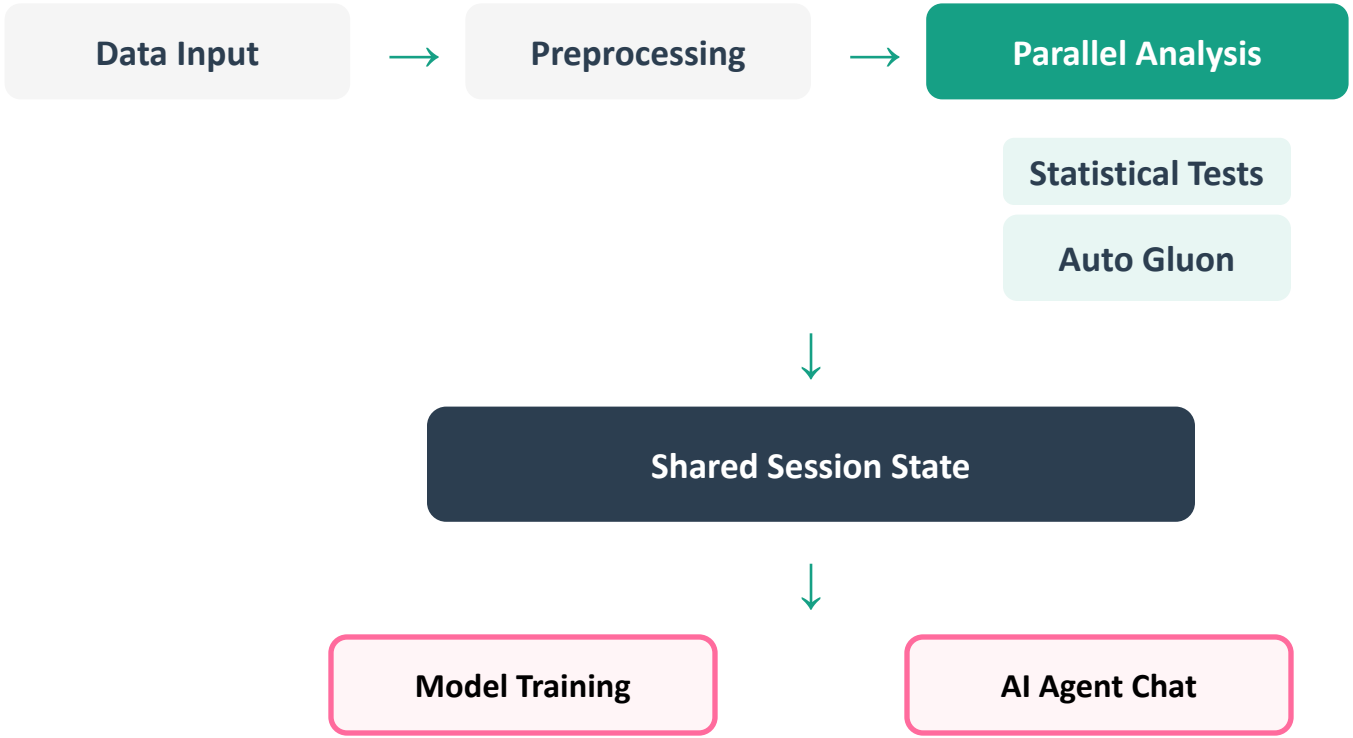
Statistical AI Agent for Dataset Analysis

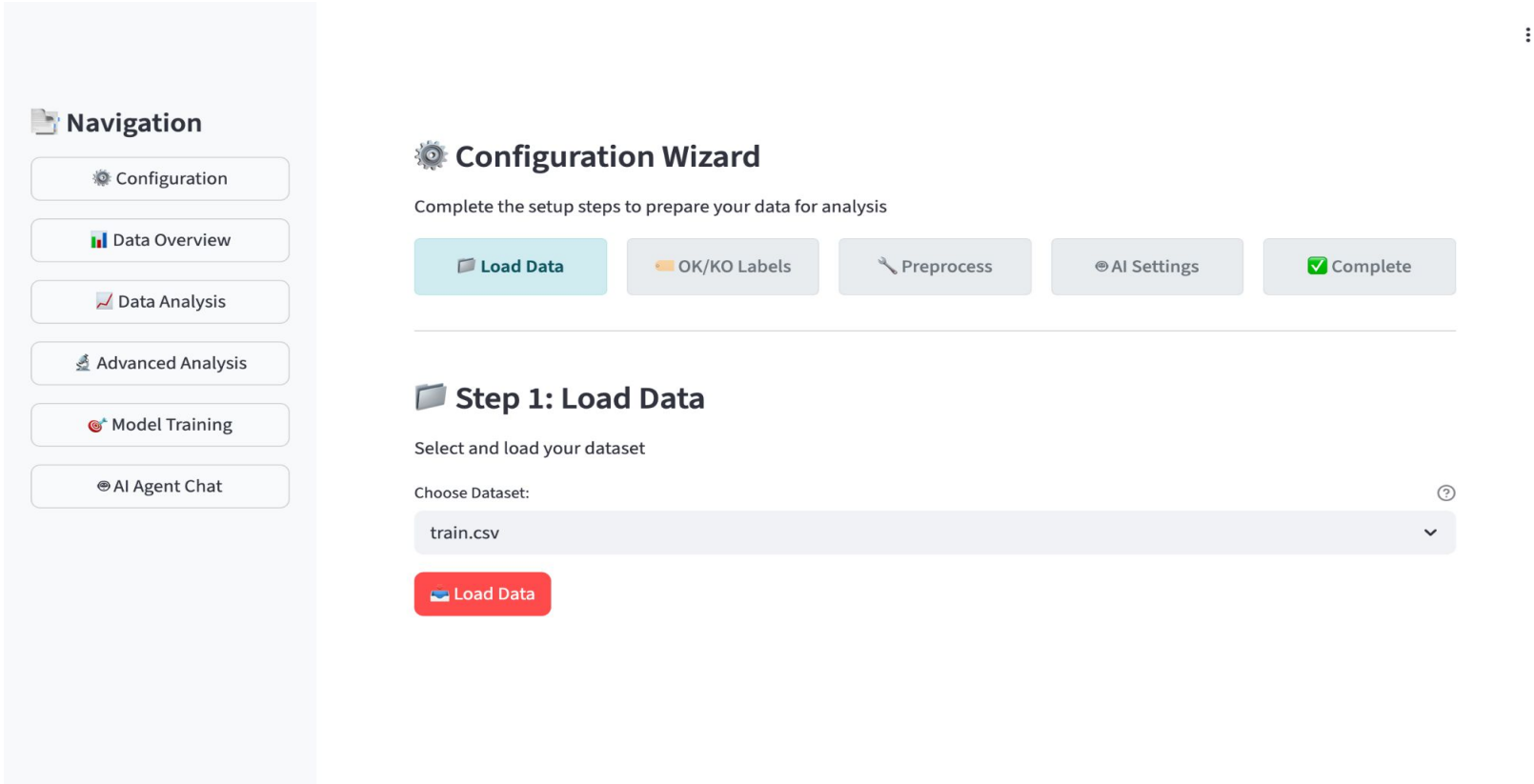
A Parallel-Validation Statistical Analysis System with AI-Powered Chat Interface

