or unnecessary were deleted in the relevant sections.
- The structure of the CTQ Factor draft was also changed as instructed.
- The “Specific Issues” column was deleted, and the content that was originally in the “Specific Issues” column was reflected in the “Risk” column.
- No changes were observed in the other output content.

Discussion_1

- A draft CTQ factor was generated that appropriately captured the concept of CTQ factors, accurately recognized the provided information, and produced comprehensive and specific results from multiple perspectives.
- This draft is considered useful as a foundation for identifying CTQ factors and for further discussion of corresponding actions.

Discussion_2

Usefulness of System Prompts Configured for Conversational AI

- The system prompt was configured to allow the AI to ask follow-up questions when information is missing, in order to draft more precise CTQ factors.
- This approach helped reduce the risk of hallucinations and improved the comprehensiveness and specificity of CTQ factors by ensuring that outputs were generated only after sufficient information had been gathered.
- Furthermore, the content of the AI's questions enabled the user to recognize what information was missing from the initially provided documents.

Granularity of Information and Instructional Approaches

- The additional risk mitigation measures proposed after reading the SAP were highly abstract in nature.
- It may also be useful to provide the AI with additional instructions regarding the desired granularity of information in the response, for example, by presenting sample cases of the expected answers.

Improvements Identified from This Prototype

- By having the AI also output the rationale for the proposed CTQ factors—such as the study objectives and the impact on patient protection—it may have been possible to gain a deeper understanding of the AI's reasoning process.

Discussion_3

Importance of Human-in-the-Loop

a. In the [Generated Result 1], there were contents that did not align with the information provided about the trial, as well as unnecessary descriptions.
   This was likely caused by user inputs containing general information from another AI. To enable more precise generation, **the content-specific information should be selected for input.**

b. The AI suggested “thorough confirmation procedures for drug removal (checklists, photography, etc.).” as a risk mitigation measure. While photographing the relevant site each time a drug is removed may be useful for confirming adherence to the study plan, it could impose a significant burden on participants or caregivers.
   In clinical trials, operational characteristics including feasibility should also be considered. To comprehensively assess complex factors such as the magnitude of risk, the effectiveness of mitigation measures, and the impact on trial operations, **it is essential that a team of experts deliberate and determine the optimal solution.**

Discussion_4

Consideration of AI Utilization Methods in Actual Operations

- It is ideal to utilize AI at the trial start-up stage and when making significant protocol amendments.
  - During the start-up stage, where frequent updates to specifications and plans occur, AI is particularly effective in promptly generating draft CTQ factors that reflect those updates.
  - These draft CTQ factors may then serve as a basis for updating plans and specifications to improve the quality and safety of the trial.
- Utilization Methods
  - Add information stepwise (from trial plan outline → SAP draft → similar trial information)
  - By including specialized documents from each department involved in drug development, the comprehensiveness and specificity of the output can be enhanced.

Prototype 2: AI to identify critical issues in the data based on a clinical trial protocol using large language models

Purpose

- Examine the usefulness and feasibility of efficient and rapid data review using large language models, applying the principles of risk-based approaches

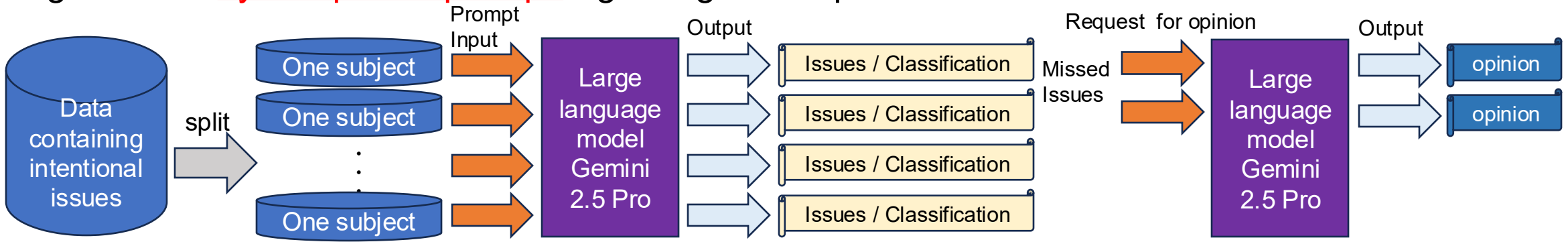
Method

1. Data Preparation
   - Based on clinical trial data (CDISC sdtm-adam-pilot-project), 38 intentional errors (e.g., date reversals, inappropriate dosing, protocol deviations, etc.) were set in the data for 30 subjects.
2. Detection of Issues Using Large Language Models
   1. The data for each subject, split into individual JSON files, were input into a large language model to detect issues.
   2. The detected issues were classified as “Critical,” “Major,” or “Minor”

