AUTONOMOUS MULTIMODAL AIR TRAFFIC CONTROL SYSTEM WITH ROLE-CONDITIONED DECISIONING AND REAL-TIME RISK ESCALATION

Christian Strommen

4 Sept 2025

Abstract

A unified, multimodally trained decision system is proposed to assume the functions of human air traffic controllers across tower, ground, and terminal radar approach control (TRACON) operations, with subsequent extension to en route centers. The system ingests (i) speech-to-text transcripts from a dedicated automatic speech recognition (ASR) model; (ii) surveillance streams including ADS-B, Mode S, multilateration (MLAT), and ADS-C; (iii) airport procedures and diagrams; (iv) meteorological products including METAR/TAF and local AWOS/ATIS; and (v) filed/active flight plans. A single multimodal model is conditioned at runtime with a role token to execute a specific job (e.g., Ground, Tower, Arrival, Departure) and to coordinate with peer role instances. The design directly addresses current constraints faced by regulators and service providers: controller fatigue, staffing shortages, and hiring/training limitations, while respecting the operational separation and handoff boundaries used today. Safety is assured through rule-aware decoding, monitors for separation minima and conformance, and a calibrated risk-to-human escalation mechanism. Initial validation is performed in high-fidelity simulation environments and then as a shadow "parallel controller" in live operations, with latency targets at or below human controllers. A staged deployment path is described, beginning with TRACON roles, followed by ground, and finally tower. A quantitative analysis indicates substantial direct cost exposures in salaries/training and large indirect costs from delays and inefficiencies; even modest delay reductions produce system-level savings in the hundreds of millions annually in major regions. Industry datasets such as ATCO2 and ATCOSIM are referenced for baseline tasks, but expanded multilingual, accent-diverse, and facility-diverse corpora are required. The intended end state is a system that exceeds human performance targeted at an order-of-magnitude improvement in safety and efficiency while integrating with existing automation platforms and procedures.

Table of Contents

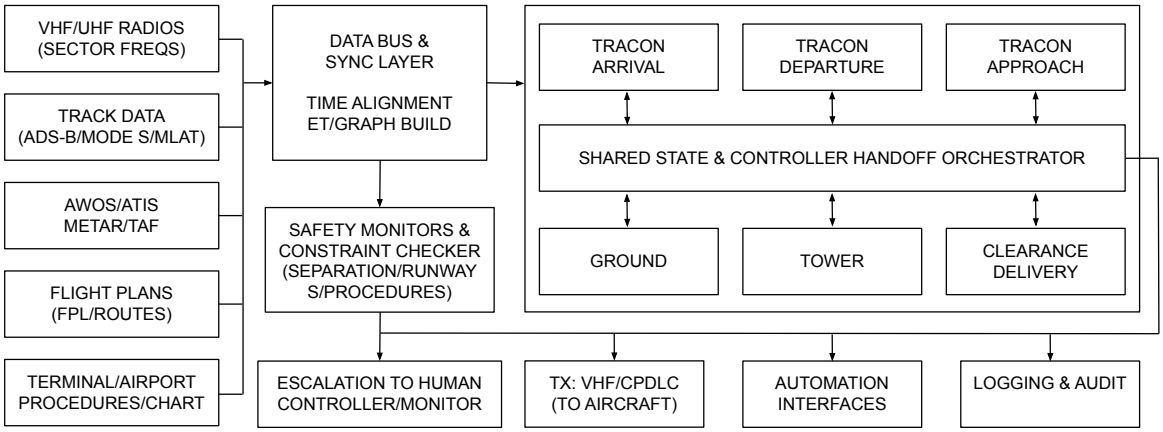
Problem Context and Motivation

Aviation network performance continues to be constrained by controller availability, fatigue, and elevated training timelines. Salary and employment data show air traffic controllers in the United States earned a median annual wage of approximately $144,580 in May 2024, with employment near 24,100; replacement demand is persistent despite slow net growth. Globally, air navigation service providers (ANSPs) report controller staffing constituting roughly 38% of their workforce, and newly issued controller licenses represent ~3.5% of total controller staffing per year, indicating ongoing training and backfill needs. Delay costs drive significant indirect losses: in 2023, U.S. passenger airlines' direct operating cost averaged $100.80 per block minute, and overall delay costs (airlines + passengers + lost demand) were estimated at $33B in 2019. In Europe, the network experienced tens of millions of minutes of ATFM delay in 2023, translating to multi-billion euro annual impacts.