Lei GAO (s327756) • Lan DENG (s338219)

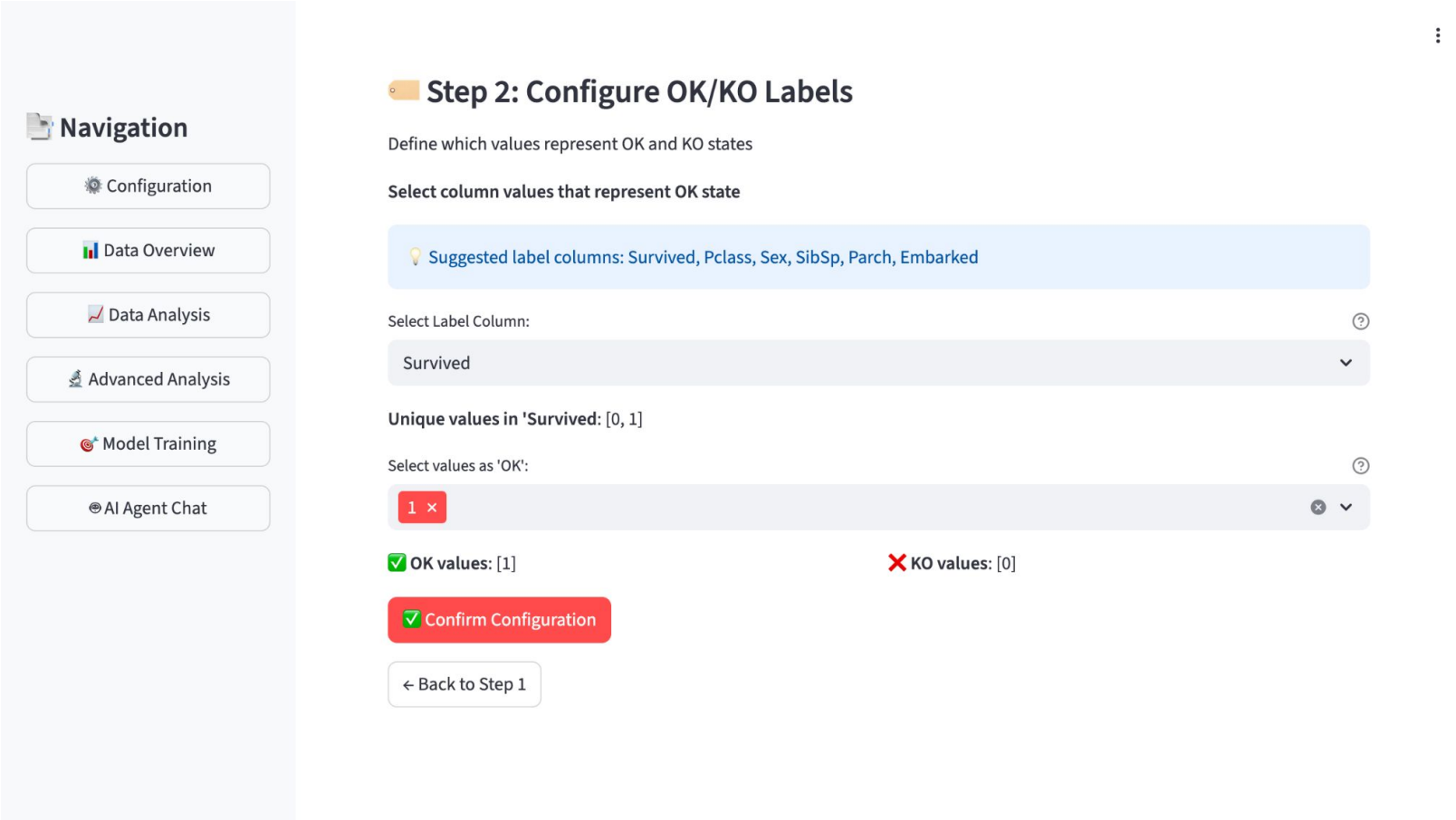
System Architecture Overview



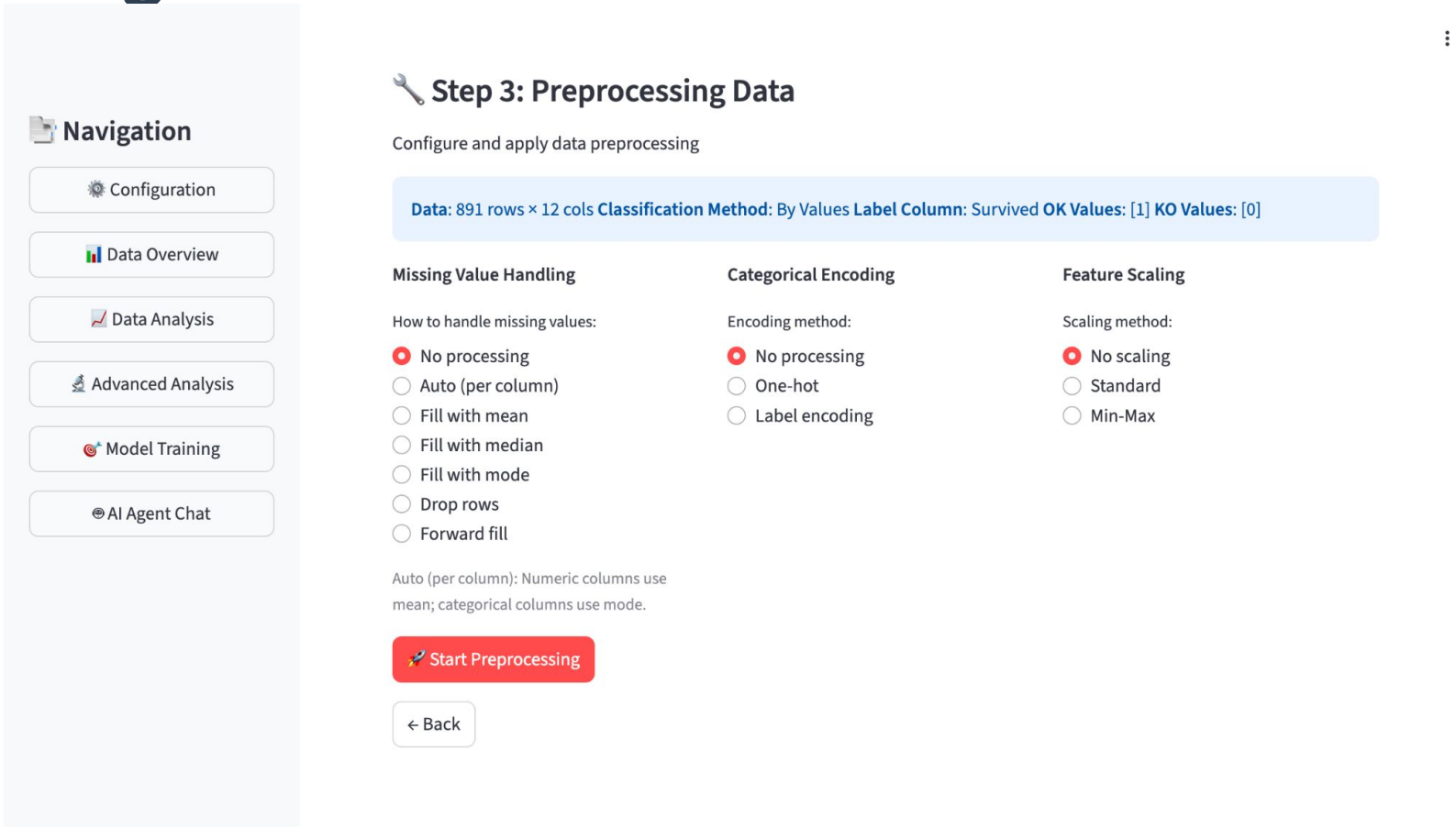
Configuration Interface



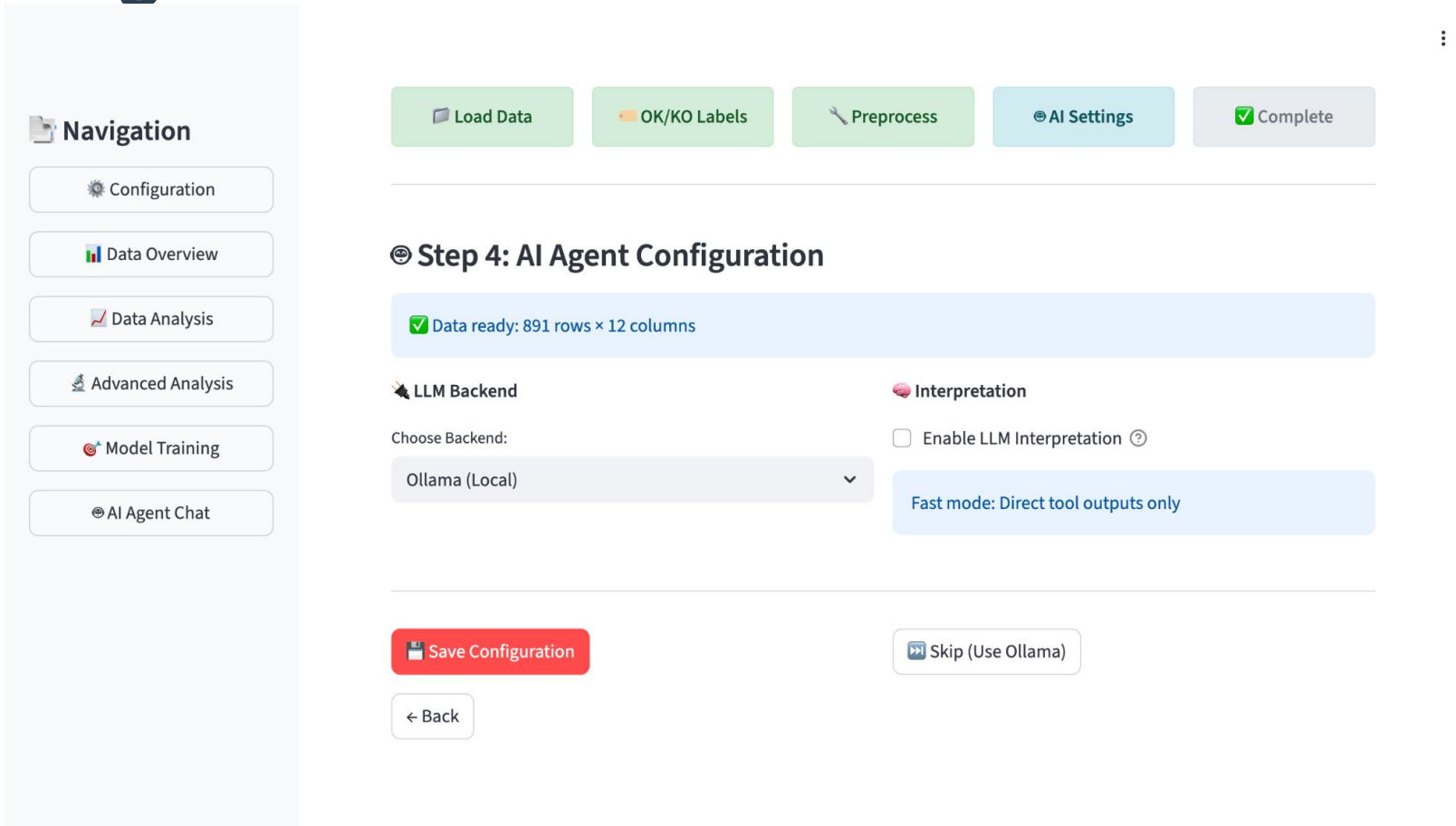
Configuration Interface



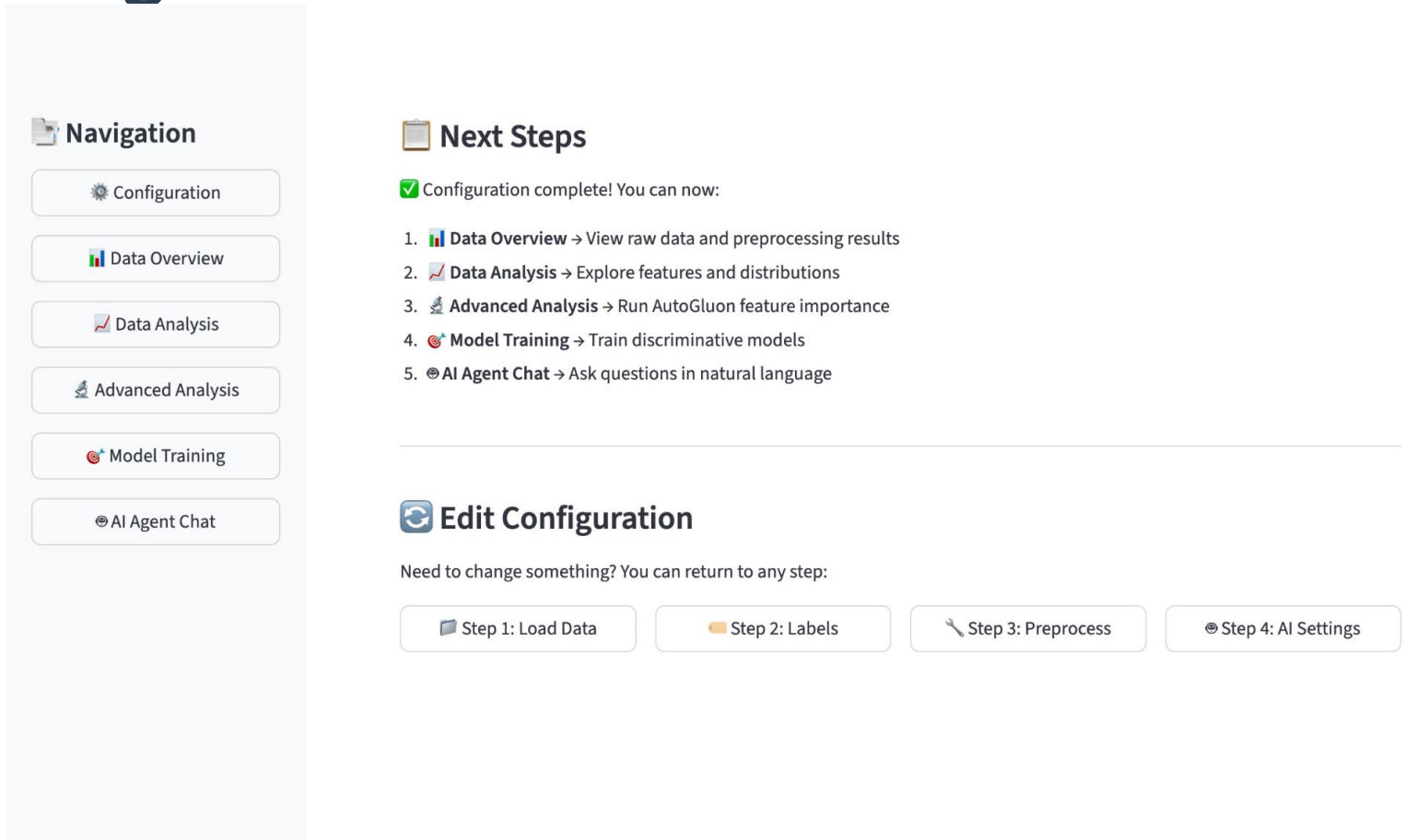
Configuration Interface



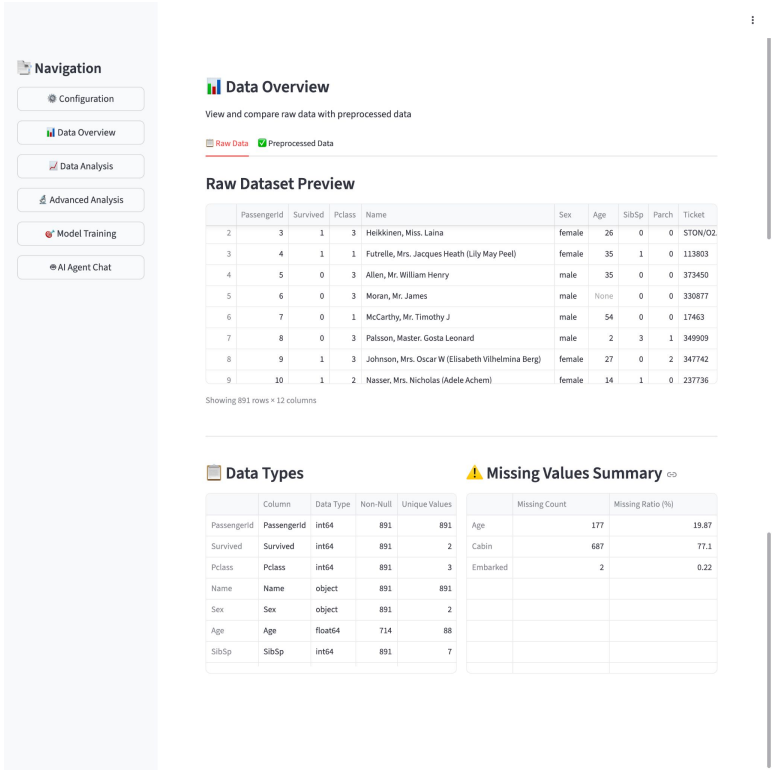
Configuration Interface

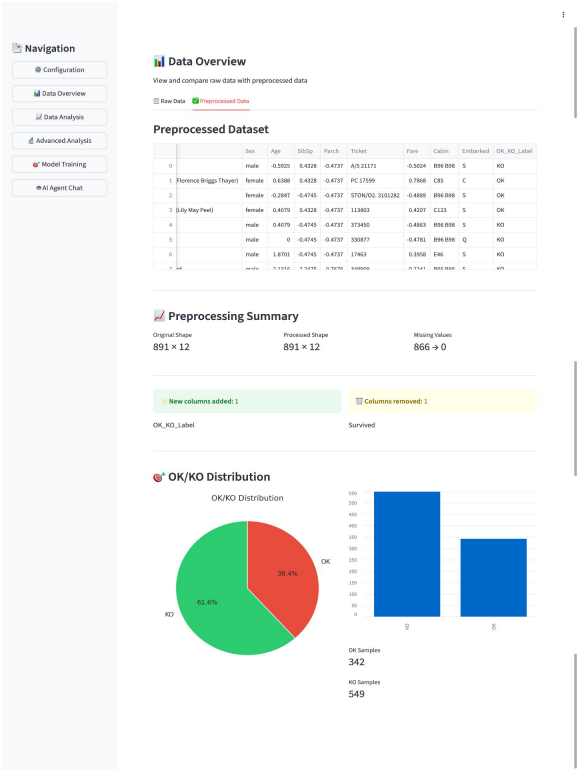


Configuration Interface



Data Overview





Dataset statistics and initial exploration

Data Analysis

Compare feature distributions between OK and KO groups