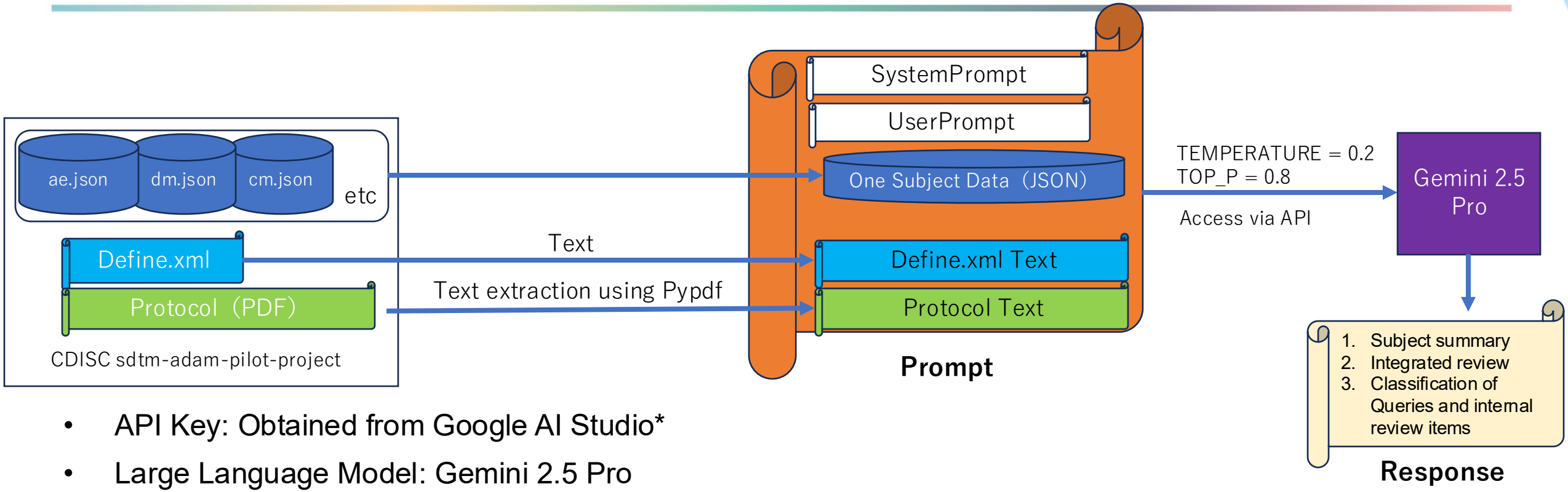
- Critical: Issues with a significant impact on safety or primary endpoints
- Major：Issues affecting secondary endpoints or important safety evaluation items
- Minor：Minor inconsistencies or data entry errors, etc.

Task Force named its detection prompt “Smart Clinical Data Validator”

3. For cases where Issues were not detected as expected, an opinion was generated by a separate prompt regarding the output results.



Overview of the Program



- API Key: Obtained from Google AI Studio*
- Large Language Model: Gemini 2.5 Pro
- Programming Language: Python

*Note: Since this prototype used public data, it was performed in a public environment. When handling data containing confidential information, the process should be conducted in a secure environment where the data is not used for AI training and security is ensured.

- The program is available at the following location:

JPMA2024TF1-2/2. Smart Clinical Data Validator/2.1. Code/code.ipynb at main · Takumi173/JPMA2024TF1-2 · GitHub

Prompt Structure

JPMA Co-creation

1. **System Prompt**
   - Overview: define the role and core principles (Presenting the rationale for judgments, Prioritizing safety and protection of rights, Considering the impact of inconsistencies and protocol deviations.). Also, specify prerequisite knowledge (utilization of SDTM, Define.xml, protocol, and medical knowledge) and error handling.
   - Purpose: To provide the basic role and prerequisite knowledge for the AI, serving as a guide for utilizing user-provided data.

