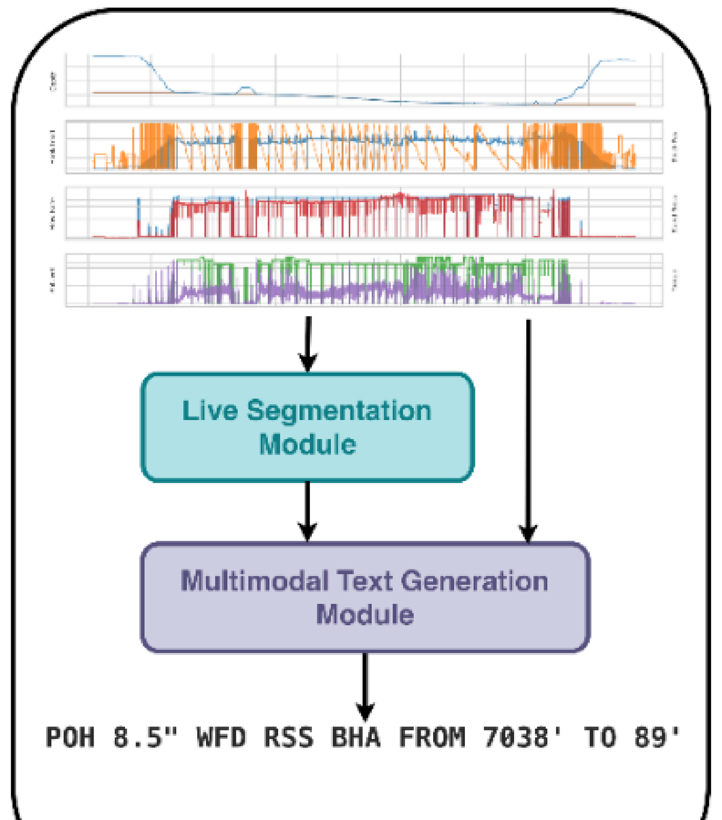
LiveDrill: Multimodal Segment-Triggered Data-to-Text for Time Series Foundation Models

Soumyadipta Sengupta*, Amine El Khair*, Sebastiaan Buiting*, Imane Khaouja, Yahia Salaheldin Shaaban, Abdallah Benzine*
AIQ

NEURAL INFORMATION PROCESSING SYSTEMS 2025

Introduction & Methodology



a) Complete LiveDrill Pipeline

Research Question:

Can Time-Series Foundation Models (TSFMs) be used to automate live industrial reporting by coupling real-time activity segmentation with segment-grounded multimodal data-to-text generation?

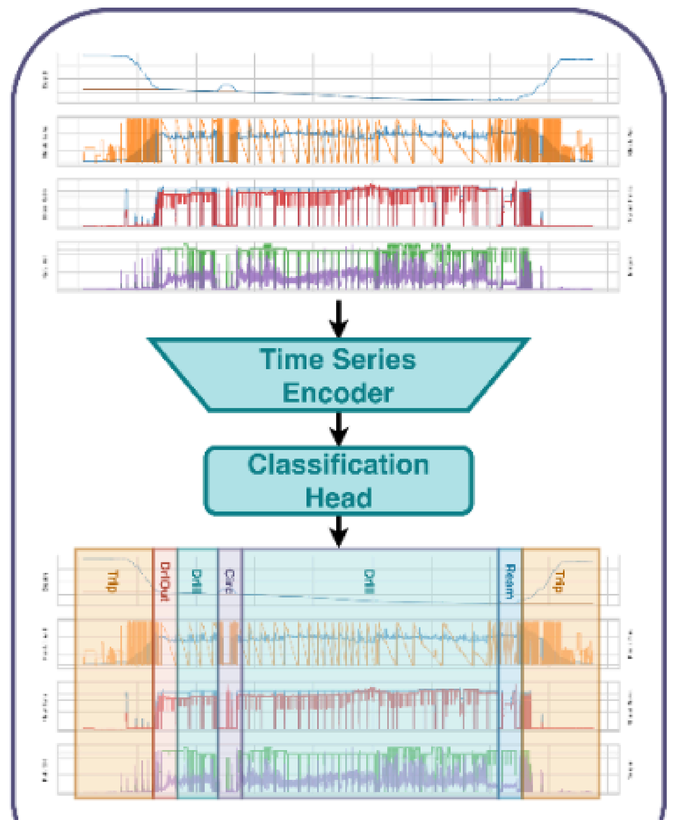
Methodology:

LiveDrill: A Streaming Pipeline for Real-Time Reporting

LiveDrill transforms continuous **multivariate sensor** streams into automated **Daily Drilling Reports (DDRs)** in **real-time**. It introduces a segment-triggered architecture that decouples activity detection from text generation.