System Overview

A single role-conditioned multimodal model is proposed. The ASR front-end performs transcription only. The main model handles speaker attribution (e.g., controller vs. pilot), phraseology parsing, spatial/temporal reasoning on surveillance tracks, weather and procedure conformance checks, conflict detection/resolution, runway/taxi routing, and inter-role coordination. A "role token" specifies the active job (Ground, Tower, TRACON Arrival/Departure/Approach positions), and handoffs are executed by synchronized role instances. TRACON roles are implemented by the same model with different conditioning, matching current procedures where the pilot is transferred between positions as the stage of flight changes. Clearance Delivery and Flight Data can be deployed as separate instances. AWOS/ATIS generation remains as is, which are then consumed live by the model.

FIG. 2 - ATC SYSTEM ARCHITECTURE WITH ROLE-CONDITIONED MODELS



Input Modalities and Data Acquisition

Speech

VHF ATC audio is collected globally via a network of receivers positioned near major airports and air corridors. Raw audio is chunked by frequency and facility, time-aligned, and diarized to segment transmissions. A dedicated ASR model is fine-tuned on domain-specific phraseology, then the multimodal model consumes only the transcript with aligned timing and channel metadata.

Surveillance and Tracks

ADS-B (1090ES and 978 UAT), Mode S, MLAT, and ADS-C streams are ingested. Each track record includes time, callsign/ICAO24 correlation, groundspeed, baro/GNSS altitude, vertical rate, heading/track, and latitude/longitude. TIS-B and ADS-R are exploited where available to improve traffic picture completeness. All surveillance streams are synchronized to a common clock and fused at ~1–2 Hz controller-facing cadence with higher-rate internal buffers for conflict probing.

Weather

METAR/TAF data and local AWOS/ATIS audio/text are ingested continuously. The model conditions on active runway use, wind components, ceiling/visibility, LLWS/microburst advisories, and convective SIGMETs when present.

Procedures, Charts, and Airport Data

Airport diagrams, hot spots, SID/STAR, and instrument approach procedures are ingested from public digital products as PDFs on the 56-day cycle. The ingestion strategy uses PDF vector and text extraction to produce structured airport graphs (runways, taxiways, holds, hot spots) and procedure state machines (minima, altitude/speed constraints, path terminators). This approach aligns with the public PDF sources while yielding machine-readable airport and procedure knowledge.