Dataset Summary & Feature Availability

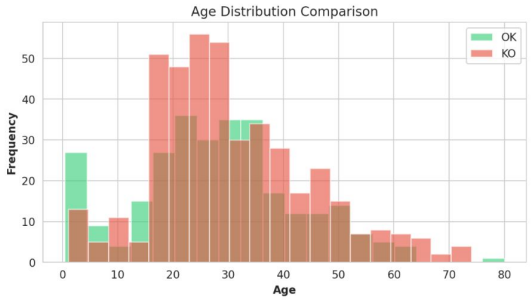
Numerical Features | Categorical Features

Select numerical features to analyze:

Age ×

Age - Statistical Comparison

	OK	KO
Mean	28.3437	30.6262
Std Dev	14.951	14.1721
Median	28	28



Dataset Summary & Feature Availability

Numerical Features | Categorical Features

Select categorical features to analyze:

Sex ×

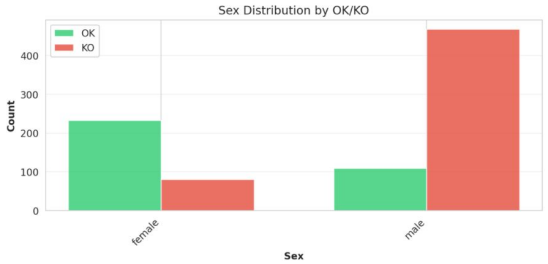
Sex - Category Distribution

Count Table:

Sex	KO	OK
female	81	233
male	468	109

Percentage Table:

Sex	KO	OK
female	14.75%	68.13%
male	85.25%	31.87%



Statistical Ranking

Numerical		Categorical	
p_{MW}	Mann-Whitney U test p-value (reliability)	**$p_{chi\text{-}square}$**	chi-square test p-value (reliability)
Cohen's d	effect size (strength of difference)	**Cramér's V**	effect size (association strength)
Difference Ratio	relative magnitude (normalized by max mean)		
$-\log_{10}(p_{MW})$ * Cohen's d * Difference Ratio		**$-\log_{10}(p_{chi\text{-}square})$ * Cramér's V**	

ML Ranking

AutoGluon Configuration & Logic

AutoGluon Tabular Predictor
Metric: Accuracy | Preset: Medium | Time Limit = 120s

