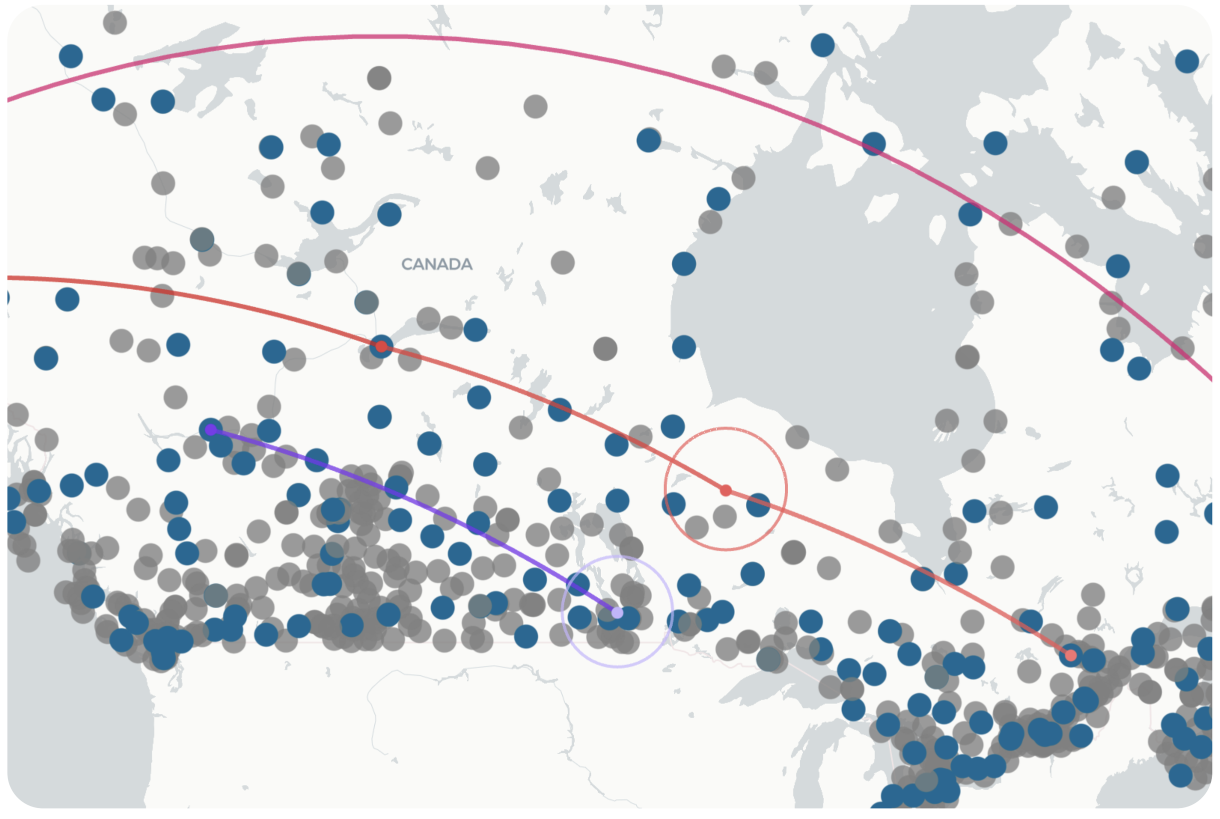
AVIATION AI RESEARCH

AeroFormer — Real-Time Weather Intelligence for Pilots

We pair a route-semantic weather matrix with dual LLM engines—one translator, one executive co-pilot—to collapse the time between encoded bulletins and actionable flight decisions.



RAW AVIATION INPUTS

Isolated data from different aerodromes.

METAR
CYYZ 131900Z 07010KT 15SM SCT045 OVC055 18/10 A3014 RMK SC3SC5 SLP208 DENSITY ALT 800FT

TAF
CYYZ 131743Z
1318/1424 08008KT P6SM SCT015 BKN030
TEMPO 1318/1323 OVC015
FM132300 09008KT P6SM FEW015 SCT030
...

SIGMET
WSSP32 LEMM 210741 LECB SIGMET 5 VALID 210740/211100 LEVA- LECB BARCELONA FIR/UIR EMBD TS OBS AT 0740Z
...

NARRATIVE OUTPUTS

Based on customized flight-path.

