U.S. FOOD & DRUG ADMINISTRATION

Generative AI-Powered Regulatory Efficiency and Excellence

Jae (Mike) Lee, Ph.D.

Division of Quantitative Methods and Modeling, Office of Research and Standards, Office of Generic Drugs

CDER | U.S. FDA

June 11, 2025

Disclaimer

This presentation reflects the views of the authors and should not be construed to represent the FDA's views or policies.

Outline

- Background
- Method
- Pilot Study
  - Retrieval Augmented Generation (RAG)
  - Prompt Engineering
- Agentic Workflow
- Takeaway

Enhancing Regulatory Excellence Through Generative AI

- Generative AI, particularly large language models (LLMs), introduces new ways to enhance how we access, interpret, and apply knowledge.

- These technologies do not replace expertise– they complement it, helping us work more efficiently, consistently, and insightfully.

- Integrating LLMs into regulatory workflows to support excellence at scale.

Potential Use Cases in Regulatory Workflows

- Automating the creation of reports
- Summarization of internal documents
- Data-driven insights to support regulatory decision-making processes
- Extracting relevant information from FDA documents
  - Leveraging LLM to extract maximum daily dose

Leveraging LLM to Extract Maximum Daily Dose

- Maximum daily dose (MDD) is important in excipient evaluation and in establishing FDA-approved maximum daily exposure (MDE) to excipients posted on FDA's Inactive Ingredient Database (IID), especially during maximum use conditions as compared to IID.

- Accurately capturing MDD information is a crucial step to avoid potential toxicity or adverse reactions caused by the excipients.

- However, current MDD databases are built, maintained, and updated manually based on the information from drug labelings and FDA internal documents, and this is a laborious task.

Aim

- This project aims to develop and leverage a LLM to extract the MDD from FDA documents to automate the information retrieval and database management.
- Establish a workflow for the LLM to best reflect FDA's internal guideline for determining the MDD.
- Extract MDD from FDA's drug labelings.

Progressive Approaches

- Retrieval augmented generation (RAG)[1,2] – generate responses based on facts fetched from relevant sections from the provided drug label to enhance the reliability of the model.

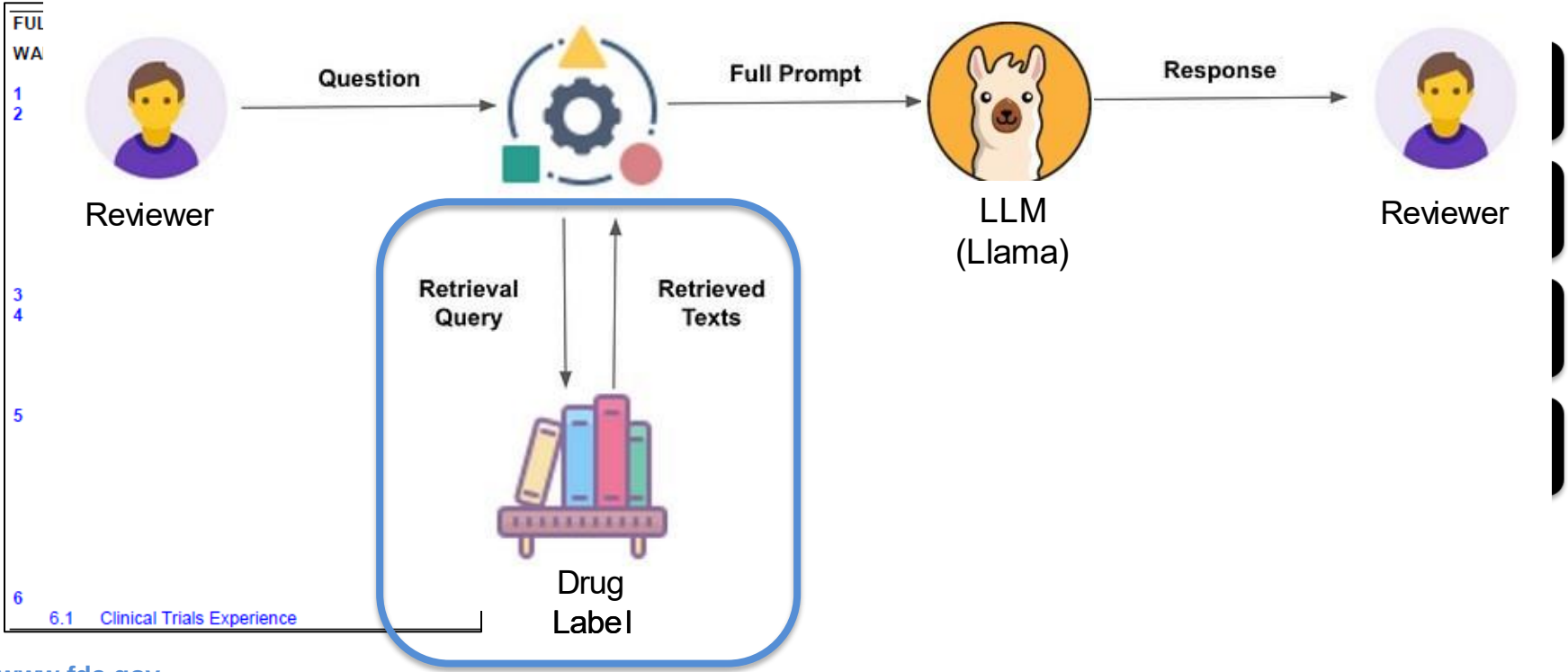
- Prompt engineering – instruct the model to infer the scenario with the correct MDD based on the retrieved content.