2. **User Prompt**
   - Overview: Detailed instructions and output templates for three tasks (subject summary creation, Integrated review, Classification and criticality assessment of queries/internal review items.)
   - Purpose: To provide concrete review instructions.

3. **Subject Data (in JSON format)**

4. **Define.xml**
   - Purpose: To input information on the meaning, label, and coding dictionaries.

5. **Study Protocol (text extracted from PDF)**
   - Purpose: To input study information such as SAE definitions, assessment schedules, and prohibited medications. Enables protocol compliance checks and citation of rationale.

Transition of Prompts

JPMA Co-creation

Initial Prompt

添付のJSONファイルは臨床試験の症例データです。このデータには誤りが含まれている可能性があります。
臨床試験の専門医として、以下の観点からJSONデータを確認し、回答してください。
時系列を考慮に入れて回答することが非常に重要です。また、症例本人の主訴も重要な情報です。
有害事象は、被験薬との因果関係を問わないあらゆる好ましくない事象や兆候を指します。
JSONデータをレビューする観点
1. 臨床の経過から見て疑わしい情報
臨床の経過から見て疑わしい情報をすべて列挙し、その理由を説明します。
2. 有害事象として報告されている情報の整合性
報告された有害事象が、関連するその他の情報と整合が取れない場合、その理由を説明します。
3. 有害事象以外のデータにある安全性上の懸念
安全性上の懸念が認められる情報をすべて列挙し、それが有害事象として記載されているかどうかを評価します。
4. その他の不整合
上記以外に不整合が疑われるデータをすべて列挙し、その理由を説明します。
回答フォーマット
*- **疑義事項**: 理由*
*- **疑義事項**: 理由*

1. When the task force members created 10 cases of test data with intentional inconsistencies from scratch, a certain number of issues were successfully detected.

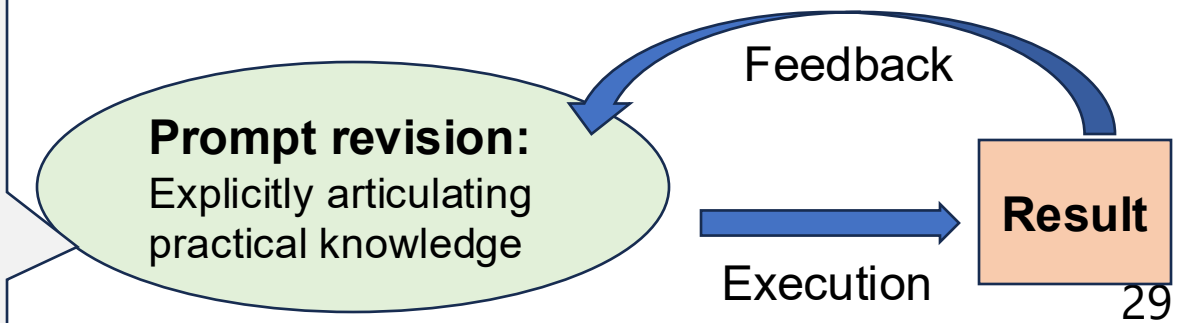
2. However, when applied to publicly available clinical trial data with a large volume and variety of data, **the desired results were not obtained.**

・This was considered to be due to the broader range of data interpretation and the ambiguity of the prompts.
Examples:
- The structure and definitions of the JSON data were not clearly specified.
- There was no explicit description of “information suspicious from the clinical course” or “safety concerns.”

By repeatedly revising and executing the prompt, and supplementing missing information, the prompt on the previous page was created.

Taking the following points into consideration, the prompt was refined through dialogue with AI.

•What is the ultimate objective of this prompt? What is the essential purpose to be achieved?
•What are the actions or outputs that should be avoided?
•What are the key points to consider in making judgments and generating outputs?