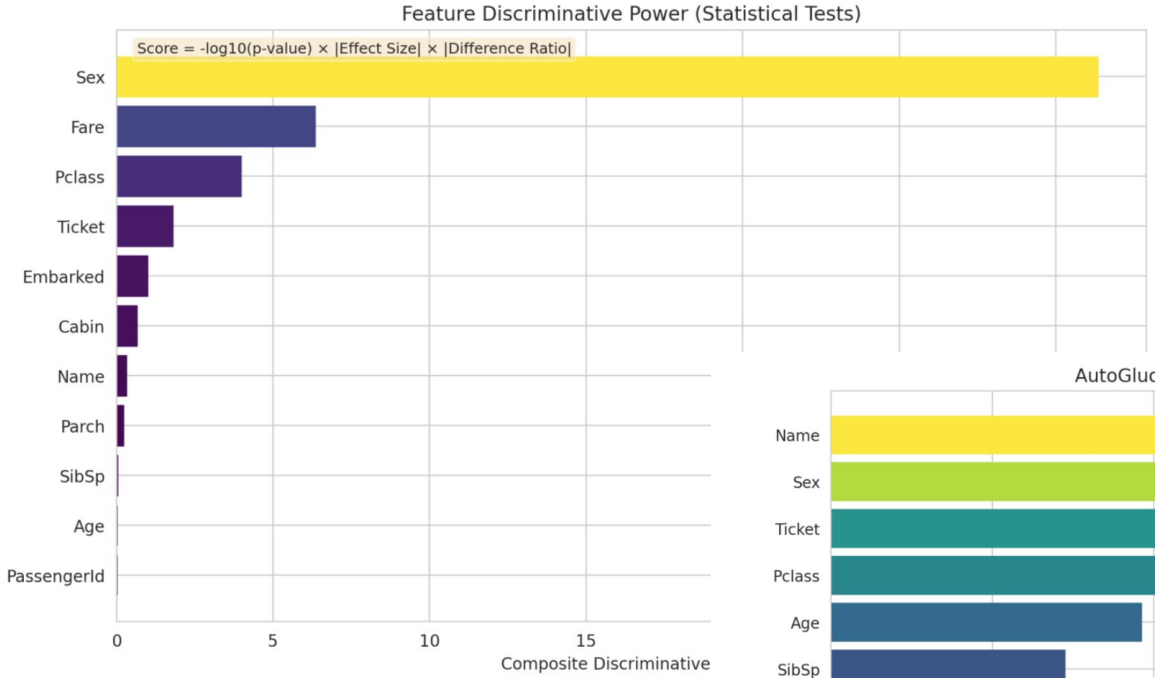
Data Splitting (Auto)
Train/Validation split by Auto Gluon

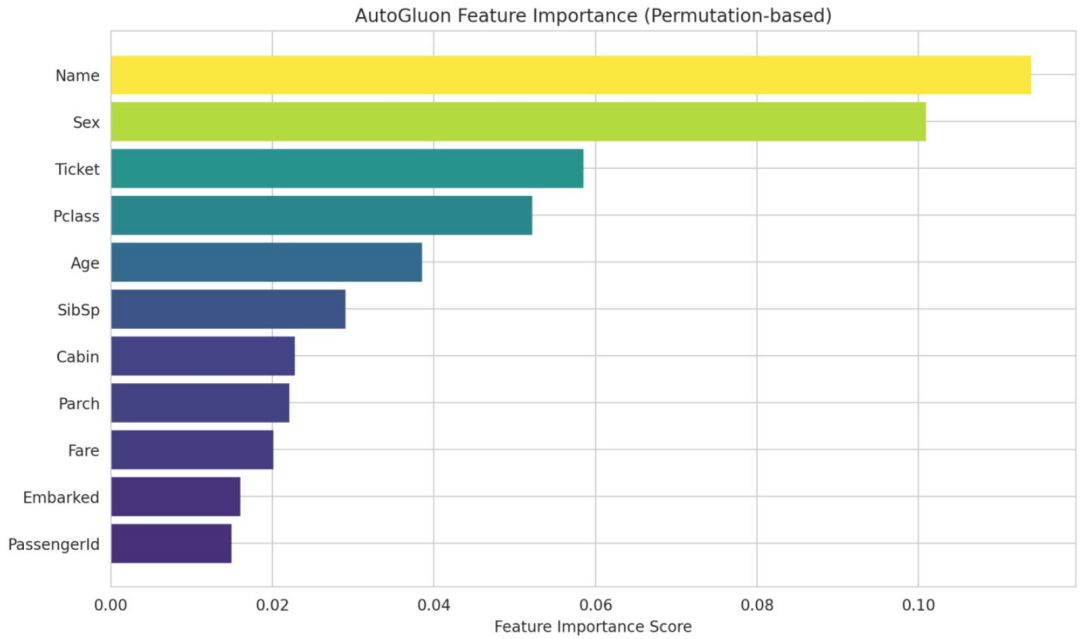
Ensemble Learning (Stacking)
Stack Level 2; base models + WeightedEnsemble_L2