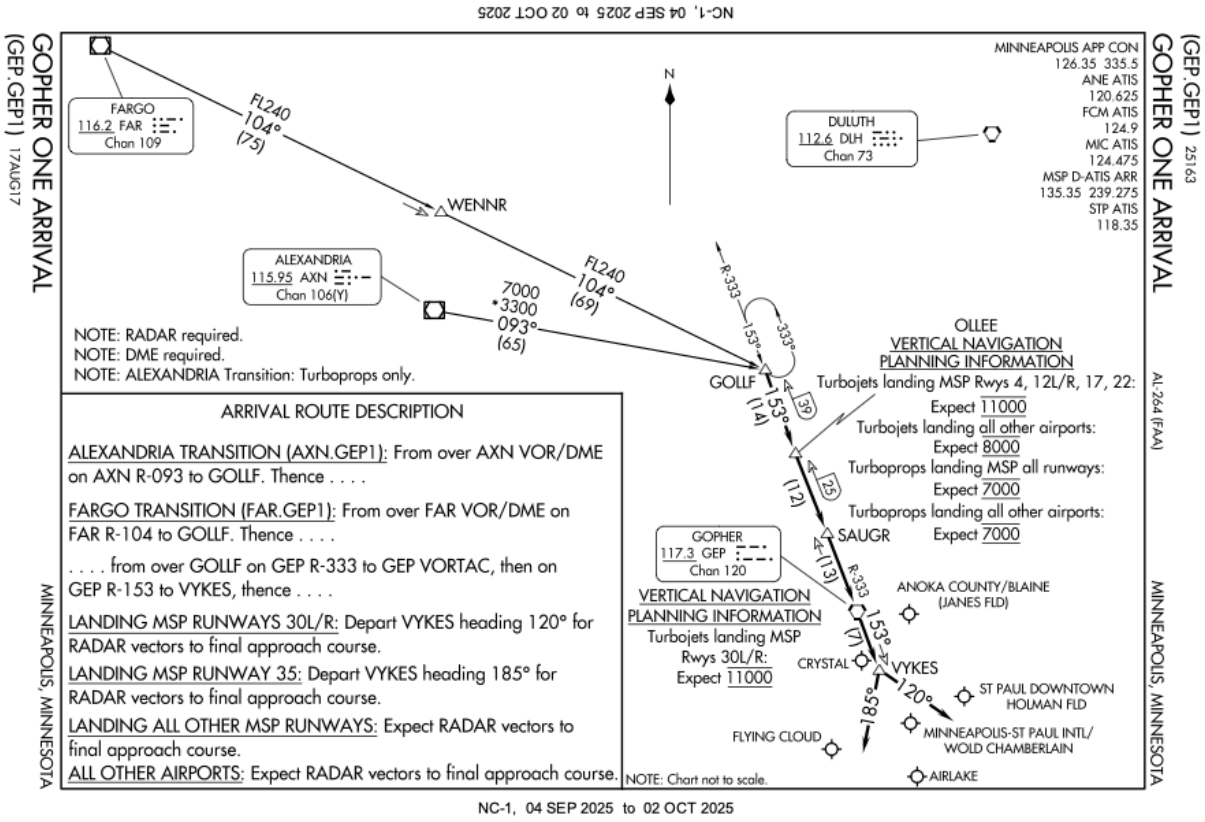


Figure: Example STAR procedure to be ingested by the model

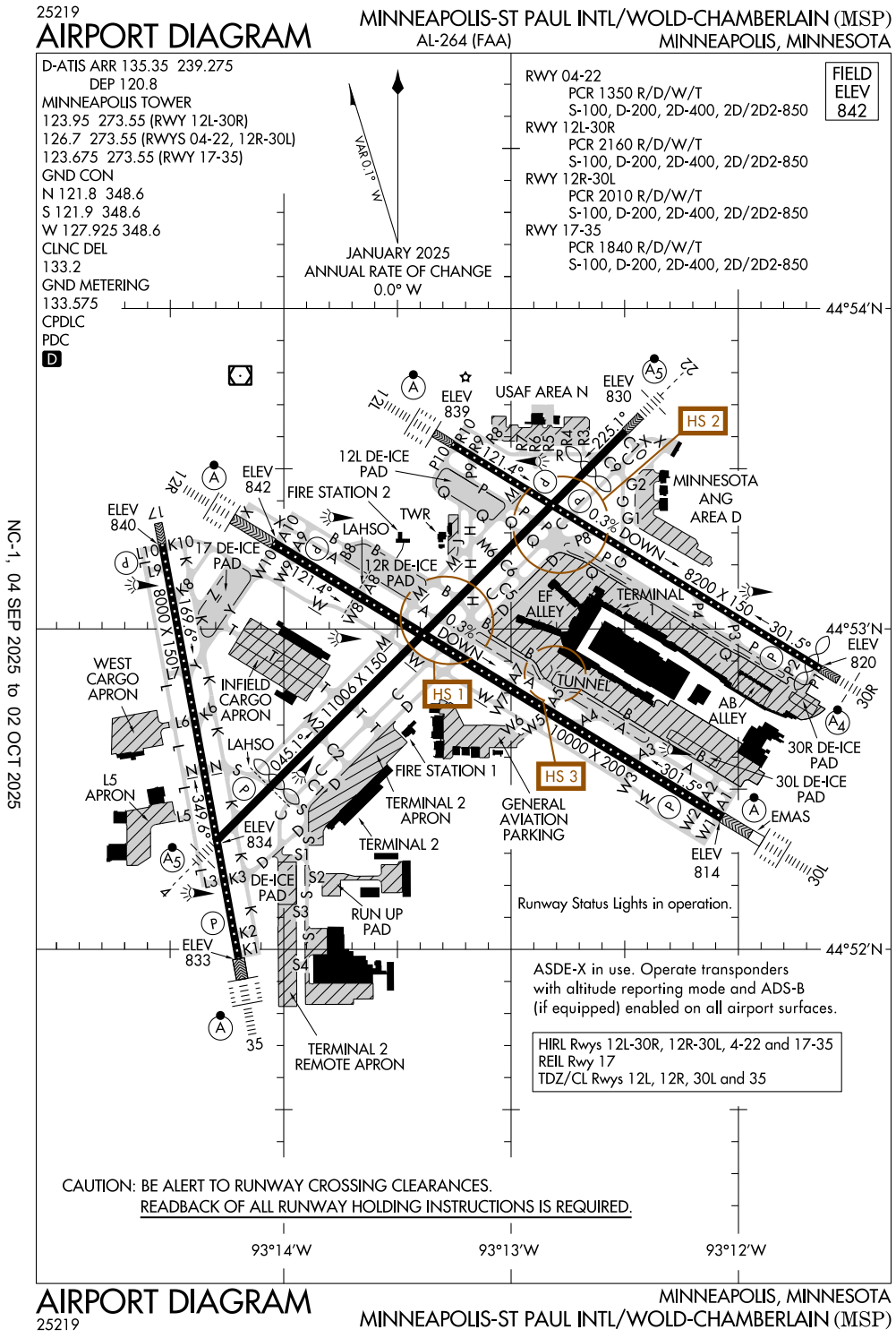
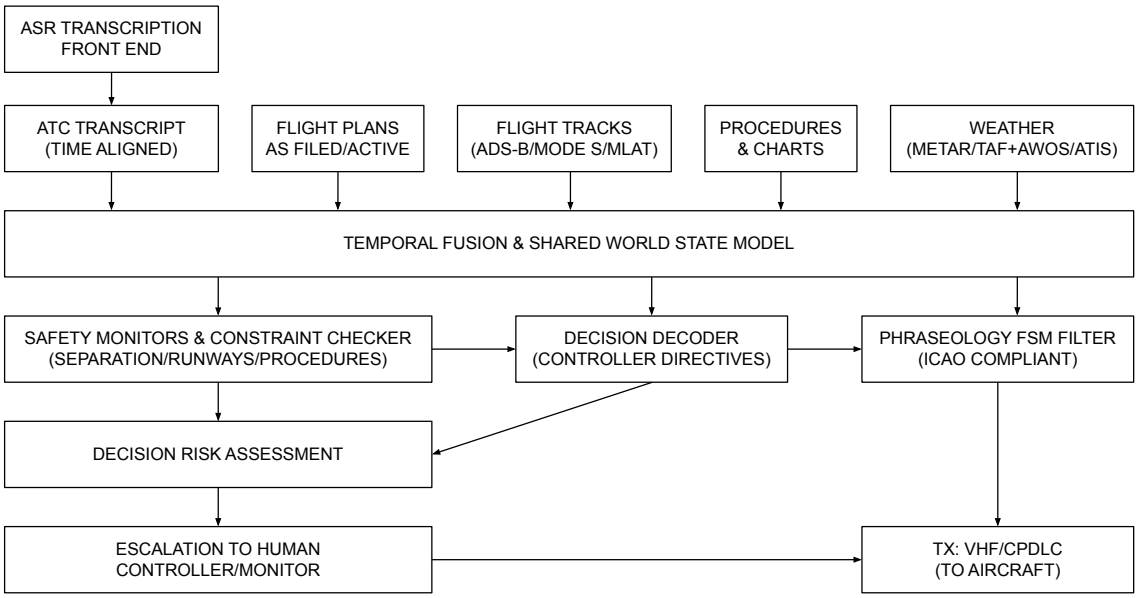


