

Hotel Bias Detection - Quick Guide

Folder Structure

```
BIAS/
├── .env
├── requirements.txt
├── processed_boston_reviews.csv      # Input: Hotel reviews
├── hotel_bias_detection.py          # Core: Bias detection logic
├── generate_response/
│   └── generate_chatbot_response.py  # Step 1: Generate AI summaries
├── pipeline/
│   └── stage_bias_detection.py       # Step 2: Run bias detection
├── response/
│   ├── chatbot_responses.csv        # Generated AI responses (CSV)
│   └── chatbot_responses.parquet    # Generated AI responses (Parquet)
└── evaluation/results/
    ├── hotel-bias-scores.json       # Detailed bias data per hotel
    ├── hotel-bias-results.csv       # Results table (Excel-friendly)
    ├── hotel-bias-results.parquet   # Results table (binary)
    └── bias-detection-summary.json  # Overall statistics
```

Quick Start

1. Setup

```
# Install dependencies
pip install -r requirements.txt
```

```
# Add your API key to .env
echo "XAI_API_KEY=xai-your_key_here" > .env
```

2. Generate AI Responses

```
python generate_response/generate_chatbot_response.py
```

Output: `response/chatbot_responses.parquet`

3. Run Bias Detection

`python pipeline/stage_bias_detection.py`

Output: 4 files in `evaluation/results/`

4. View Results

Summary
`cat evaluation/results/bias-detection-summary.json`

Detailed
`open evaluation/results/hotel-bias-results.csv`



What Each File Does

File	Purpose	When to Use
<code>generate_chatbot_response_xai.py</code>	Create AI summaries of hotel reviews	Run first, once per dataset
<code>stage_bias_detection.py</code>	Check AI responses for bias	Run second, analyzes responses
<code>hotel_bias_detection.py</code>	Core detection logic	Auto-imported by pipeline
<code>chatbot_responses.parquet</code>	AI-generated summaries	Generated by step 1
<code>hotel-bias-results.csv</code>	Bias detection results	View in Excel
<code>bias-detection-summary.json</code>	Statistics overview	Quick status check



What Gets Detected

3 Types of Bias:

1. **Over-reliance on Negative** - AI too negative when reviews are mixed
 2. **Missing Data Acknowledgment** - AI doesn't mention limited reviews (<4)
 3. **Rating Disparity** - AI negative despite 4+ star ratings
-

Configuration

Change number of hotels:

```
# In generate_chatbot_response_xai.py
NUM_HOTELS = 5 # Change to 10, 20, etc.
```

Adjust bias thresholds:

```
# In hotel_bias_detection.py
self.config = {
    'neg_sentiment_threshold': 0.7,    # 70% negativity = bias
    'min_reviews_threshold': 4,       # <4 reviews = sparse data
    'rating_disparity_threshold': 4.0 # 4+ stars = good rating
}
```

requirements.txt

```
pandas>=2.0.0
pyarrow>=12.0.0
python-dotenv>=1.0.0
openai>=1.0.0
```