Go/No-Go Summary

> "
> *This is a 'Go', but it's high effort: misty departure, turbulent ridge crossing, and a gusty crosswind landing in thin air. If you're not sharp today, stay on the ground.*

Route-Based Narrative

> "
> *Picture the takeoff. It's a typical marine-layer morning with only 4 knots on the surface, 5 miles visibility, and scattered clouds at 900 feet. Launch decisively, snake through the gaps, and expect brilliant sunshine above 2,000 feet.*

Dongxu Liu · Debarghya Barik · Praytay Mondal

Supervisor · Andreas Lehrmann

LLM TRANSLATOR

Route-native narratives

Turns weather grids into conversational explanations.

EXECUTIVE SUMMARY ENGINE

BLUF co-pilot

Prioritizes go/no-go + regulatory checks sourced from the narrative stream.

COVERAGE

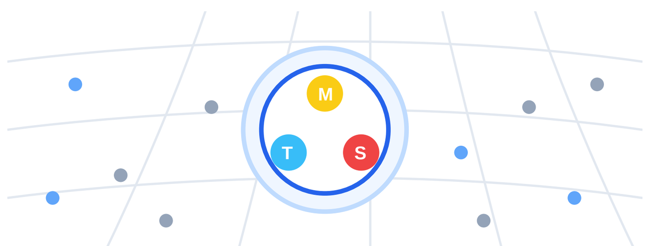
210 aerodromes from Canada

METAR · TAF · Winds Aloft · SIGMET with intelligent interpolation.

OPERATIONAL FLOW

Route-aware intelligence pipeline.

Data Fusion Layer



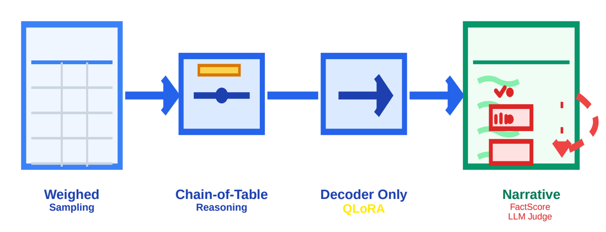
Real-Time Multi-source Data Acquisition

Route Semantic Matrix

SEMANTIC ROWS	CYTZ	ENR1	ENR2	CYOW
Winds	Calm	100 25kt	250 30kt	VAR 20 G35kt
Visibility	10 SM	20 SM	3/4 SM	1/4 SM
Warnings	-	ICE	TURB	TURB
Rules	PASS	CHECK	CHECK	FAIL

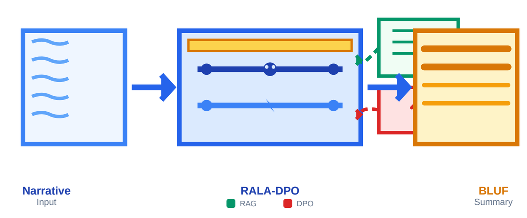
Flight Route-Aware Contextual Alignment

Narrative Decoding Engine



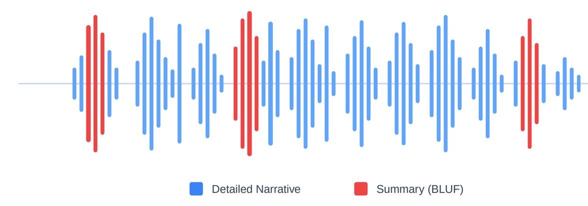
Qwen 3 + QLoRA: High-fidelity tabular-to-text with FactScore validation

Executive Summary Engine



Qwen 3 + RALA-DPO: Thinking Mode with external regulations(RAG).

Delivery & TTS



Natural-sounding audio for hands-free consumption

SYSTEM ARCHITECTURE

Data preprocessing and intelligence pipeline.

DATA

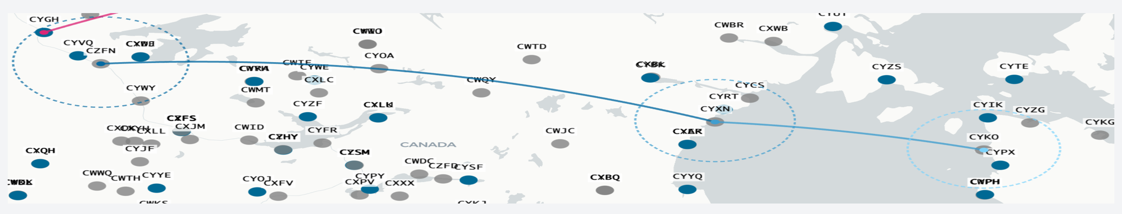
Real-Time Data Acquisition

- Aggregates live METAR, TAF, Winds Aloft, and SIGMET data from 210 Canadian aerodromes.
- Fetched every 30 minutes, stored 70,000+ METARs, 23,000+ TAFs, 15,000+ Winds Aloft records.

