# **Data Transformer** Mode-specific **Flight Input Data** Normalization Airlines, airports, For delays, durations, etc. aircraft types, schedules, delays **Column Tokenization** Airlines, airports, aircraft Regularization Target Masking Prevents data copying

## **GPT-2 Autoregressive Model**

### Model Architecture

6 transformer decoder layers 768-dimensional embeddings

#### **Row-wise Autoregressiv Generation**

Token-by-token autoregressive

# Invers Data Transformer

Converts tokens to flight data

# Synthetic Flight Data

Realistic tabular flight information

## **Overfitting Detection**

 $Q\delta$  statistic and DCR

Statistical bootstrapping