Figure: Example airport diagram to be ingested by the model

Model Architecture

FIG. 1 - UNIFIED MULTIMODAL MODEL ARCHITECTURE



Front-End ASR

A separate ASR model is trained/fine-tuned for ATC phraseology, accents, and noisy VHF. The ATCO2 and ATCOSIM resources inform labeling schemas and benchmarking, but expanded multilingual, accent-rich, facility-diverse corpora are required for global robustness; ATCO2 is valuable yet insufficient in scale/diversity for production-grade worldwide coverage.

Multimodal Encoder-Reasoner

Text (transcripts), tabular/temporal tracks, and document-derived airport/procedure graphs are embedded and fused. A temporal attention backbone maintains sector state and intent estimates. Phraseology is parsed to structured directive/acknowledgement frames tagged with entities (callsigns, runway/taxiway identifiers, squawk codes, headings, altitudes, speeds). A differentiable constraint checker projects candidate clearances against procedure constraints, terrain/obstacle buffers, protected surfaces, wake vortex categories, and separation minima. The decoder issues controller-style utterances, with a grammar/phraseology filter enforcing ICAO/State phraseology norms before transmission.

Role Conditioning and Multi-Agent Coordination

A single model is role-conditioned via soft prompts and per-facility configuration. Instances for Ground, Tower, and TRACON positions coordinate through a shared world model. Handoffs inherit the flight's latent state (intent, conformance flags, constraints, and outstanding clearances) and mirror existing system boundaries (e.g., STARS-based handoff conventions in U.S. TRACON).

