

Navigator MaaS Model Card [DRAFT]

Gretel Navigator

Model Overview

Gretel Navigator generates production-quality synthetic data optimized for AI and machine learning development from prompts, schema definitions, or seed examples. Unlike single-LLM approaches to data generation, Navigator employs a compound AI architecture specifically engineered for synthetic data, combining top open-source SLM models fine-tuned across 10+ industry domains. This purpose-built system creates diverse, domain-specific datasets at scales of hundreds to millions of examples while preserving complex statistical relationships and offering increased speed and accuracy compared to manual data creation.

Top use cases:

- Creating synthetic data for LLM training and fine-tuning
- Generating evaluation datasets for AI models and RAG systems
- Augmenting limited training data with diverse synthetic samples
- Creating realistic PII/PHI data for model testing

Documentation and Resources

- [Getting Started Guide](#)
- [SDK Documentation](#)
- [Example Notebooks](#)
- [API Reference](#)
- [Best Practices Guide](#)

Input Examples

Natural Language Prompts

```
1 Generate customer bank transaction data with the following columns:
2 - customer_name: Full names in Western format
3 - transaction_date: Dates within the last 30 days
4 - transaction_amount: Dollar amounts between $1-$10,000
5 - transaction_type: Either 'debit' or 'credit'
6 - transaction_category: Common banking categories like 'dining', 'retail', 'utilities'
7 - account_balance: Running balance after each transaction
```

Schema-Based Input

```
1 CREATE TABLE transactions (
2     customer_name VARCHAR(100),
3     customer_id CHAR(8),
4     transaction_date DATE,
5     transaction_amount DECIMAL(10,2),
6     transaction_type VARCHAR(6),
7     transaction_category VARCHAR(50),
8     account_balance DECIMAL(10,2)
9 );
```

Core Capabilities

Data Generation Architecture

- Agentic workflow system for synthetic data generation
- Multi-modal support (tabular, text, time-sequence)
- Scalable generation (up to millions of records)
- Underlying LLMs fine-tuned by Gretel on 10 different industry data and formats including healthcare, life sciences, financial, manufacturing, retail

Key Features

- Natural language interface to specify data requirements
- Schema-based data generation
- Real-time and streaming data generation
- Dataset augmentation and modification
- Structured data supported as LLM inputs and outputs

Example Open Datasets

High-quality open synthetic datasets created using Navigator available on HuggingFace:

- [Text-to-SQL Dataset](#): Large-scale dataset for SQL generation
- [GSM8K Math Problem Solving Dataset](#): AI reasoning dataset
- [Multilingual Financial PII Dataset](#): Financial services training data

Coming Soon: AI Data Designer

The AI Data Designer functionality, coming soon to Model-as-a-Service (MaaS), will provide an end-to-end synthetic data pipeline using Navigator, enabling iterative improvements, validation, and evaluation of generated datasets directly through the Navigator SDK.

Service Limitations

- Fine-tuning capability is not yet available in MaaS. See [Navigator Fine-tuning](#) in Gretel Cloud.
- [Batch SDK](#) for large-scale generation currently limited to Gretel Cloud

Support and Resources

- Technical Documentation: [Gretel Navigator | Gretel.ai](#)
- SDK Documentation: [Gretel Navigator Inference API | Gretel.ai](#)
- Community Support: [Join the Synthetic Data Community Discord Server!](#)
- Email Support: support@gretel.ai

Licensing and Data Ownership

- Llama 3.1 Community Licensed; Model suite supporting Apache 2.0 under the Open Data Commons (ODC-BY) license coming soon to Model-as-a-Service

Model Governance

Navigator is designed to democratize synthetic data generation while upholding high standards of responsible AI development. The system incorporates automated alignment checks to detect the generation of harmful or discriminatory data while respecting legitimate use cases across industries.

Navigator is trained exclusively on high-quality, license-compliant datasets spanning 10+ sectors, ensuring both legal compliance and output quality. However, like any advanced AI system, Navigator may occasionally produce unexpected or biased outputs. We therefore recommend that users conduct appropriate testing and validation for their specific use cases.

Gretel's governance framework includes privacy-preserving architecture, regular security audits, and continuous monitoring for bias and quality control. Through ongoing model updates and strict access controls, we maintain alignment with responsible AI principles while protecting against potential misuse. Users are encouraged to review our [Responsible Use Guidelines](#) and implement appropriate safety measures based on their specific applications and industry requirements.