Contents of the System Prompt - 1

- **Define the Role**
  - You are an AI assistant supporting the integrated review of clinical trial data.
  - Your primary objectives are to protect the rights and safety of participants, and to ensure the scientific validity of the trial and the reliability of the data.
  - To achieve this, you utilize the provided information (such as JSON data, Define.xml, and the protocol) to perform medical assessments, verify data integrity, and check protocol compliance.
- **Core Principles**
  - Your assessments must be based on the provided information and general medical knowledge; personal speculation or hallucination is strictly prohibited.
  - The highest priority is given to participant safety and the protection of their rights.
  - You must consider the impact of data inconsistencies or deviations on your evaluation.
  - When raising issues, you must state the basis for your findings (data, protocol, or medical knowledge).

Contents of the System Prompt - 2

JPMA Co-creation

- **Prerequisite Knowledge**
  - SDTM: Clinical trial data standards (e.g., DM, AE, VS, LB, etc.)）
  - Define.xml: Metadata defining data structures and labels
  - Protocol: Study plan (including eligibility criteria, dosing plan, assessment schedule, etc.)
  - Utilization of Medical knowledge: Clinical significance of diseases, drug actions, adverse effects, laboratory values, etc.
- **Task Execution Notes**
  - Follow the specified output format.
  - If a judgment cannot be made due to insufficient or conflicting information, clearly indicate this.
- **Error Handling**
  - If any required file is missing or in an invalid format, report the error and terminate processing.

- The actual prompt is written in Japanese and includes more detailed descriptions.
- In addition, to provide structured data to the AI, prompts are written using Markdown notation (for example) 「*」 = bullet points 「**」 = emphasis (bold)

Contents of the User Prompt - 1

JPMA Co-creation

Define the role

- As an integrated reviewer of clinical trial data, you are responsible for consolidating the perspectives of the Medical Monitor, Data Manager, and CRA, and reviewing the provided information (protocol, JSON data, Define.xml).

Subject Summary Creation

- Refer to JSON and Define.xml to describe patient background (using Define.xml labels) and major events in chronological order.
- Extract abnormal fluctuations or clinically significant changes in adverse events, laboratory values, vital signs, and efficacy assessments using medical knowledge.
- Record dates based on variables such as Study Day.

Integrated Review

- detect issues from three perspectives and assign a level of criticality (Critical/Major/Minor):
- **Medical Review (Medical Monitor perspective):** Validity of safety and efficacy assessments, unreported risks
- **Data Integrity (Data Manager perspective):** Cross-domain/intra-domain inconsistencies, outliers, missing data, data deficiencies
- **Protocol Compliance (CRA perspective):** Inclusion/exclusion criteria, dosing plan, assessment schedule, consent procedures

Contents of the User Prompt - 2

JPMA Co-creation

Classification of Queries and Internal review items

- Queries to study sites (for ensuring safety and reliability) or Internal review items (e.g., minor inconsistencies)
- Assign a level of Criticality (Critical/Major/Minor) and organize them sequentially

Output Format

- Describe the subject summary, integrated review results (M-, D-, P-), and queries/internal review items (Q-, I-) in Markdown format according to the template.

Each template

- The actual prompt is written in Japanese and includes more detailed descriptions.
- In addition, to provide structured data to the AI, prompts are written using Markdown notation
- (for example):
  「*」 = bullet points 「**」 = emphasis (bold)