Permutation Importance
Feature importance via shuffling

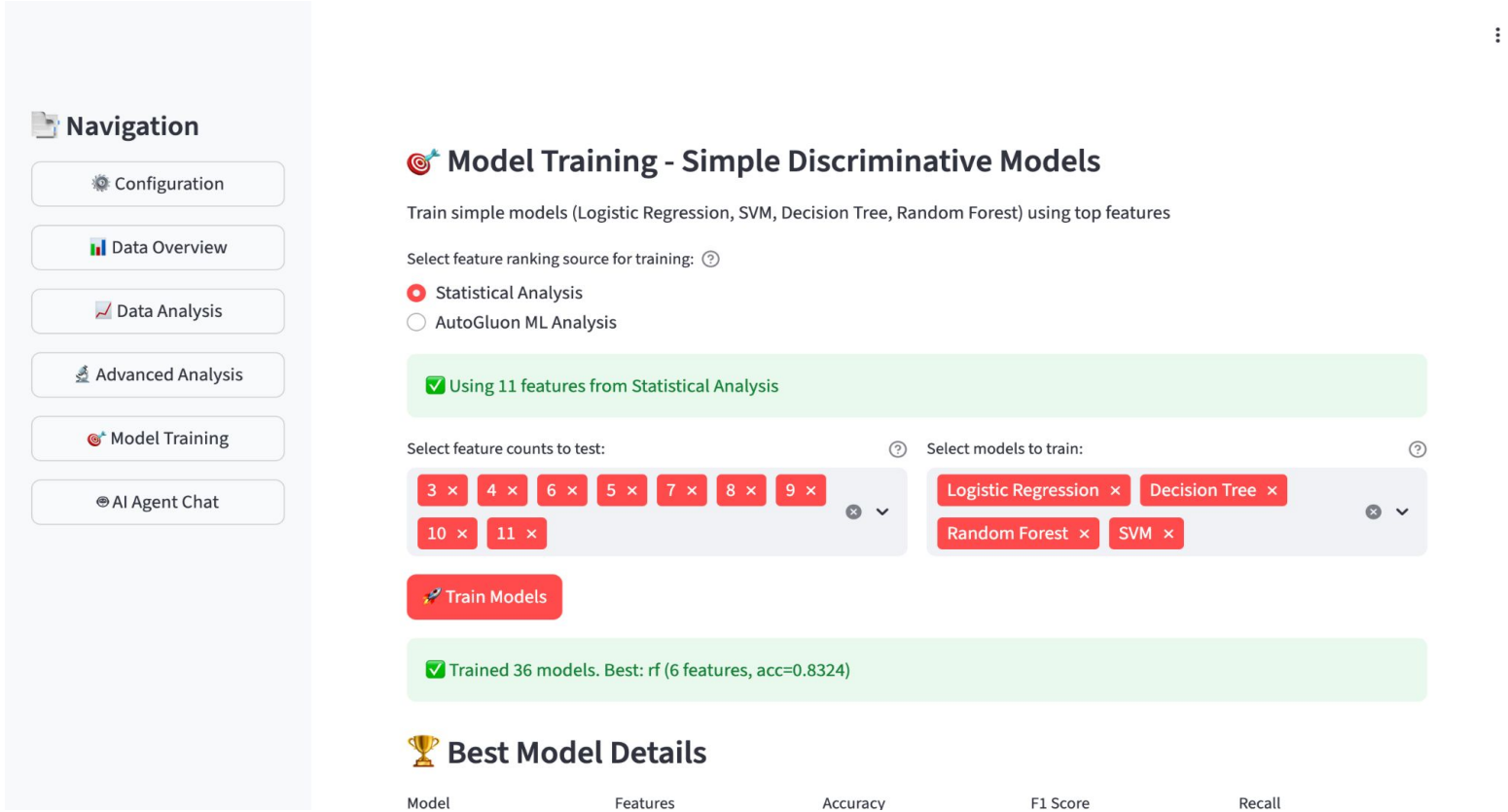
Training Outputs

Leaderboard	**Full model ranking**
Best Model	**WeightedEnsemble_L2**
Validation Accuracy	**0.8771**
Training Time	**≤120s total time**


Feature Discriminative Power (Statistical Tests)
Score = -log10(p-value) × |Effect Size| × |Difference Ratio|
Sex
Fare
Pclass
Ticket
Embarked
Cabin
Name
Parch
SibSp
Age
PassengerId
0
5
10
15
Composite Discriminative


AutoGluon Feature Importance (Permutation-based)
Name
Sex
Ticket
Pclass
Age
SibSp
Cabin
Parch
Fare
Embarked
PassengerId
0.00
0.02
0.04
0.06
0.08
0.10
Feature Importance Score

Model Training

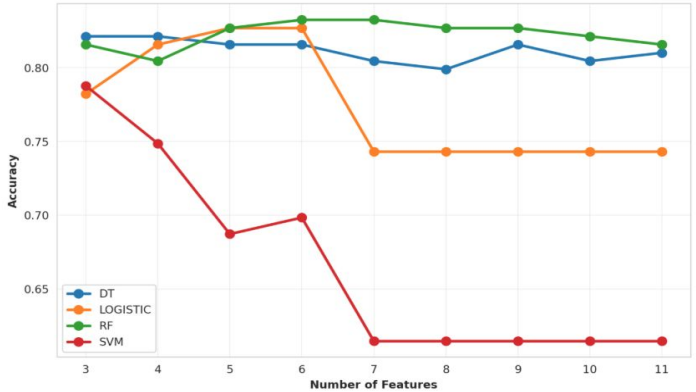


Model Training Results - Statistical

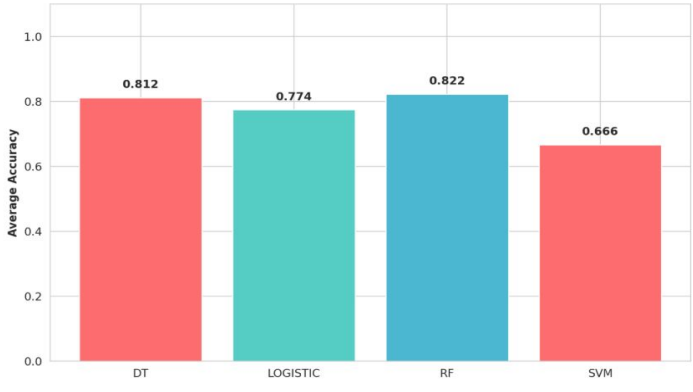
🏆 Best Model Details

Model	Features	Accuracy	F1 Score	Recall
RF	6/11	0.8324	0.7727	0.6892

📈 Accuracy vs Feature Count



🎯 Model Comparison



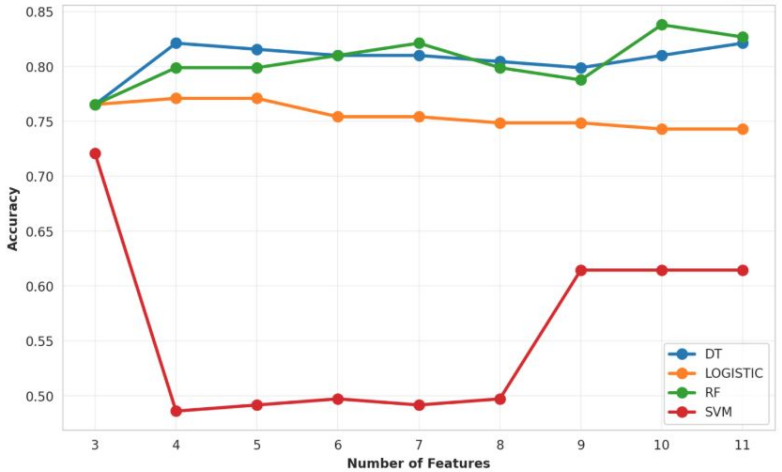
Model performance comparison: Different algorithms × Feature selection methods

Model Training Results - AutoGluon

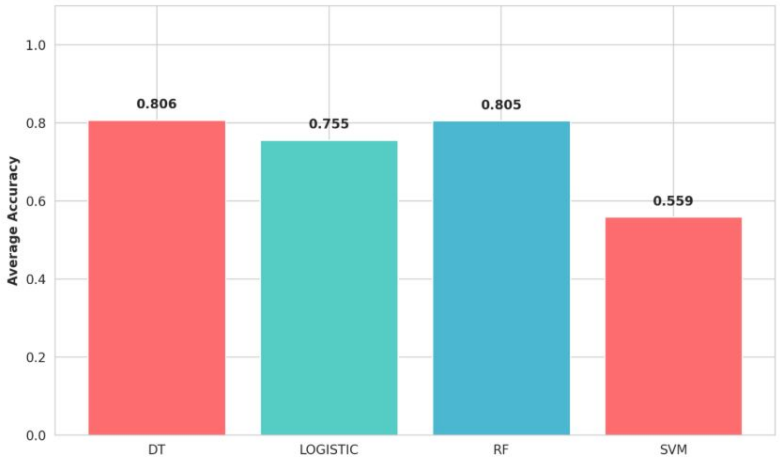
🏆 Best Model Details

Model	Features	Accuracy	F1 Score	Recall
RF	10/11	0.8380	0.7786	0.6892

📈 Accuracy vs Feature Count



🎯 Model Comparison



AI-Powered Chat Interface: Architecture

User Natural Language Query

↓

StatisticalAgent (Intent Parser)