Route-Based Data Alignment

- Captures current/newest METARs from departure Aerodrome.
- Fetches Available Forecasted TAFs and Winds Aloft for en-route and destination Aerodromes.
- Checks SIGMETs All Over the Route.
- Instantly interpretation using deterministic rules, e.g. Pressure Gap Awareness, VFR Check before landing...

Gap Filling & Interpolation

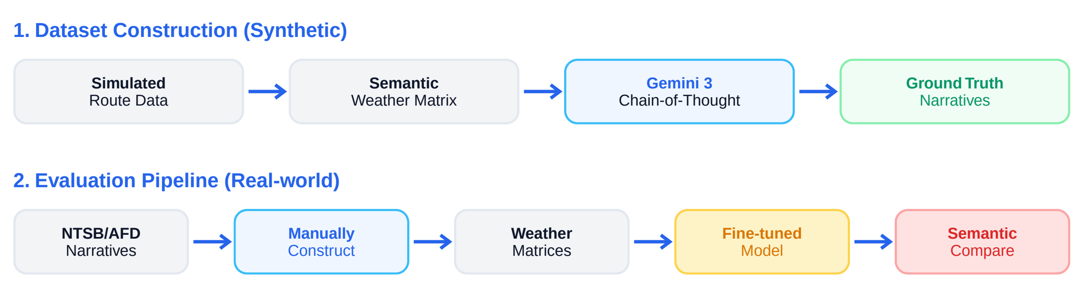


Spatial interpolation techniques fill sensor gaps by leveraging nearby station data. Generates weather estimates for aerodromes lacking direct sensors.

Data Type	Interpolation Method
Sea Level Pressure, Wind Vectors	Barnes Objective Analysis
Surface Temperature, Altimeter (Complex Terrain)	GIDS (Gradient + Inverse Distance Squared)
Surface Temperature (Flat Terrain)	IDW (p=2)
Ceiling, Visibility	Nearest Neighbor (Conservative)

Zero-Shot Ground Truth

Utilizes Gemini 3 to generate high-quality reference data (ground truth) for model training without manual labeling. Validated fine-tuned model with NTSB/ASRS accident narratives and NWS-AFD-AVIATION weather reports.



- NWS-AFD dataset contains both detailed discussions and summaries for aviation, suitable for both matrix generation and summarization fine-tuning.
- The weather matrices constructed from NTSB accident narratives are critical for validation and therefore require careful verification.

NARRATIVE ENGINE

Convert weather matrices to detailed plain English narratives.

DATA ENGINEERING

Weighted Sampling

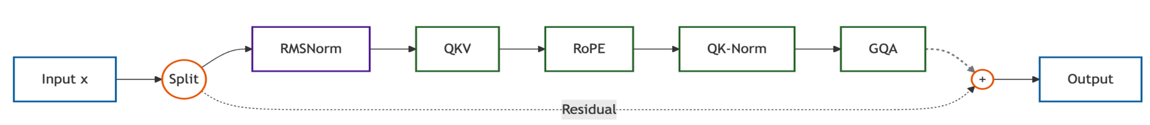
- Ensure the selection of simulated routes is balanced between high-risk weather conditions (such as icing and turbulence) and normal weather conditions.
- Using NTSB accident reports to generate realistic weather conditions, serving as critical supplementary samples that enrich the existing training and validation datasets.

Targeted Oversampling

Aggressively oversamples high-risk examples (thunderstorms, IFR conditions) in every epoch, forcing the model to master difficult patterns.