- Agentic workflow – handle more complex decision-making processes.

Pilot Study

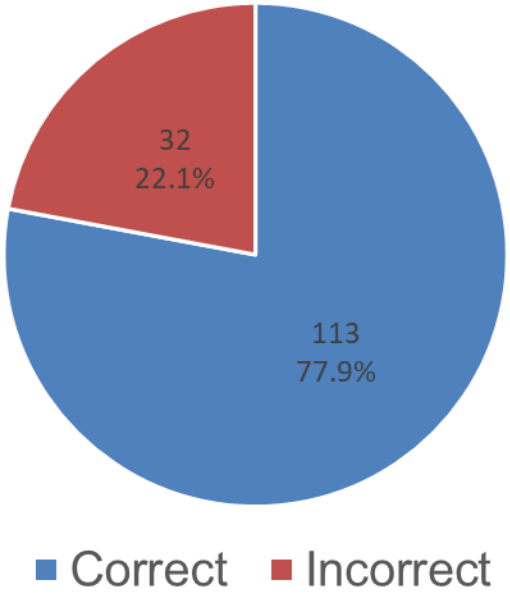
- A total of 148 drug products randomly selected from Immunology Interest Group (IIG) MDD Database with annotated MDD for various dosage forms.
- Drug labelings are obtained from Drugs@FDA, which are publicly available.
- The highest possible MDD value is an important consideration for the IIG.

Retrieval Augmented Generation



Prompt Engineering

- Simply asking “What is the maximum daily dose of ____?” will not provide correct answer.



Prompt Engineering

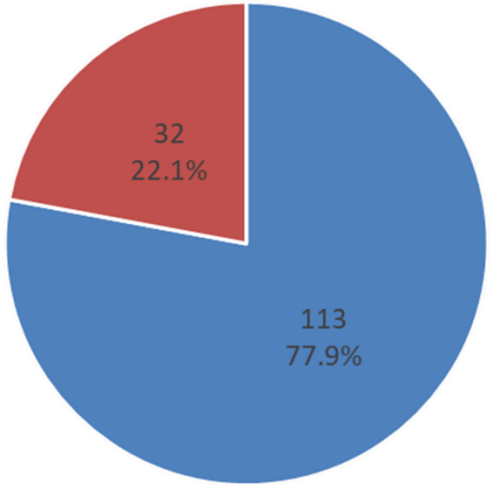
- Simply asking “What is the maximum daily dose of ____?” will not provide correct answer.

- Additional instructions need to be provided.