Rule-based + LLM Fallback (LLaMA-3)

↓

Tool Routing

- get_feature_importance()
- get_statistical_summary()
- plot_distribution()
- plot_time_series()
- plot_frequency_spectrum()

Data Context

← Session State

- processed_df
- analysis_results
- feature_ranking
- statistical_summary

↓

Response Generation: Plot + Text Summary

LLM Backend Design

Backend	Deployment	Latency	Use Case
Ollama (LLaMA-3)	Local	2-5s	Privacy-first deployment
OpenAI GPT-4	Cloud	0.5-1s	High accuracy research
Claude / Gemini	Cloud	0.5-1s	Alternative cloud options
DeepSeek	Cloud	1-2s	Cost-effective option

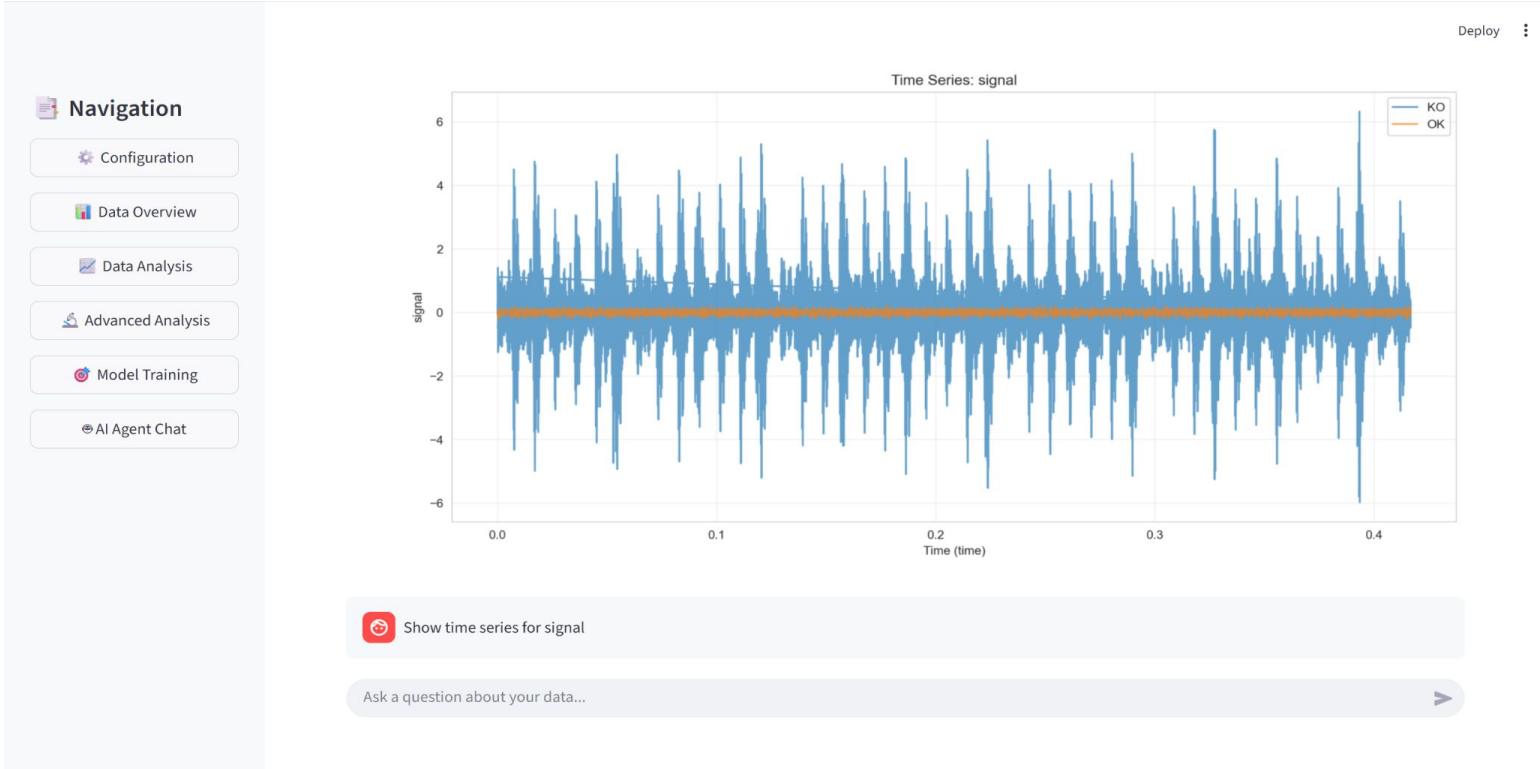
ConversationManager

Chat History
Maintains context

Context Window
Manages tokens

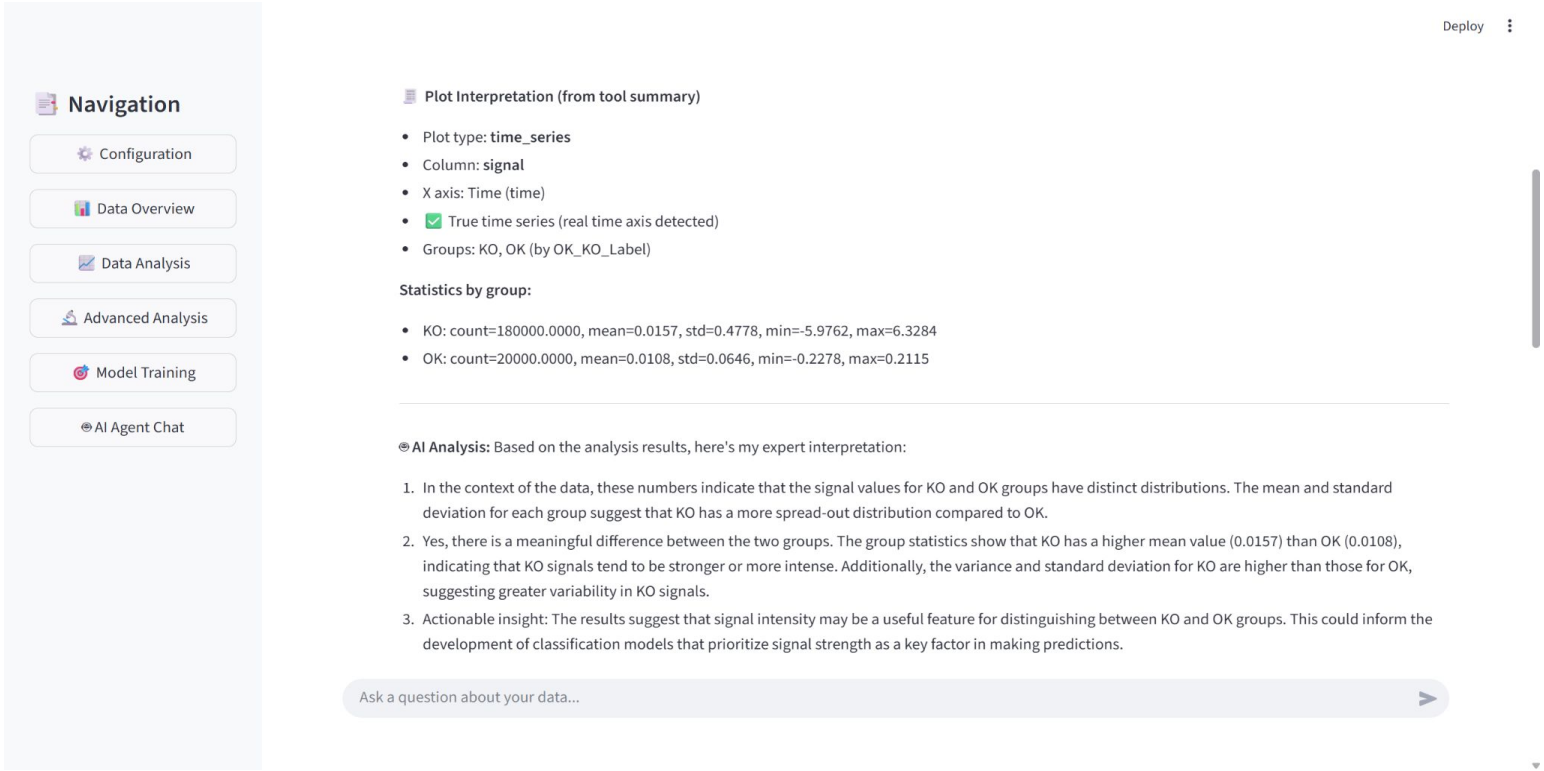
Session State
Links to analysis

AI Agent Use Case : Time Series Visualization



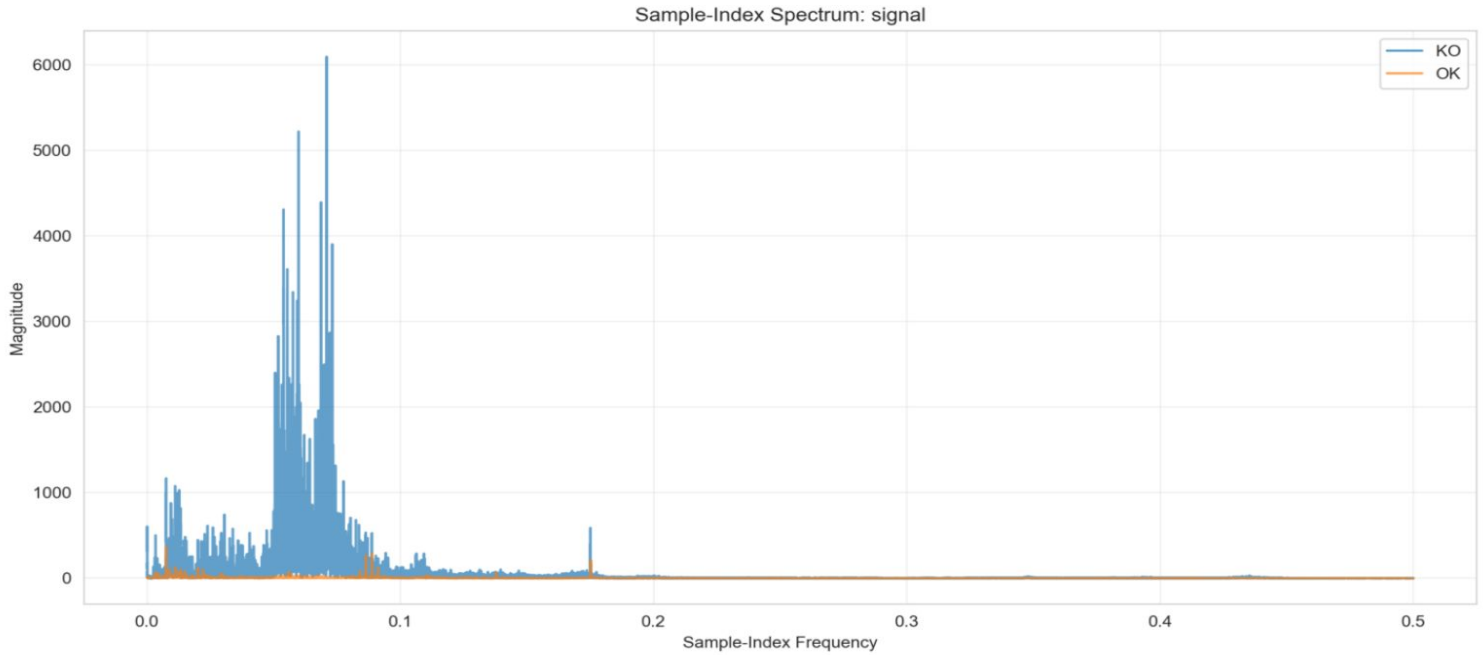
User query: "Compare distribution of rms between OK and KO"

AI Agent Use Case : Time Series Visualization



User query: "Show time series of signal"

AI Agent Use Case: Frequency Spectrum (FFT)



Plot FFT for signal

User query: "Show frequency spectrum" - FFT with peak detection

AI Agent Use Case: Frequency Spectrum (FFT)

✅ Generated frequency spectrum (FFT) plot for **signal**

Plot Interpretation (from tool summary)

- Plot type: **sample_index_spectrum**
- Column: **signal**
- ⚠️ Feature table data (NOT physical waveform)
- Groups: KO, OK (by OK_KO_Label)

Dominant peaks (sample-index):

- KO: 0.07, 0.07, 0.07
- OK: 0.01, 0.09, 0.09

Note: ⚠️ Feature table spectrum (NOT physical frequency). This shows patterns in sample order, not real Hz. Top-5 peaks by magnitude.

