

AI Persona Mapper

Turn your words into patterns, habits, and a data-driven digital twin.

Overview

AI Persona Mapper is a production-quality Streamlit application that analyzes text to extract personality traits, thematic patterns, and habit formation projections. Using only Python standard libraries and basic data science tools, it provides deep insights into writing patterns without requiring external NLP/ML libraries.

Features

- **Text Pattern Analysis:** Token frequency, bigram detection, co-occurrence matrices
- **Theme Detection:** Identifies 12+ themes including work, social, tech, exercise, sleep
- **Personality Traits:** Computes 7 personality dimensions (Focus, Chaos, Energy, Discipline, Openness, Socialness, NightOwlness)
- **Habit Projection:** Monte Carlo simulation of habit formation curves with uncertainty bands
- **Visual Analytics:** Interactive charts including radar plots, heatmaps, and trend curves
- **History & Comparison:** Save and compare multiple analyses
- **Export Options:** CSV summaries, persona cards (PNG), and batch exports

Setup

Requirements

- Python 3.10+
- NumPy
- Pandas
- Matplotlib
- Streamlit

Installation

1. Clone or download the repository
2. Install dependencies:

```
bash
pip install streamlit numpy pandas matplotlib
```

3. Ensure the following file structure:

```
ai_persona_mapper/  
├── app.py  
├── core/  
│   ├── text_processing.py  
│   ├── themes.py  
│   ├── patterns.py  
│   ├── traits.py  
│   ├── habit.py  
│   ├── visuals.py  
│   └── export_utils.py  
├── assets/  
│   └── stopwords.txt  
└── README.md
```

Run Instructions

Navigate to the project directory and run:

```
bash  
  
streamlit run app.py
```

The app will open in your default browser at `http://localhost:8501`

Usage Guide

Basic Usage

1. Enter at least 20 characters of text in the main text area
2. Click "Analyze My Vibe" to generate analysis
3. Explore results across four tabs: Patterns, Personality, Habit, History/Compare

Advanced Options

- **Habit Keyword:** Specify a custom habit to track (defaults to dominant theme)
- **Horizon Days:** Set projection period (30-180 days)
- **Monte Carlo Runs:** Adjust simulation iterations for uncertainty bands
- **Random Seed:** Set for reproducible results
- **Bigrams/Heatmap:** Toggle additional visualizations

Export Features

- **CSV Summary:** Complete analysis metrics and insights
- **Persona Card:** Single-image summary with traits, themes, and habit curve

- **Markdown Report:** Formatted analysis report

How It Works

Text Processing Pipeline

1. Lowercasing and punctuation removal (preserves emojis)
2. Regex-based tokenization
3. Stopword filtering (120+ common English words)
4. Sentence segmentation for structural analysis

Personality Trait Computation

Each trait uses weighted rule-based scoring:

- **Focus:** Productivity keywords, low questions, structured writing
- **Chaos:** Vocabulary diversity, exclamations, topic jumping
- **Energy:** Active keywords, exercise themes, short sentences
- **Discipline:** Planning terms, consistent structure
- **Openness:** Curiosity markers, questions, theme diversity
- **Socialness:** People references, social themes
- **NightOwlness:** Night-related keywords and themes

Habit Projection Model

- Logistic growth curve with parameters derived from personality traits
- Monte Carlo simulation for uncertainty quantification
- Event detection: plateaus, slumps, milestones
- Personalized recommendations based on trait profile

Example Use Cases

- **Personal Journaling:** Track emotional patterns and themes over time
- **Writing Analysis:** Understand your writing style and vocabulary patterns
- **Habit Formation:** Project success rates for new habits based on personality
- **Self-Discovery:** Identify dominant themes and traits in your communication

Technical Details

- **Deterministic:** Results are reproducible with the same random seed
- **Modular Architecture:** Clean separation of concerns across modules
- **PEP8 Compliant:** Follows Python coding standards

- **Well-Documented:** Comprehensive docstrings and inline comments
- **Error Handling:** Graceful handling of edge cases and invalid inputs

Limitations

- English language only
- No external NLP models (rule-based approach)
- Minimum 20 characters required for meaningful analysis
- Theme detection based on keyword matching
- Personality traits use transparent but simplified scoring

License

This project is provided as-is for educational and personal use.

Acknowledgments

Built with Streamlit, NumPy, Pandas, and Matplotlib - powerful open-source tools for data science and visualization.