Safety Monitors and Risk Escalation

Independent monitors operate alongside the decoder: separation monitors (longitudinal/lateral/vertical), runway incursion monitors using the airport graph and surface surveillance, and instruction-conformance monitors. An decision risk score is computed from conflict probability, time-to-conflict, model uncertainty, phraseology ambiguity, and traffic complexity. If the risk score exceeds a calibrated threshold, the decision is escalated to a human controller for adjudication; this threshold is increased over time as validated performance improves, reducing the share of escalated decisions.

Data Engineering and Alignment

Global collection infrastructure is established for VHF audio and 1090/978 surveillance at major airports and en route vantage points. Each site runs synchronized time sources, local buffering, and secure uplink. Audio, surveillance, filed route messages, and weather are joined by time and callsign/ICAO24 correlation. Chart PDFs are ETL-processed into airport graph and procedure constraint databases. TIS-B/ADS-R are incorporated opportunistically to fill non-equipped traffic gaps in terminal areas.

Training Regimen

A staged curriculum is proposed. Phase 1 trains instruction understanding and intent prediction from transcripts plus historic tracks. Phase 2 trains action generation with rule-aware decoding and loss penalties for separation/procedure violations. Phase 3 adds multi-role coordination and handoff consistency. Fine-tuning is performed per region to capture local phraseology variants (e.g., ICAO standard vs. CAP 413 deviations), then adapted per facility. The ASR model is trained separately on expanded ATC speech corpora; public sets such as ATCO2/ATCOSIM seed initial models but are augmented by the collected global corpus to achieve accent and environment coverage.

Evaluation, Simulation, and Shadow Operations

Safety-critical evaluation precedes any operational use. Fast-time and human-in-the-loop simulation environments are used to stress the system with high-density traffic, convective weather, runway closures, and non-nominal events. BlueSky and NASA agent-based NAS simulators are suitable candidates for fast-time and scenario generation. Performance metrics include loss-of-separation rate, conflict resolution latency, runway occupancy conformance, taxi-time and hold-time distributions, and phraseology accuracy. After simulation, a shadow mode is executed in live operations where the model makes real-time decisions while a human controller remains in full authority; discrepancies are analyzed against the human "gold standard."

Real-Time Constraints and Systems Integration

Operational latency targets match or beat human controllers for surveillance update cycles and readback loops. Integration aligns with existing automation (e.g., STARS in TRACONs/Towers) and surveillance (radar + ADS-B). The system ingests live AWOS/ATIS but does not generate them. Interfacing adheres to current handoff and identification processes, preserving existing safety nets (e.g., minimum safe altitude warnings, conflict alerts) as separate monitors during initial deployment.

Safety Assurance and Precautions

The system is deployed with layered safeguards: rule-aware decoding to prevent illegal clearances, independent monitors to veto unsafe outputs, conservative confidence gating, and human escalation on high-risk decisions. Extensive precautions are applied during rollout, including restricted operational domains, time-of-day limits, and traffic complexity caps determined in collaboration with service providers. Regulatory implementation specifics are deferred, but it is anticipated that multiple precautions and staged approvals will be required.

Deployment Strategy

Local needs dictate sequence, but a likely path begins at TRACON roles (arrival/departure/approach) where benefits accrue from improved metering, vectoring, and sequencing; proceeds to Ground, where surface routing and hotspot avoidance reduce taxi time and incursions; and finally Tower, where runway crossings, line-up-and-wait, and departure/arrival spacing are managed. The role-conditioned single model simplifies fleet management and allows staff to assign roles dynamically per facility.

Performance Targets

The final version is targeted to outperform human controllers by an order of magnitude on safety and efficiency composites. Safety goals are framed as at least 10× reduction in predicted loss-of-separation risk under matched traffic/weather, with maintained or reduced controller-equivalent latency. Efficiency gains target measurable reductions in vectoring mileage, level-offs, taxi-out time, and missed-approach rates, subject to weather and traffic constraints.

Financial Analysis: Direct and Indirect Savings

Direct Costs: Salaries and Training

A concrete U.S. reference indicates median controller wages near $144.6k and employment of roughly 24k. If the system ultimately enables a proportional reduction in controller staffing for specific roles through phased substitution and natural attrition, direct wage exposure is on the order of ~$2B annually in the U.S. alone (wages only, before benefits/overhead). Globally, controller headcount is not published as a single figure; however, sector data indicate controllers comprise ~38% of ANSP headcount and replacement licensing runs ~3.5%/year. Assuming a global controller population on the order of 50–60k and average fully loaded costs ranging from $100k to $150k, direct annual labor exposure falls around $5–$9B. Training replacement demand at 3.5% yields approximately 1.8–2.1k new licenses per year; an illustrative all-in training pipeline cost of $70k–$150k per trainee implies $125M–$315M annually before accounting for opportunity costs of on-the-job training throughput constraints. These figures are presented as order-of-magnitude exposures and are sensitive to regional labor economics and staffing policies.

