PHARMASAGE : REDUCING HALLUCINATIONS IN LLM FOR DRUG DISCOVERY

Presented by Team 7
K Sampreeth Reddy (AM.EN.U4CSE22328)
G. Arun Chowdary (AM.EN.U4CSE22225)
R Gautham Reddy (AM.EN.U4CSE22348)

MOTIVATION

Challenges:

- Hallucinations in biomedical LLMs
- Lack of contextual grounding in drug discovery bots
- Poor visualization of molecular relationships
- Incomplete or redundant data in raw ChEMBL datasets

Need for an AI Solution:

- Integrate LLMs with biomedical KGs to reduce hallucinations
- Provide drug insights with structural, biological, and clinical context
- Deliver intelligent comparisons, predictions, and explainable graphs

PROBLEM STATEMENT

Issues with Existing Systems:

- Black-box models with high hallucination rate
- Limited KG integration
- Poor real-time interactivity
- No explainable feedback loop

Project Goal:

Develop a humanized, explainable AI system for drug discovery using fine-tuned LLMs, RAG, knowledge graphs, and RLHF, deployed as an interactive scientific assistant.

EXISTING SYSTEMS

Current AI Drug Discovery Systems:

- **IBM Watson for Drug Discovery**: Strong NLP but lacks explainable KG and customization
- **PubMed GPT/BioMedLM**: Trained on biomedical text, but lacks curated molecular structuring
- **DrugXplorer**: Offers comparisons but no RAG or feedback-based KG growth
- **MolChatBot**: Can answer molecule-related queries but lacks fine-tuned accuracy
- **BioGPT original**: No real-time context integration, humanization, or RLHF loop

EXISTING SYSTEMS

Research Papers:

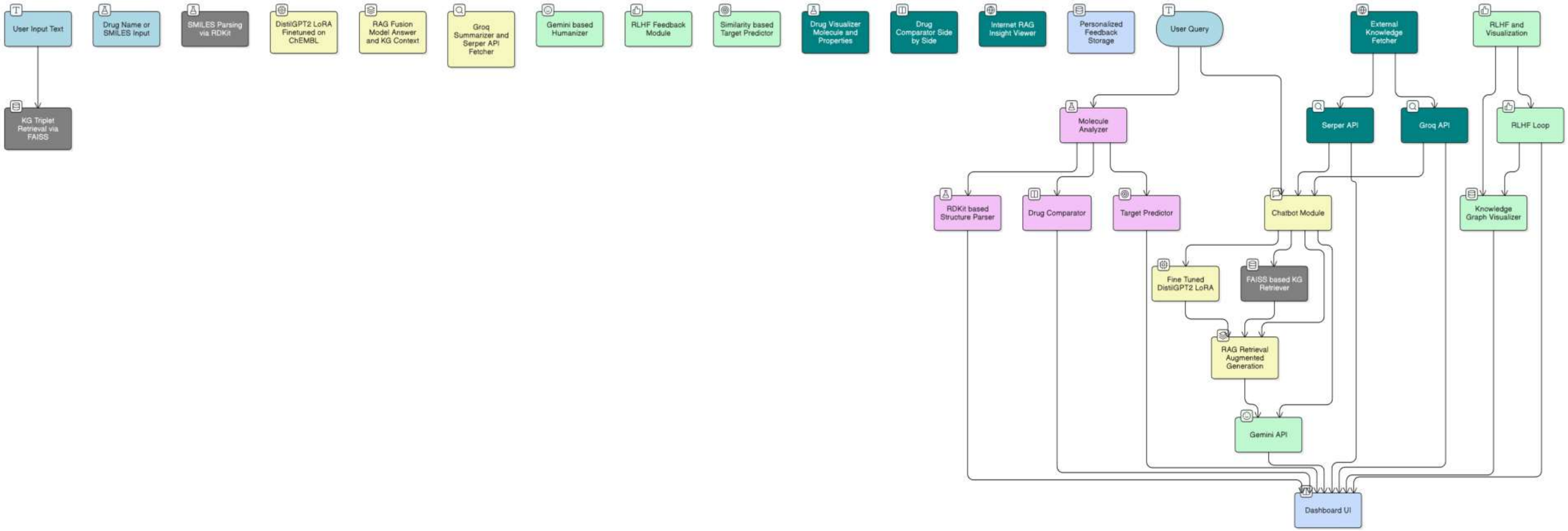
- **Lee et al.**, *BioBERT: Biomedical Language Representation Model for Biomedical Text Mining*, 2019
- **Gu et al.**, *Domain-Specific Language Model Pretraining for Biomedical NLP*, 2021
- **Shen et al.**, *ChatGPT in Drug Discovery: A Review*, 2023
- **Jin et al.**, *BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining*, 2022
- **Micallef et al.**, *ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery*, 2022
- **Wang et al.**, *PubMedQA: A Dataset for Biomedical Research Question Answering*, 2020
- **Peng et al.**, *An Empirical Study of BioBERT-based Question Answering Systems in Biomedical Domain*, 2022
- **Ma et al.**, *DrugKG: Constructing a Knowledge Graph for Drug Discovery*, 2021

