

Ai-Powered Ethical Decision Making Simulations With **Mistral**

Names: K. Ramakrishna Siddhardh,22BIT0012(VIT Vellore)

A. Sai Ganesh,22BIT0023(VIT Vellore)

R. Subramanya Harideep,22BCE3391 (VIT Vellore)

M. Varun Teja,22BDS0415(VIT Vellore)

1. Introduction

- AI-Powered Ethical Decision-Making Simulations With Mistral is a cutting-edge web application that uses Mistral's generative AI models and Streamlit to simulate real-world ethical dilemmas. By allowing users to enter natural language scenarios, the system generates detailed ethical evaluations, enabling better understanding of the implications of policy decisions, social issues, or corporate strategies. This makes complex moral reasoning accessible even to non-experts and supports education, governance, and research use cases.

2. Technologies Used

- This project leverages modern tools to deliver ethical AI simulation via an interactive platform:
- **Python** – backend logic and integration
- **Streamlit** – for building the responsive, stylish web interface
- **Mistral AI** – generates simulated ethical scenarios using natural language understanding
- **dotenv (optional)** – to securely manage API keys
- **LangChain (future scope)** – for enhancing context management and chaining reasoning steps

3. Application Context

- The app does not use a traditional database schema. Instead, it focuses on free-text ethical challenge inputs and AI-driven scenario generation. Users provide prompts like:

- "Should autonomous cars prioritize passenger safety over pedestrian safety?"
- "Is it ethical to use AI surveillance in public spaces?"
- The Mistral model evaluates and responds with:
- Simulated consequences
- Ethical principles at stake
- Possible real-world impact

4. Prompt Design

- Ethical scenarios are generated by prompting Mistral with user input and a predefined system role. While this app uses a simple user-initiated prompt, a future improvement may include few-shot prompting or chain-of-thought reasoning via LangChain.
- Example prompt:
User: *"Should a company track employees' keystrokes to ensure productivity?"*
Mistral Output: A nuanced breakdown of privacy vs. productivity, legal implications, psychological impact, and potential consequences.

5. Application Workflow

- The working of this ethical simulator follows a clean and intuitive flow:
- **User Input:** A natural language ethical question is entered.
- **Model Invocation:** The Mistral API receives the query using the mistralai Python package.
- **AI Processing:** Mistral generates a detailed ethical scenario.
- **Output Rendering:** The response is presented in a stylized Streamlit interface with visual emphasis.
- **User Feedback Loop:** Users can iterate or test different prompts interactively.

6. Mistral Prompt Example

- The system currently sends a raw user message to the Mistral model. Below is a basic example:
- **Prompt:**

"Is it ethical for hospitals to use AI to prioritize treatment based on survival probability?"

➤ **Response Output (by Mistral):**

- Discusses ethical implications of triage AI
- Balances efficiency with equity
- Explores fairness, medical ethics, and public trust

7. Streamlit UI Design

- The interface is crafted with futuristic design principles:
- Animated headers
- Glowing neon buttons
- Glassmorphism for scenario boxes
- Orbitron and Roboto Mono fonts
- Dark gradient theme with CSS transitions
- This makes the user experience modern, immersive, and intuitive for exploring sensitive ethical content.

8. Installation Requirements

- To run the app locally, ensure the following Python libraries are installed:
- `pip install streamlit mistralai langchain`