This design shifts reporting from delayed, manual summaries to continuous, segment-grounded narratives, ensuring updates are timely and explicitly tied to operational events.

LSM	TSE: Moment-large			TSE: Moirai-large		
	Seg $F1_{IoU}$	Avg LLM	Combined	Seg $F1_{IoU}$	Avg LLM	Combined
Moment-large	0.467	0.359	0.405	0.467	0.340	0.394
Moirai-large	0.484	0.359	0.412	0.484	0.340	0.399

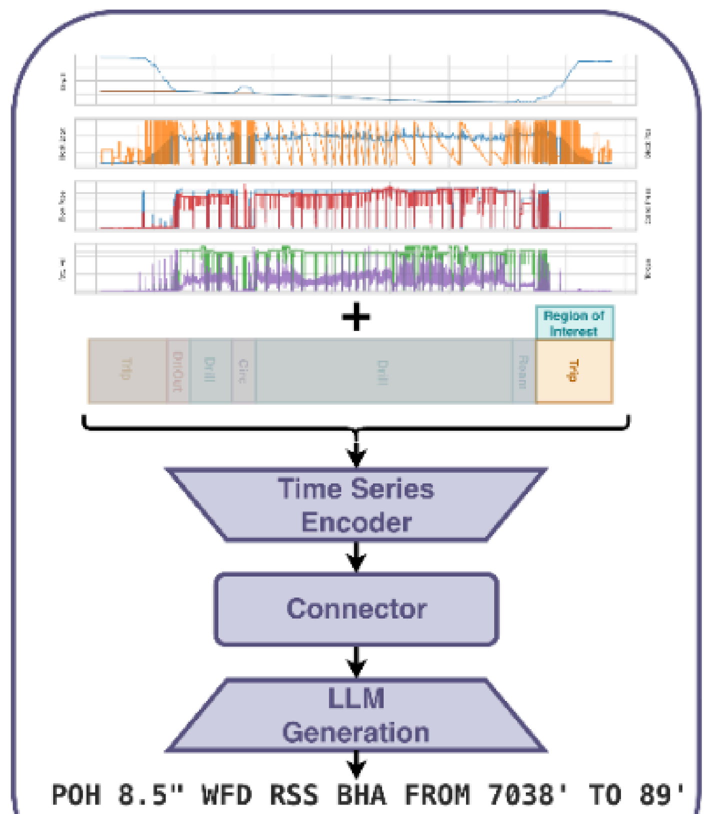


b) Live Segmentation Module

Live Segmentation Module (LSM):
This module comprises a **pretrained time-series foundation model encoder (Moment or Moirai)** followed by a **dense classification head** to assign per-step **activity** labels in **real-time**.

Segmentation Model	Overall	
	$F1_{IoU}$	$F1_{pw}$
Moment-large	0.467	0.827
Moirai-large	0.484	0.809

Fast and Efficient: It is highly optimized for speed (~192ms latency), easily handling messy, high-frequency data streams without lagging behind live operations.
Strong Generalization: It generalizes well across different drilling phases, achieving high point-wise accuracy and qualitatively strong alignment with ground truth despite signal noise.
Small Segment Sensitivity: It struggles with very short or fragmented segments, where minor signal noise can cause the model to split a single activity into multiple pieces.
Rare Class Detection: Performance drops for infrequent or complex anomaly classes (like Stuck Pipe), which are harder to capture than stable operations.