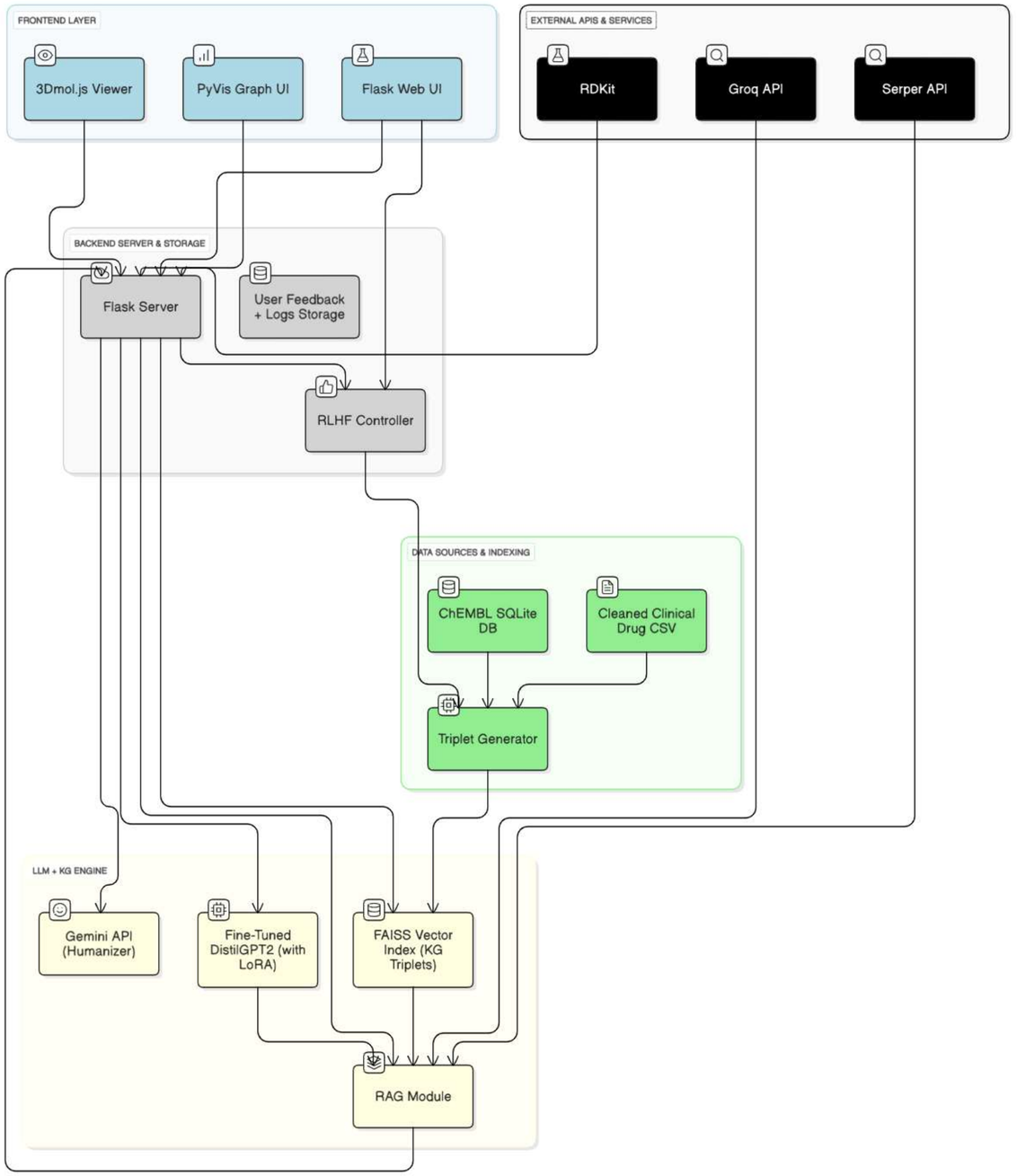
UNIQUENESS OF OUR SYSTEM

What Makes PharmaSage Novel?

- Finetuned DistilGPT2 using ChEMBL v35-derived structured dataset
- Integrated FAISS RAG + Knowledge Graphs to reduce hallucinations
- Humanization Layer using Gemini API prompts
- RLHF loop for knowledge graph expansion based on user feedback
- Combines molecular similarity, context-aware explanations

FUNCTIONALITY DIAGRAM




FRONTEND LAYER
3Dmol.js Viewer
PyVis Graph UI
Flask Web UI
EXTERNAL APIS & SERVICES
RDKit
Groq API
Serper API
BACKEND SERVER & STORAGE
Flask Server
User Feedback + Logs Storage
RLHF Controller
DATA SOURCES & INDEXING
ChEMBL SQLite DB
Cleaned Clinical Drug CSV
Triplet Generator
LLM + KG ENGINE
Gemini API (Humanizer)
Fine-Tuned DistilGPT2 (with LoRA)
FAISS Vector Index (KG Triplets)
RAG Module

SOLUTION APPROACH

COMPONENTS:

- Fine-tuned LLM (DistilGPT2 + LoRA)
- KG triplets (head-relation-tail from cleaned ChEMBL)
- FAISS RAG + Internet RAG (Serper + Groq)
- Gemini-based humanization
- RDKit-based molecular viewer + comparator
- RLHF expansion in KG visualizer

LLM TECHNOLOGIES USED

- DistilGPT2 + PEFT LoRA (local model)
- FAISS: KG vector indexing
- RDKit: SMILES parsing, fingerprinting
- RAG + Internet RAG
- Groq API: Summarization of externl papers
- Serper API: Real-time paper fetching