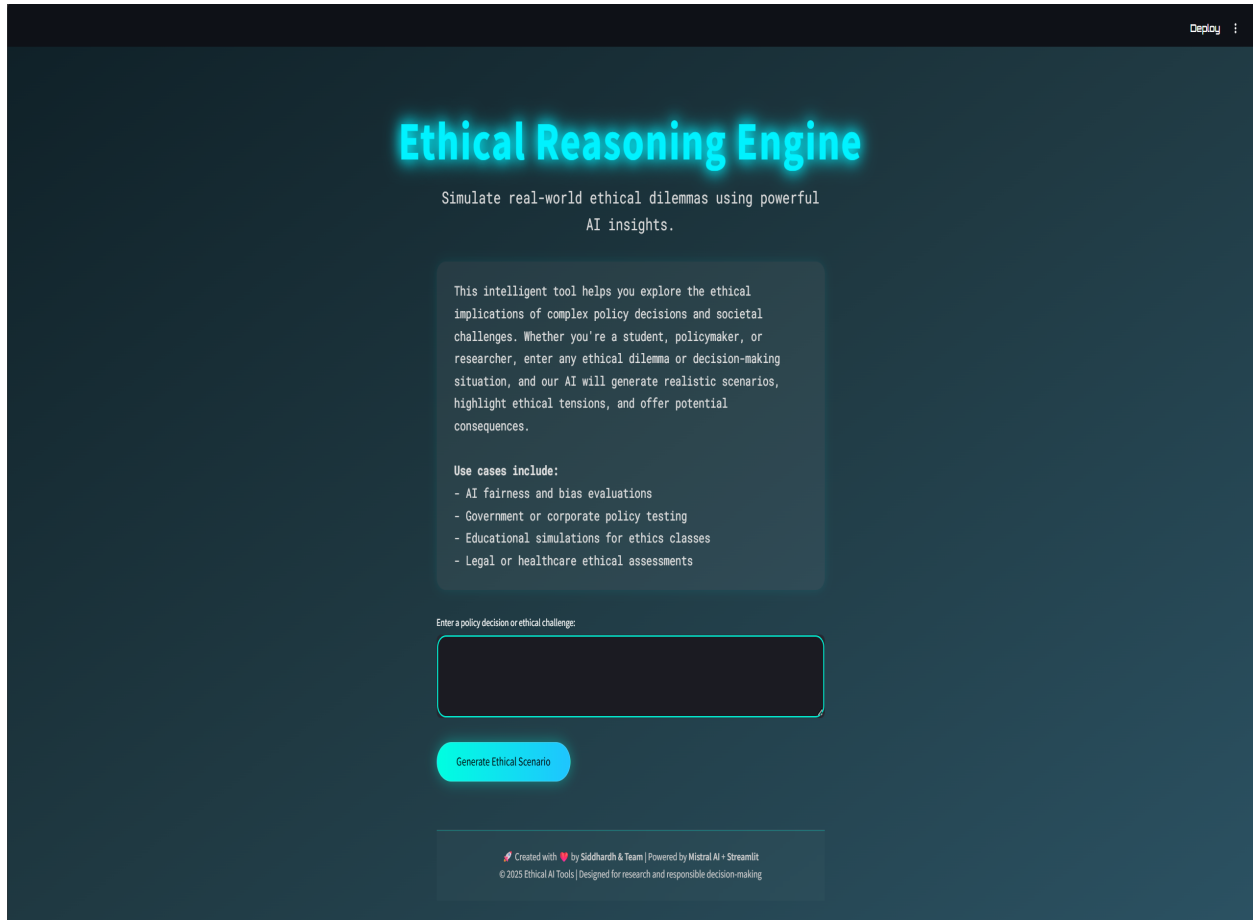
9. Conclusion

- AI-Powered Ethical Decision-Making Simulations With Mistral is a practical and engaging tool to democratize ethical thinking. Whether for educational simulations, policy research, or responsible AI testing, it empowers users to explore critical dilemmas and understand potential consequences using LLM-based insights. Future enhancements may include scenario chaining, role-based personas, or integration with real-world datasets.

10. Project Demo Video Link:

<https://drive.google.com/file/d/1e6mFxlQx0iw9Yb8EbuZfS4zVzeQGNn32/view?usp=sharing>

11. Demo Photos of project:



Enter a policy decision or ethical challenge:

Is it ethical to deploy predictive policing systems in minority neighborhoods where historical crime data is biased?



Generate Ethical Scenario

Ethical Scenario Ready

The deployment of predictive policing systems in minority neighborhoods, particularly where historical crime data is biased, raises significant ethical concerns. Here are key considerations:

1. Bias in Historical Data

Predictive policing relies on historical crime data, which often reflects systemic biases in policing. Minority neighborhoods have been over-policed for decades, leading to disproportionate arrest rates not because of higher crime rates, but due to discriminatory practices. Using this data to predict future crime risks reinforcing these biases, creating a feedback loop of over-policing.

2. Reinforcement of Structural Inequality

If predictive policing disproportionately targets minority communities, it can exacerbate racial disparities in the criminal justice system. This could lead to increased surveillance, harassment, and unjustified stops, further eroding trust between law enforcement and marginalized groups.