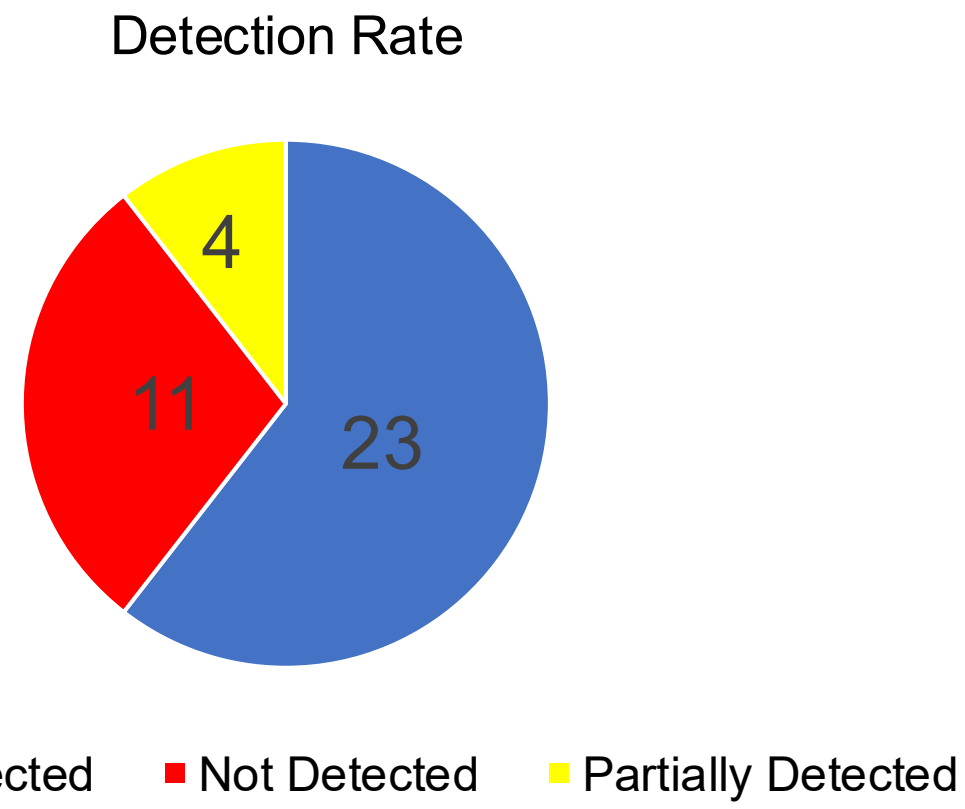
About the User Prompt

Prompt Review Role Settings

- We expected that by incorporating multiple perspectives into AI would help address points that might be overlooked when relying on a single perspective.
- Initially, we prepared three roles as a completely independent AI, however, this resulted in overlapping queries across the three roles.
- Therefore, we included all three roles in a single prompt and instructed the AI to review from each perspective.
  As a result, this approach worked as expected, enabling us to consolidate three review roles into a single prompt. The output closely matched that of separate prompts for each role, while largely eliminating duplicate queries.
  → This made it possible to reproduce, with a single prompt, the data review by three people as conducted in actual clinical trials.

Result

- We evaluated the detection accuracy for 38 issues set within 30 subjects.
  - Of these, 23 issues (60.5%) were accurately detected, and 4 issues (10.5%) were partially detected. The remaining 11 issues (28.9%) were not detected.



AI's Opinion of Undetected Cases: Case Studies

- **Cases Where Seriousness Was Judged as Non-SAE but SAE Is Suspected**
  - The following five subjects were expected to be detected as serious cases; however, Smart Clinical Data Validator achieved only partial detection (△) or no detection (×).

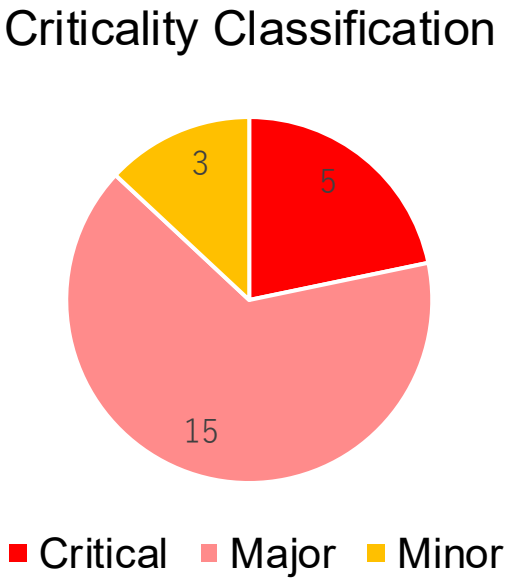
USUBJID	Summary of Issues for Clarification	Detection
01-701-1034	The disease name : MALIGNANT HYPERTENSION.	△
01-701-1047	Vital signs: Diastolic Blood Pressure = 121, 124; Systolic Blood Pressure = 185, 183.	△
01-701-1180	The disease name : SUDDEN DEATH.	△
01-704-1017	The disease name : Late effects of cerebral infarction	×
01-704-1017	The disease name : Brain death	△

AI's Opinion

- For USUBJID 01-701-1034, 01-701-1047, and 01-704-1017, AI interpreted as correct and no query was required, as the data indicated that the site reported the subject did not meet any SAE definition
- For USUBJID 01-701-1180 (sudden death) and 01-704-1017 (brain death), inconsistencies between the event and related data were detected during the initial review, and urgent confirmation was requested. However, the response differed after additional clarification was provided.