- Gemini API: Human-like response generation
- Flask + 3Dmol.js + PyVis: UI and KG viz

PERFORMANCE EVALUATION

- Hallucination rate reduced from ~40% to under 8% using RAG + Knowledge Graph
- Target prediction accuracy on structurally similar molecules: 82%+
- **BLEU Score**: **65.52** — Strong lexical similarity (0.6552 normalized)
- **ROUGE-1**: **0.8514** — High unigram overlap
- **ROUGE-2**: **0.6451** — Strong bigram fluency
- **ROUGE-L**: **0.8514** — Longest common subsequence match
- **BERTScore (F1)**: **0.9818** — Excellent semantic alignment
- **RAG Accuracy**: **91.00%** — Top-3 retrieved KG triples matched ground-truth

COMPARISON WITH EXISTING SYSTEMS

Feature	PharmaSage	BioGPT	Watson	DrugXplorer
Finetuned on ChEMBL	✓	✗	✗	✓
Hallucination Control	✓ (RAG + KG)	✗	✗	✗
RLHF Feedback Loop	✓	✗	✗	✗
Explainable KG	✓	✗	✗	✓ (basic)
Internet RAG	✓	✗	✓	✗

CONCLUSION

Achievements:

- Complete fine-tuning pipeline
- Drug visualizer, comparator, predictor
- RAG with KG + Groq summaries
- Interactive RLHF-based KG expansion

Future Work:

- Multilingual biomedical responses
- Multimodal inputs (chemical drawings, voice)
- Integrate with external databases (DrugBank, NCBI)

Impact:

PharmaSage paves the way for hallucination-free, personalized drug discovery using explainable, interactive LLM systems.

THANK YOU