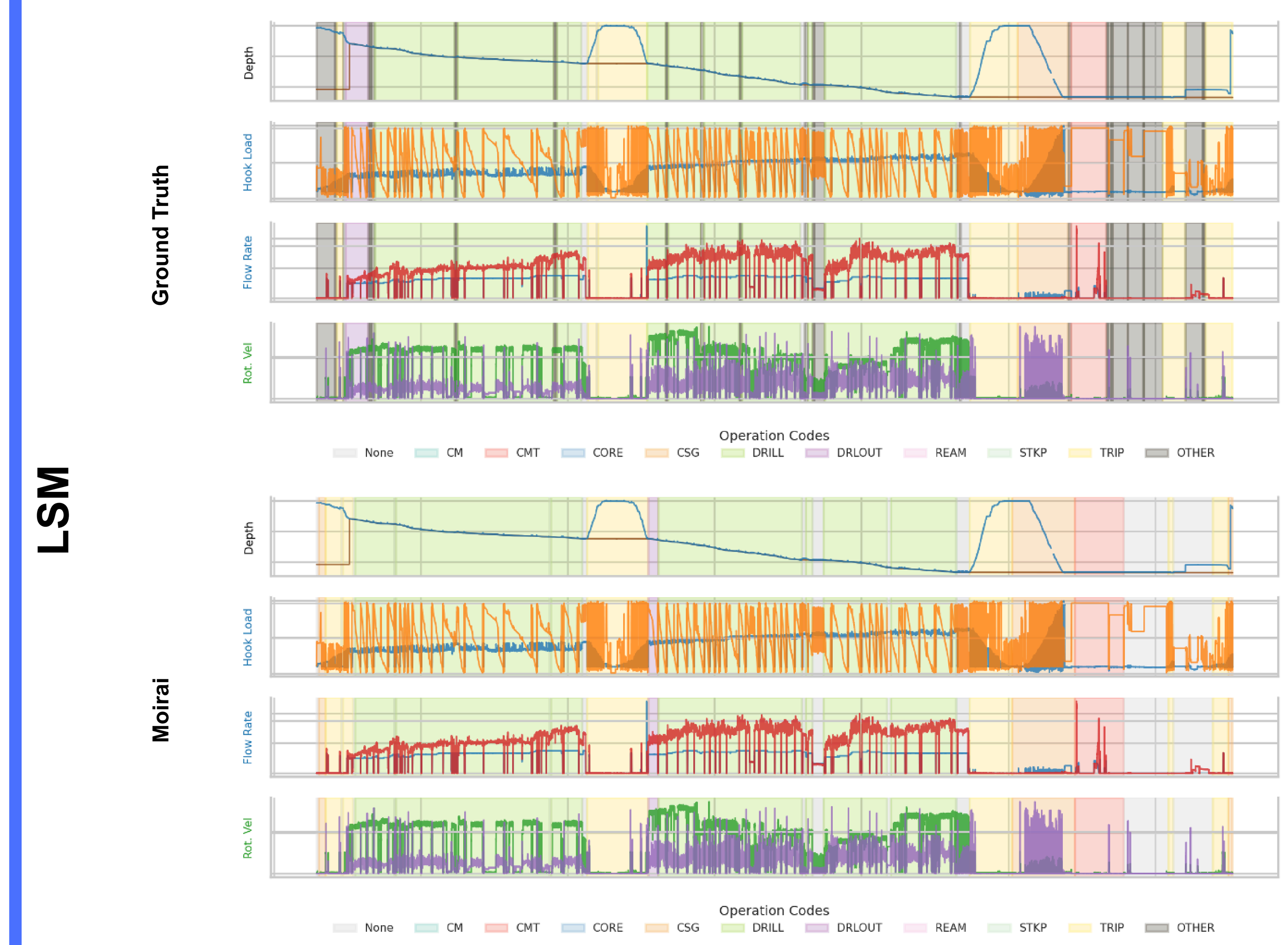
c) Multimodal Text Generation Module

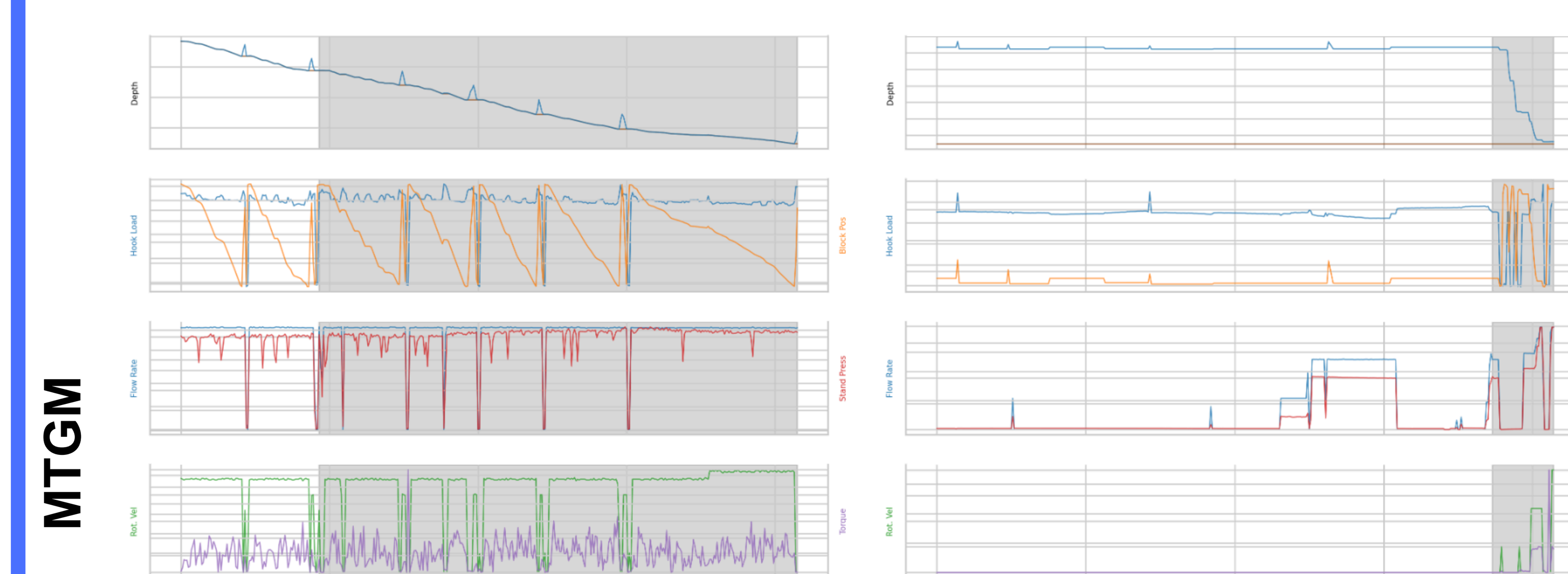
Multimodal Text Generation Module (MTGM):
An **event-triggered generator** that fuses sensor data with a Region-of-Interest (ROI) mask via a **pretrained time-series foundation model encoder** and **connector** to **soft-prompt** a **frozen LLM** for text creation .

Qualitatively Grounded: It produces coherent, segment-grounded descriptions that accurately reflect the sensor dynamics, proving effective even when applied to messy field data.
Model Agnostic: The system is robust regardless of the backbone used; swapping between Foundation Models (Moment vs. Moirai) results in negligible performance differences.
Dependency on Segmentation: It is sensitive to input quality; if the LSM detects a "very small" or fragmented segment, the MTGM lacks the context to generate a meaningful report.
Transition Complexity: While good at stable phases, it struggles to generate accurate narratives for complex, infrequent transition events where data patterns are less distinct.