AI Analysis: Based on the analysis results:

1. The top-5 peaks by magnitude in the feature table spectrum indicate patterns in sample order, which may reflect differences in signal characteristics between KO and OK groups.
2. Since groups exist (KO, OK) and there are meaningful differences between them, this suggests that the signal features captured in the spectrum can be used to distinguish between KO and OK samples.
3. The actionable insight is that these signal features can potentially be used as input features for a classification model to predict whether a sample belongs to the KO or OK group.

User query: "What are the statistics for rms?" - Multi-metric view

AI Agent: Technical Highlights

Data Encapsulation

No DataFrame in prompts ⟶ prevents hallucination

Robust Time Handling

Auto-detects time columns with graceful fallback

Context-Aware

Receives full analysis results for comprehensive queries

Flexible Visualization

Natural language ⟶ plot type selection automatically

Structured Output Pattern

Plot + Summary + AI Interpretation ⟶ Deterministic computation (no AI guessing)

System Design Philosophy

1

Parallel Validation

Statistical + ML methods
run independently
→ user selects best

2

Shared State

One analysis
→ multiple consumers
No duplication

3

Anti-Hallucination

Tool-based execution
→ deterministic results
No AI guessing

Thank You

Key Contributions

✓ Parallel-validation architecture for feature ranking

✓ Multi-backend LLM integration with local deployment

✓ Context-aware AI agent with anti-hallucination design

✓ Interactive visualization toolkit for exploratory analysis

Lei GAO (s327756) • Lan DENG (s338219)