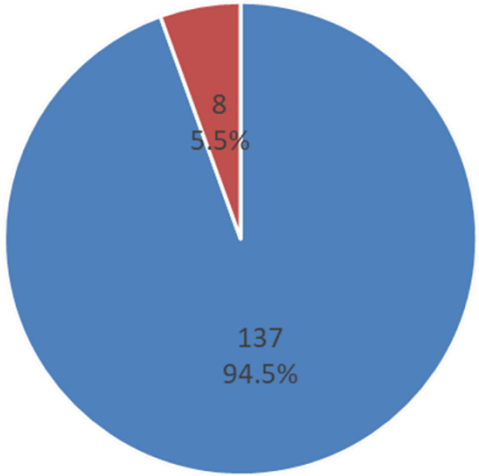
NDA #	Trade Name	Generic Name	MDD	MDD (LLM)	Notes
205525	SYNDROS	DRONABINOL	75.6 mg/m^2	16.8 mg/m^2	Special condition: “…8.4 mg twice daily for anorexia…and 12.6 mg/m^2 per dose for 4 to 6 doses per day for nausea and vomiting…”

Results

- 3 out of 148 cases were opioid products (the maximum theoretical daily dose instead of MDD) → excluded from the dataset.
- Specific instructions on handling dose modifications based on IIG's MDD selection criteria were provided to the model.



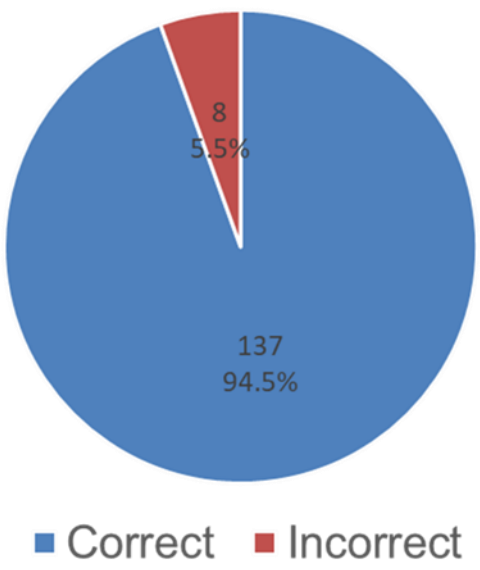
Baseline Prompt



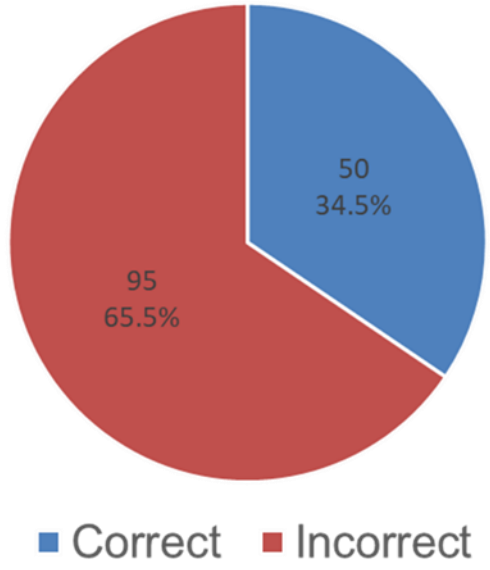
New Prompt

RAG vs. Pre-training

- The new prompt but without RAG (results based purely on the pre-trained knowledge)
- LLM answered strictly based on drug labelings as intended.



New Prompt



New Prompt (without RAG)

Ongoing Effort

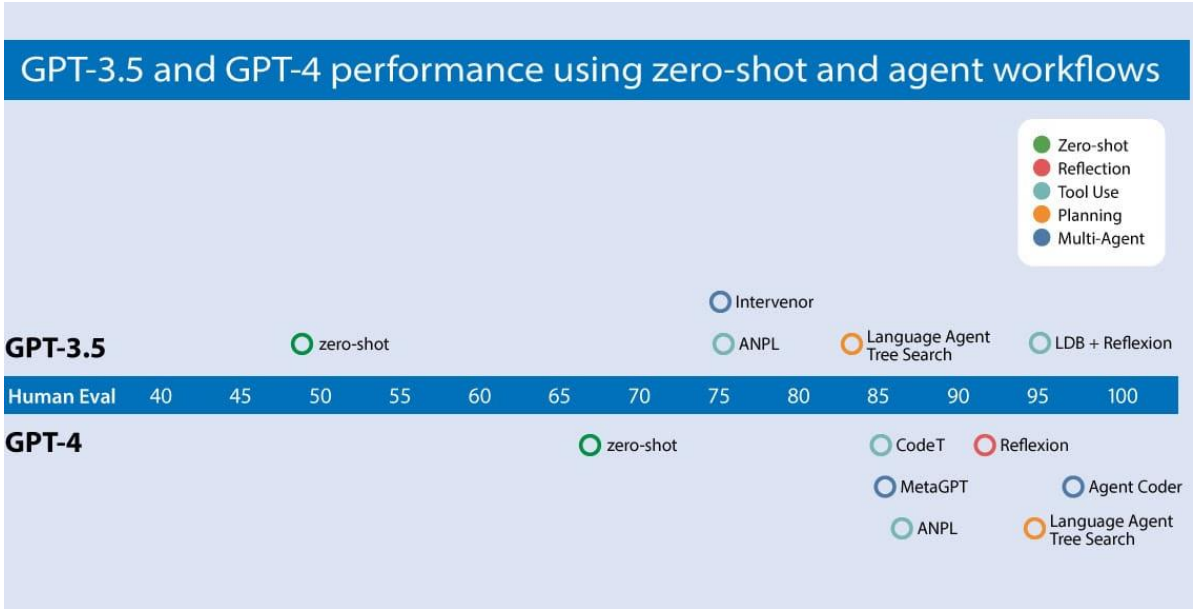
- The standard clinical principles for determining and documenting the MDD of drug products in the internal FDA guideline outline more processes and considerations than simply extracting the highest dose allowed in a day as done in the pilot study to determine the MDD.

