

Synthetic Healthcare Data Generator

A comprehensive tool for generating, analyzing, and visualizing synthetic and artificial healthcare data.

🌟 Features

- **Artificial Data Generation:** Create artificial data using various statistical distributions.
- **Synthetic Data Generation:** Generate synthetic data based on real data while preserving statistical properties.
- **Advanced AI Methods:** Leverage modern AI techniques including:
 - GAN (Generative Adversarial Networks)
 - VAE (Variational Autoencoders)
 - LLM (Large Language Models via Mistral API)
- **Comparative Analysis:** Visualize and compare real vs. synthetic data distributions.
- **Dynamic Configuration:** Interactive interface for customizing data generation parameters.

🚀 Getting Started

Prerequisites

- Python 3.8+
- Required Python packages (see Installation)

Installation

1. Clone the repository:

```
bash
```

📄 Copier

```
git clone https://github.com/your-username/synthetic-healthcare-data.git
cd synthetic-healthcare-data
```

2. Create a virtual environment:

```
bash
```

📄 Copier

```
python -m venv venv
source venv/bin/activate # On Windows: venv\Scripts\activate
```

3. Install the required packages:

bash

 Copier

```
pip install -r requirements.txt
```

Running the Application

Start the Streamlit application:

bash

 Copier

```
streamlit run app.py
```

Project Structure

 Copier

```
synthetic-healthcare-data/  
├─ app.py           # Main application entry point  
├─ demo_app.py      # Core functionality for data generation  
├─ ai_methods_tab.py # AI-based methods implementation  
├─ requirements.txt  # Project dependencies  
├─ README.md        # This file  
└─ .gitignore       # Git ignore file
```

Usage Examples

Generating Artificial Data

1. Navigate to the "Données Artificielles" tab.
2. Add variables and configure their properties (type, distribution, parameters).
3. Set up correlations between numerical variables if needed.
4. Click "Generate Data" and explore the visualizations.
5. Download the generated data in CSV format.

Creating Synthetic Data

1. Navigate to the "Données Synthétiques" tab.
2. Upload a real dataset in CSV format.
3. Choose the generation method (Bootstrap or Gaussian Copula).

4. Generate synthetic data and analyze the comparison metrics.
5. Download the synthetic dataset.

Using AI Methods

1. Navigate to the "Méthodes d'IA" tab.
2. Choose between different AI-based generation approaches.
3. Configure the model parameters.
4. For LLM-based generation, provide your Mistral API key.
5. Generate data and analyze the results.

Privacy and Ethics

- This tool is designed for research, testing, and educational purposes.
- No real patient data is included or required.
- Generated data should not be used to make clinical decisions.
- Always ensure compliance with relevant regulations when using synthetic data.

Contributing

Contributions are welcome! Please feel free to submit a Pull Request.

License

This project is licensed under the MIT License - see the LICENSE file for details.

Contact

For questions or feedback, please open an issue on GitHub.