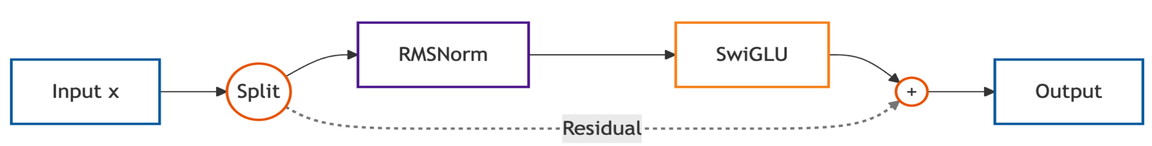
MODEL QWEN3

Attention Mechanism



- RoPE: Encodes position via vector rotation to master long-context understanding.
- QK-Norm: Normalizes attention signals to ensure stability during large-scale training.
- GQA: Groups attention heads to significantly boost inference speed and efficiency.

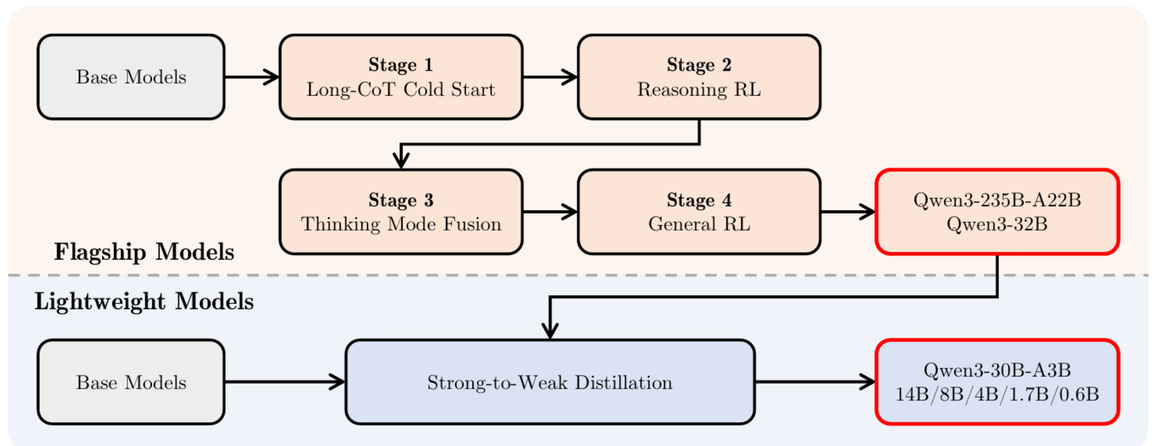
Feed-Forward Network



- Normalizes feature magnitude without re-centering, reducing computational overhead while maintaining stability.
- Split-Add: Divides feature streams to enhance gradient flow and prevent signal degradation.
- SwiGLU: Uses a gating mechanism to selectively filter information for richer expressivity.

MODEL QWEN3

Post Training



- Integrates Thinking and Non-Thinking modes, allowing users to toggle reasoning and control depth.
- Direct logit distillation enhances student models without sacrificing control over reasoning processes.

QLORA

Low Rank Adaptation

LoRA is a technique used to fine-tune massive Large Language Models (LLMs) without needing to retrain every single parameter. It freezes the backbone parameters and trains low-rank adapters.

$$\underbrace{\begin{bmatrix} \cdot & & \cdot \\ & \Delta W & \\ \cdot & & \cdot \end{bmatrix}}_{\text{High VRAM (Dense)}} \approx \underbrace{\begin{bmatrix} \cdot \\ B \\ \cdot \end{bmatrix}}_{\text{Tall}} \times \underbrace{\begin{bmatrix} \cdot & A & \cdot \end{bmatrix}}_{\text{Wide}}$$

- Assumes that the change in weights (delta W) has a low intrinsic rank, can be broken down into the product of two much smaller matrices, A and B.
- Freezes the backbone parameters and trains low-rank adapters.

Quantized LoRA

Integrates LoRA with aggressive quantization to make fine-tuning massive models feasible on consumer-grade hardware.

SUMMARY ENGINE

Generate executive briefings and Go/No-Go suggestions.

MODEL QWEN3

Thinking Mode

Qwen3 integrates dual reasoning modes within a unified framework. In thinking mode, the model allocates a "thinking budget" to reason step-by-step. For weather summaries, this mode triggers on complex queries like "Evaluate departure feasibility under icing conditions."