- Although the guideline provides detailed processes and standard principles for reviewers to ensure consistency in determining the MDD for similar drug products, there is still potential for reviewer-dependent variability.

- Implement the process and principles in LLM to minimize the variability and to automate the MDD determination process.

Agentic Workflow

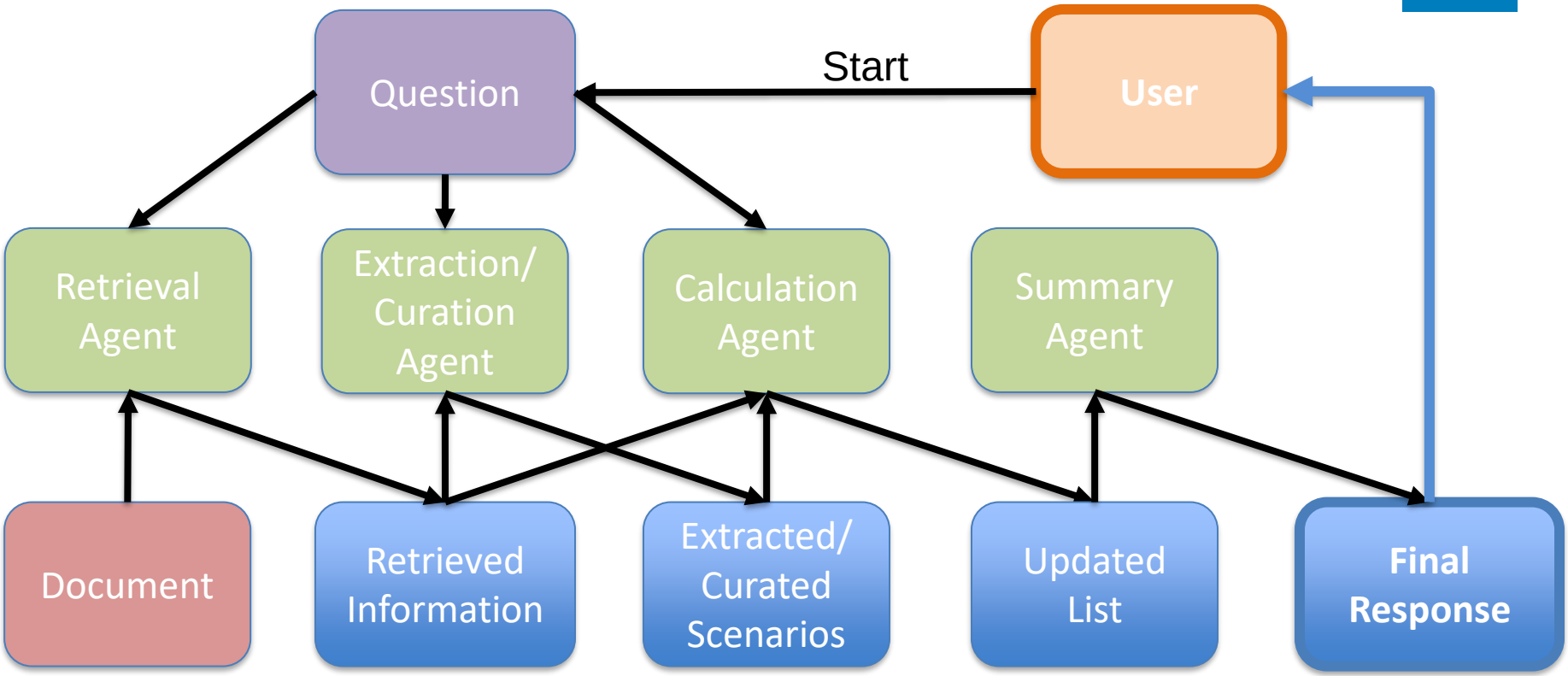
- Agentic workflow can be leveraged to improve the accuracy of LLM in deducing answers through complex decision processes.