TSE Model	Avg LLM [0-1]	CM	CMT	CORE	CSG	DRILL	DRLOUT	REAM	STKP	TRIP
Moment-large	0.359	0.137	0.333	0.291	0.353	0.590	0.456	0.499	0.114	0.200
Moirai-large	0.340	0.212	0.344	0.309	0.354	0.563	0.368	0.454	0.210	0.157

Qualitative Results:





Reference: cont drilling 8.5" deviated hole w/ rss bha f 8815' to 9303' [r-2]

Prediction: cont rss drilling 8-3/4" hole with hlb rss bha from 8,300' to 9,200' (1000' incline). - full wash rop, 15-20klbs.m

Reference: rih halliburton 12-1/4" rss (icruise) bha from 6,213 ft to 6,401 ft. - washed down last stand at 700 - 950 gpm / 2,100 - 3,500 psi.

Prediction: rih 12 1/2" hole with 12.5" motor head from 6,250 ft to 6,350 ft. - drill 12.5" hole with 12.5" motor head from 6,25

Future Work:

- **Pretrain sensor–text foundation models** directly on aligned pairs (not just forecasting) to improve zero-shot transition detection.
- Investigate **differentiable connectors** that **allow text generation loss** to backpropagate and **refine segmentation boundaries**, minimizing error propagation.

📱 **Scan QR code** for the full paper.

LiveDrill LLMaaJ