

# Synthetic Data Generation for Real- World Simulations

Presented by

Team Betelgeuse

Gokul Raja R

Syed Farith C

Krishnaprasath K

Padmavathi S



# Problem Statement

## Data Acquisition Challenges

Acquiring high-quality real-world data is difficult due to privacy concerns, high costs, and the risk of exposing sensitive information.

## Impact on Industries

This poses a significant problem for industries like healthcare, finance, and customer service, where using real data can lead to compliance issues or breaches of privacy.



# Project Overview

- 1 Provide businesses with realistic synthetic data for safe and efficient training, testing, and development.
- 2 Capture and understand user needs, new service requests, and customer feedback through conversation analysis.
- 3 Ensure data privacy by avoiding the use of real personal or sensitive information.
- 4 Automate the process of notifying users via email once their synthetic data is generated, ensuring timely access and communication.





# System Flow

1

## Data Capture

Conversations between users and service providers are captured and processed by our system.

2

## Function Calling and Data Extraction

Using natural language processing (NLP) techniques, the system dynamically calls specific functions to extract key information from these conversations.

3

## Data Storage

The extracted information is stored in MongoDB, our primary database, which is ideal for handling structured data efficiently.

4

## Synthetic Data Generation

Our system generates synthetic datasets based on the extracted information.

5

## Email Notification

Once the synthetic data is generated, an automated email is sent to the user's email address, informing them that their data is ready.

# Technologies and Tools Used

## MongoDB

MongoDB serves as our primary database, storing structured data such as business details and customer feedback.

## AI Models

- Llama 3.1 70B Instruct Turbo
- Llama 3 Groq 70B Tool Use
- Mixtral 8x7B

## Other Tools

We use Celery, a distributed task queue, to manage background tasks like data extraction and processing. Redis acts as the message broker, ensuring smooth and efficient task execution.

# Business Value



## Enhanced Decision-Making

By simulating various scenarios, businesses can make better decisions and predict customer behavior more accurately.



## Cost Efficiency

Synthetic data eliminates the need for costly data collection efforts, offering a scalable and cost-effective solution.



## Privacy Compliance

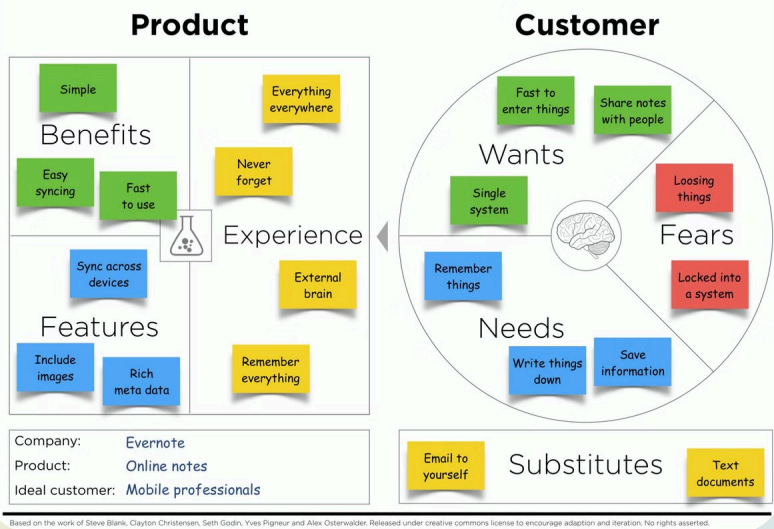
Using synthetic data helps companies comply with data protection regulations, avoiding the use of real personal information.



## Edge Device Deployment

Once the synthetic data is generated, an email is automatically sent to the user's email address, informing them that their data is ready. This ensures seamless communication and allows users to access their data promptly.

### Value Proposition Canvas







# Conclusion

Our synthetic data generation project combines cutting-edge AI models, dynamic function calling, efficient data processing, and seamless user communication to create a powerful tool for businesses. It enables the simulation of real-world scenarios, enhances decision-making, ensures data privacy, and supports deployment on edge devices. We are Team Betelgeuse, and we're proud to have developed this innovative solution. Thank you for your time!