GPT-3.5 (zero shot) was 48.1% correct. GPT-4 (zero shot) does better at 67.0%. However, the improvement from GPT-3.5 to GPT-4 is dwarfed by incorporating an iterative agent workflow. Indeed, wrapped in an agent loop, GPT-3.5 achieves up to 95.1%.

https://www.deeplearning.ai/the-batch/issue-241/

Agentic Workflow



Example - Information Retrieval

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use ALECENSA safely and effectively. See full prescribing information for ALECENSA.

ALECENSA® (alectinib) capsules, for oral use
Initial U.S. Approval: 2015

2 DOSAGE AND ADMINISTRATION

2.1 Patient Selection

Select patients for the treatment of metastatic NSCLC with ALECENSA based on the presence of ALK positivity in tumor tissue or plasma specimens *[see Indications and Usage (1) and Clinical Studies (14)]*. If ALK rearrangements are not detected in a plasma specimen, test tumor tissue if feasible.

Information on FDA-approved tests for the detection of ALK rearrangements in NSCLC is available at http://www.fda.gov/CompanionDiagnostics

2.2 Dosing and Administration

The recommended dose of ALECENSA is 600 mg orally twice daily *[see Clinical Pharmacology (12.3)]*. Administer ALECENSA until disease progression or unacceptable toxicity.

The recommended dose of ALECENSA in patients with severe hepatic impairment (Child-Pugh C) is 450 mg orally twice daily *[see Use in Specific Populations (8.7) and Clinical Pharmacology (12.3)]*.

ALECENSA should be taken with food. Do not open or dissolve the contents of the capsule. If a dose of ALECENSA is missed or vomiting occurs after taking a dose of ALECENSA, take the next dose at the scheduled time.

2.3 Dose Modifications for Adverse Reactions

The dose reduction schedule for ALECENSA is provided in Table 1.

Table 1: ALECENSA Dose Reduction Schedule

Dose reduction schedule	Dose level
Starting dose	600 mg taken orally twice daily
First dose reduction	450 mg taken orally twice daily
Second dose reduction	300 mg taken orally twice daily

Discontinue if patients are unable to tolerate the 300 mg twice daily dose.

Recommendations for dose modifications of ALECENSA in case of adverse reactions are provided in Table 2.

Table 2: ALECENSA Dose Modifications for Adverse Reactions

Criteria[a]	ALECENSA Dose Modification
ALT or AST elevation of greater than 5 times upper limit of normal (ULN) with total bilirubin less than or equal to 2 times ULN	Temporarily withhold until recovery to baseline or to less than or equal to 3 times ULN, then resume at reduced dose as per Table 1.
ALT or AST elevation greater than 3 times ULN with total bilirubin elevation greater than 2 times ULN in the absence of cholestasis or hemolysis	Permanently discontinue ALECENSA.
Total bilirubin elevation of greater than 3 times ULN	Temporarily withhold until recovery to baseline or to less than or equal to 1.5 times ULN, then resume at reduced dose as per Table 1.
Any grade treatment-related interstitial lung disease (ILD)/pneumonitis	Permanently discontinue ALECENSA.
Grade 3 renal impairment	Temporarily withhold until serum creatinine recovers to less than or equal to 1.5 times ULN, then resume at reduced dose.
Grade 4 renal impairment	Permanently discontinue ALECENSA.
Symptomatic bradycardia	Withhold ALECENSA until recovery to asymptomatic bradycardia or to a heart rate of 60 bpm or above. If contributing concomitant medication is identified and discontinued, or its dose is adjusted, resume ALECENSA at previous dose upon recovery to asymptomatic bradycardia or to a heart rate of 60 bpm or above. If no contributing concomitant medication is identified, or if contributing concomitant medications are not discontinued or dose modified, resume ALECENSA at reduced dose (see Table 1) upon recovery to asymptomatic bradycardia or to a heart rate of 60 bpm or above.
Bradycardia[b] (life-threatening consequences, urgent intervention indicated)	Permanently discontinue ALECENSA if no contributing concomitant medication is identified. If contributing concomitant medication is identified and discontinued, or its dose is adjusted, resume ALECENSA at reduced dose (see Table 1) upon recovery to asymptomatic bradycardia or to a heart rate of 60 bpm or above, with frequent monitoring as clinically indicated. Permanently discontinue ALECENSA in case of recurrence.
CPK elevation greater than 5 times ULN	Temporarily withhold until recovery to baseline or to less