Results of Criticality Classification

- Out of 23 issues accurately detected, 5 were classified as "Critical," 15 as "Major," and 3 as "Minor".



- Issues directly related to participant safety, such as adverse event data and investigational product dosage, tended to be classified as "Critical".
- Cases suspected to be simple data entry errors, such as the reversal of Start Date and End Date or the inversion of Systolic and Diastolic Blood Pressure values, tended to be classified as "Minor".

Results of Criticality Classification Case Studies

JPMA Co-creation

- Abnormal investigational product dosages
  - "216 mg" was classified as "Critical," while "82 mg" was classified as "Minor".
  - It is presumed that "Critical" was determined by AI because it represents a significant deviation from the study protocol and poses a safety risk.
- Inclusion/exclusion criteria
  - only 1 out of 9 cases was classified as "Critical," while the others were "Major".
  - Although definitions were provided in the prompt, there were no clear criteria, so it is presumed that the AI made a comprehensive judgment.
  - The results are likely to vary with each execution, Suggesting the need for multiple runs, a review of the classification criteria, or prompt refinement.

Discussion_1

Importance of Human Input

1. There have been cases where the AI was strongly influenced by human input (definitions) and failed to detect inconsistencies. (Example: Slide 36)

   → It is important to understand the influence of human input on the response from AI, and to strive to provide information that is neither excessive nor insufficient. Depending on the situation, withholding information and seeking the AI's viewpoint may help validate its reasoning.

2. There were instances where the AI’s conclusions differed depending on the execution. (Example: Slide 36)

   → It is necessary to control the AI by adjusting its parameters or by crafting more specific and clear prompts. In addition, it is important to consider a process in which the same prompt is executed multiple times, and the results are compared and integrated to obtain more reliable outputs.

Discussion_2

Importance of Human-in-the-Loop

- While the results seem to be comparable to a human review in some aspects, they are not perfect and cannot fully replace it.
- It is important to ensure the accuracy and quality of results, and to maintain transparency and accountability by having humans ultimately review the outputs generated by AI.

Discussion_3

AI-Based Data Review and Future Perspectives

- Although some challenges such as missed detections and over-detections remain, the AI was able to detect issues with a certain level of accuracy or higher.
- The result showing that AI demonstrated the ability to quickly detect potential issues based on user provided documents suggests an effective alternative to the traditional workflow that has relied on manual review.
- However, AI does not always meet the expected quality. Therefore, a Human-in-the-Loop process remains essential.
- In particular, when handling context-dependent information or text that allows multiple interpretations, there is a risk that AI may not generate reliable outputs.
  A practical solution in such case is for AI to generate explanations alongside its outputs, enabling human reviewers to interpret and apply them effectively.

Summary

Summary

- Both of the two prototyping initiatives yielded results that strongly suggest the potential for practical AI implementation.
- When drafting prompts, it was confirmed that clearly defining the purpose of the question and the expected scope of answers in advance, and minimizing ambiguity, enabled the AI outputs to remain focused, consistent, and highly accurate.
- By establishing a “collaborative workflow” in which AI rapidly and comprehensively drafts candidate issues based on a wide range of information, and humans then curate the outputs and validate their reliability from a professional perspective, both operational efficiency and appropriateness can be achieved.
- Building such a “collaborative workflow” is expected to enable early detection of risks and the prompt and accurate implementation of countermeasures, from the planning stage through to the execution of a clinical trial.

Task Force Members

Role	Organization	Name
Supervisor	Pfizer R&D Japan G.K.	Osamu Komiyama
Lead & Member	Pfizer R&D Japan G.K.	Hidetoshi Misawa
Lead & Member	Chugai Pharmaceutical Co., Ltd.	Naoto Awaji
	EA Pharma Co., Ltd.	Yukari Kitta
	Novartis Pharma K.K.	Yusuke Inoue
	Pfizer R&D Japan G.K.	Takumi Tanaka
	Astellas Pharma Inc.	Natsumi Shibata
	Otsuka Pharmaceutical Co., Ltd.	Yuto Kyosaka

Questions？