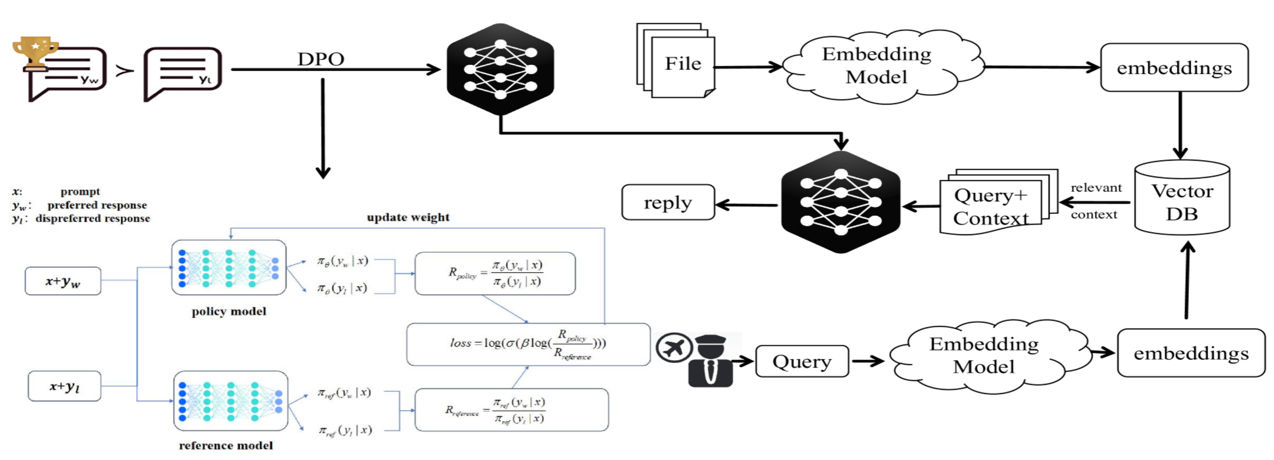
Context Window

Qwen3 8B natively supports 32,768 tokens context length. For most single-flight briefings, the native 32k window is sufficient and preferred for maximum fidelity.

Aviation Knowledge Graph

Integrated RDF/OWL-based KG structures relationships between weather phenomena and aviation impacts. When encountering terms like "Squall Line," RAG retrieves not just definitions but connected KG nodes (e.g., "Wind Shear," "Microburst," "Hail") via graph database queries alongside vector retrieval.

FINETUNE WITH RALA-DPO



We fine-tune Qwen3 using Direct Preference Optimization (DPO) to enhance the model's ability to generate more professional and reliable responses. Additionally, we integrate Retrieval-Augmented Generation (RAG), leveraging a knowledge graph to ensure that the model's outputs are both accurate and trustworthy.

Hallucination Detection and Guardrails

To ensure reliability and safety, we deploy multiple layers of validation to detect hallucinations and verify factual accuracy. These guardrails work together to ensure the model's outputs are trustworthy and aligned with source data.

- Self-Consistency Check: The 'Thinking Mode' prompt includes a self-verification step where the model compares its generated summary against the input data.
- NeMo Guardrails: We deploy NVIDIA NeMo Guardrails as a deterministic fact-checker. This layer runs Python scripts to verify numerical claims (e.g., if the summary says 'Visibility 1/2 SM,' the script regex-matches the METAR to confirm '1/2SM' exists).
- Fact Score: Deconstructs generated text into atomic facts to calculate Factual Precision (no hallucinations) and Factual Recall (no missing key data).
- Superior Model Grading: We use advanced models (e.g., GPT-4o, Gemini 3) to grade the 8B model's summaries on specific dimensions: Faithfulness, Completeness, and Safety. Research shows this correlates better with human expert evaluation.

Project Progress

Current development status across all system components.

Literature review	DONE
Data Acquisition	DONE
Route-Based Data Alignment	DONE
Gap Filling	IN PROGRESS
NARRATIVE ENGINE	IN PROGRESS
SUMMARY ENGINE	IN PROGRESS

Impact

By delivering these simplified insights during pre-flight briefings, real-time en-route updates, and landing checks, AeroFormer significantly accelerates the "Go/No-Go" decision process. This enables General Aviation (GA) pilots to effortlessly grasp critical information, thereby substantially elevating overall flight safety standards.

Authors

Dongxu Liu · Debarghya Barik · Pratyay Mondal
Supervised by Andreas Lehrmann · THWS · Würzburg