Example - Extraction & Curation

- Extract a list of all possible dosing scenarios.

	Condition	Source	Rare Condition	Population
Scenario 1	Recommended dose for adults with normal hepatic function	"The recommended dose of ALECENSA is 600 mg orally twice daily."	No	Adult
Scenario 2	Recommended dose for patients with severe hepatic impairment (Child-Pugh C)	"The recommended dose of ALECENSA in patients with severe hepatic impairment (Child-Pugh C) is 450 mg orally twice daily."	Yes	Adult
Scenario 3	First dose reduction due to adverse reactions	“First dose reduction: 450 mg taken orally twice daily”	No	Adult

Example - Extraction & Curation

- Create a curated list of dosing scenarios.

	Condition	Source	Rare Condition	Population
Scenario 1	Recommended dose for adults with normal hepatic function	"The recommended dose of ALECENSA is 600 mg orally twice daily."	No	Adult
Scenario 3	First dose reduction due to adverse reactions	“First dose reduction: 450 mg taken orally twice daily”	No	Adult
Scenario 4	Second dose reduction due to adverse reactions	"Second dose reduction: 300 mg taken orally twice daily"	No	Adult

Example - Calculation

	Condition	Source	Thought Process	Calculation Steps	MDD
Scenario 1	Recommended dose for adults with normal hepatic function	"The recommended dose of ALECENSA is 600 mg orally twice daily."	Since the dosing frequency is twice daily, this implies two administrations per day. Given that each administration is 600 mg, we will calculate the MDD by multiplying the dose per administration by the number of administrations per day.	MDD = dose per administration * number of administrations per day = 600 mg/administration * 2 administrations/day	1200 mg
Scenario 3	First dose reduction due to adverse reactions	“First dose reduction: 450 mg taken orally twice daily”	Similar to Scenario 1, since the dosing frequency is twice daily, implying two administrations per day. We calculate the MDD by multiplying the reduced dose per administration by the number of administrations per day.	MDD = dose per administration * number of administrations per day = 450 mg/administration * 2 administrations/day	900 mg
Scenario 4	Second dose reduction due to adverse reactions	"Second dose reduction: 300 mg taken orally twice daily"	Following the same logic as before, with a dosing frequency of twice daily, we calculate the MDD by multiplying the second reduced dose per administration by the number of administrations per day.	MDD = dose per administration * number of administrations per day = 300 mg/administration * 2 administrations/day	600 mg

Example - Summary and Evaluation

Condition	Recommended dose for adults with normal hepatic function
Source	"The recommended dose of ALECENSA is 600 mg orally twice daily."
Thought Process	Since the dosing frequency is twice daily, this implies two administrations per day. Given that each administration is 600 mg, we will calculate the MDD by multiplying the dose per administration by the number of administrations per day.
Calculation Steps	MDD = dose per administration * number of administrations per day = 600 mg/administration * 2 administrations/day
MDD	1200 mg
Overall Summary	For Scenario 1, the Condition is Recommended dose for adults with normal hepatic function, Source states The recommended dose of ALECENSA is 600 mg orally twice daily, resulting in an MDD of 1200 mg. For Scenario 3, the Condition is First dose reduction due to adverse reactions, Source states First dose reduction: 450 mg taken orally twice daily, resulting in an MDD of 900 mg. For Scenario 4, the Condition is Second dose reduction due to adverse reactions, Source states Second dose reduction: 300 mg taken orally twice daily, resulting in an MDD of 600 mg. Therefore, the MDD is 1200 mg from Scenario 1.

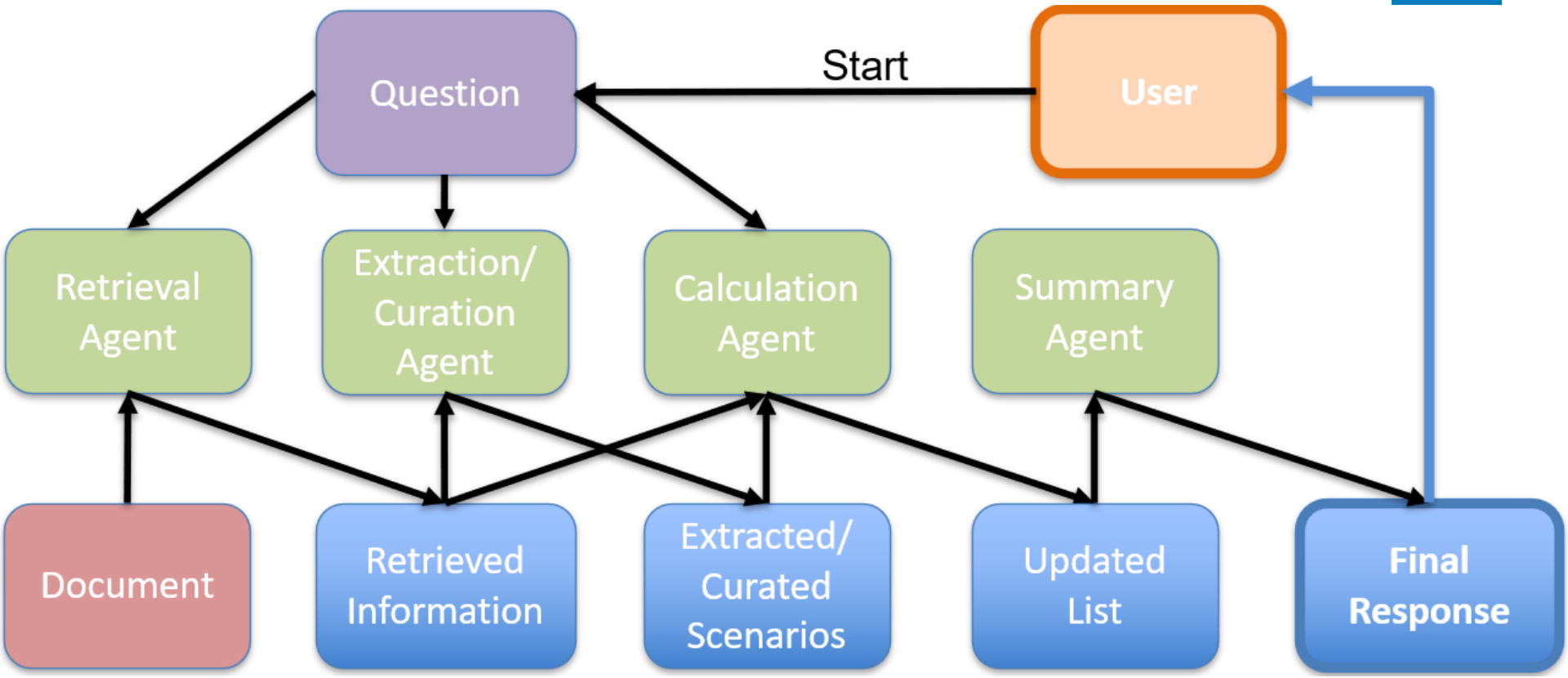
Future Effort

- Accurately incorporate FDA's internal guidelines and workflows established to determine MDD.

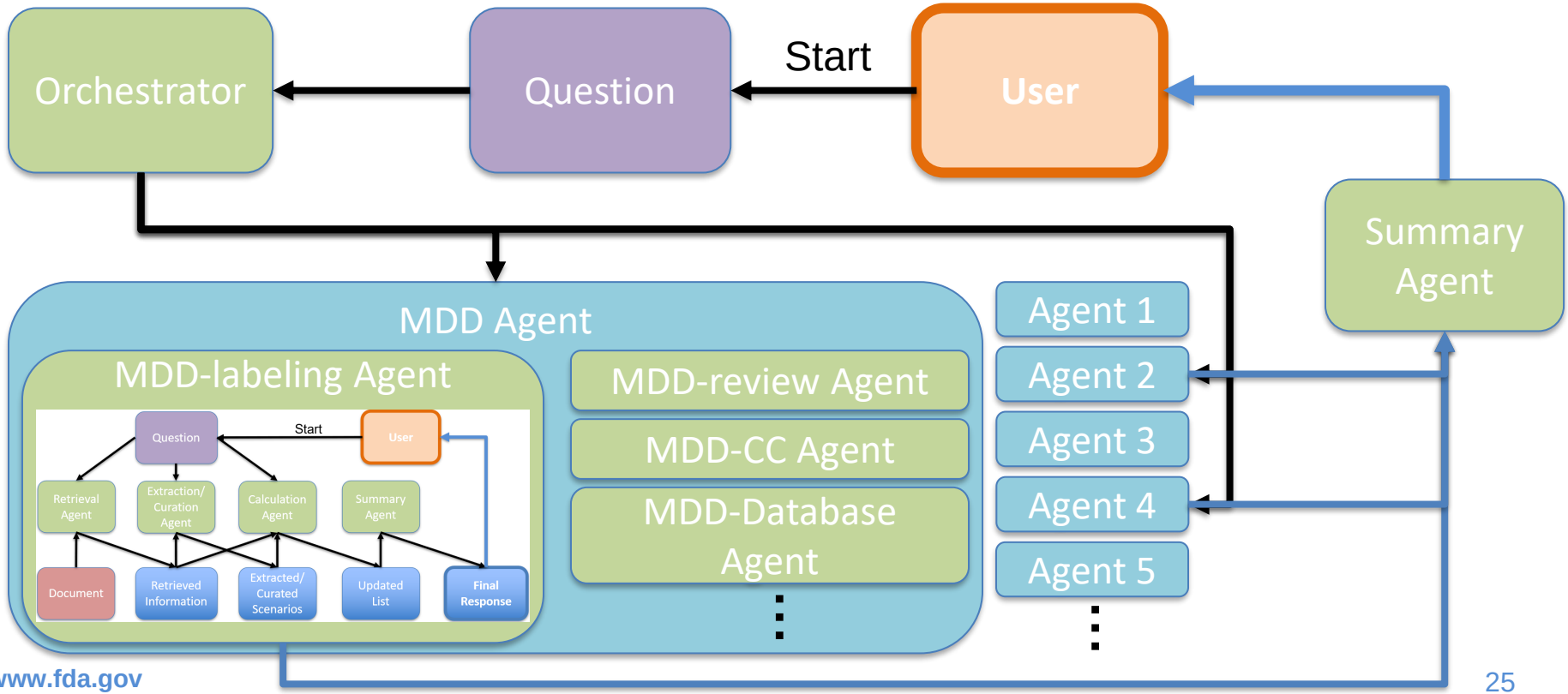
- Determine MDD for all products and compare with the existing database created by FDA reviewers during product-specific guidance (PSG) development.

- Can be expanded as a multi-agent system[3,4] that will include more supportive FDA documents such as controlled correspondences, consults, or prior reviews to enhance the MDD extraction.

Multi-Agent System



Multi-Agent System



Conclusion

- Promising potential in leveraging LLMs to streamline the current work procedure based on internal guidelines
- The established LLM pipeline can be applied/extended for pediatric populations.
- This tool will ultimately reduce the burden on the reviewers and reviewer-dependent variability in interpreting and deducing the MDD.

Acknowledgments

- **CDER/OGD/ORS**
  - Robert Lionberger
  - Lei Zhang
  - Lanyan (Lucy) Fang
  - Andrew Babiskin
  - Meng Hu
  - Jing (Jenny) Wang
  - Yusheng (Jason) Jiang
  - Markham Luke
  - Yan Wang
  - Myong-Jin Kim
  - Hye-Lim Lim
  - Karen Li
  - Qi Zhang
- **CDER/OGD/OSCE/DPTR**
  - Robert Dorsam
  - Melanie Mueller
  - Chanchal Gupta
- **CDER/OGD/OB/DBI**
  - Qing Liu
  - Amanda Jones

References

1. Gao Y et al. Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997, 2023.

2. Available on Hugging Face (https://huggingface.co/BAAI/bge-small-en-v1.5)

3. Bousetouane F. Agentic systems: A guide to transforming industries with vertical AI agents. arXiv preprint arXiv:2501.00881, 2025.

4. Fourney A et al. Magentic-one: A generalist multi-agent system for solving complex tasks. arXiv preprint arXiv:2411.04468, 2024.


FDA
U.S. FOOD & DRUG
ADMINISTRATION