Indirect Costs: Delays, Time, and Inefficiencies

In Europe, the network accumulated ~18.1 million minutes of en-route ATFM delay in 2023; when using a cost-per-minute benchmark of €127, the annual airline cost component for 22.4 million minutes has been reported near €2.8B. If the proposed system reduces only the ATC-attributable share of en-route delay by a modest 10%, savings would be roughly €280M on that base; 20% would approach ~€560M, exclusive of passenger time value and follow-on network effects. In the U.S., a $100.80 cost per block minute provides a scaling point: a hypothetical reduction of 90 seconds of taxi-out time across 6 million departures would equate to roughly $907M in direct airline operating savings, ignoring passenger time and schedule stability gains. These back-of-the-envelope calculations illustrate the sensitivity of system-level economics to small per-flight improvements.

Verification Protocol

A multi-phase verification plan is adopted. Phase A: offline backtesting on historical tapes comparing generated clearances against actual ones with safety/efficiency scoring. Phase B: fast-time simulation over seasonal traffic/weather ensembles to measure conflict rates and throughput. Phase C: controlled human-in-the-loop simulations with career controllers adjudicating edge cases. Phase D: shadow operations in active facilities with full logging and latency auditing. Escalation thresholds are tuned during Phases C–D so that only decisions assessed as significantly high risk are escalated; the threshold increases as evidence accumulates, decreasing escalations over time.

Implementation Details

Input Ingestion

All feeds are collected independently and combined through a common synchronization layer. Audio and surveillance sites are secured geographically and network-redundant. PDF ingestion of terminal procedures and airport diagrams produces both text (constraints, minima, frequencies) and geometry (taxiway/runway graph, protected areas). Updates track the FAA's 56-day cycle and analogous cycles in other FIRs.

Model Networking

Streaming encoders keep rolling context windows for each flight and sector. A per-flight memory holds last clearance, readback status, conformance residuals, and intent. A sector memory tracks sector-level metering, miles-in-trail, and runway configuration. The decoder is constrained by a phraseology finite-state machine that outputs standard radiotelephony and blocks non-standard forms. A separate "hard guard" enforces regulatory minima and terrain/obstacle clearances.

Interaction with Existing Automation

The system reads track and flight plan updates from current platforms and publishes candidate advisories/clearances via existing data paths. In U.S. TRACONs, integration with STARS is used for track association and handoff mirrors. During early deployment, the system runs in advice-only mode with human acceptance required before transmission.

Compute and Runtime

Training and inference are hardware-agnostic; throughput targets align to surveillance updates and sub-second action proposal latencies. Facility-level sharding and per-sector replicas provide resilience and scalability.

Data Resources and Prior Art

The ATCO2 project provides a multilingual ATC speech corpus with annotations (NER, code-switching) and research baselines, and the ATCOSIM corpus offers simulator-based English ATC speech; both are valuable for benchmarking and bootstrapping, but the envisioned system requires broader accent, noise, and facility coverage to serve global operations. The Whisper family enables strong base ASR that can be adapted to ATC conditions. ADS-B/TIS-B/ADS-R technical references inform surveillance fusion and coverage assumptions, and FAA digital chart products supply authoritative procedures and aerodrome geometry.

Limitations and Risk

Explainability is recognized as difficult for neural decision systems, and supplemental artifacts (e.g., rationale traces mapping from constraints to proposed actions) will be produced to aid post-hoc review. Global phraseology and procedural variation necessitate region/facility adaptation. Rare edge cases and degraded surveillance environments require conservative fallbacks and escalation. Public acceptance is acknowledged as a long-horizon consideration, although early successes in low-risk roles and transparent safety evidence are expected to aid confidence.

Conclusion

A role-conditioned, unified multimodal model paired with a dedicated ASR front-end, strong monitors, and conservative escalation offers a credible path to superior safety and efficiency relative to human baselines. By aligning with current operational partitions (TRACON, Ground, Tower), ingesting authoritative procedure PDFs, and integrating with existing automation such as STARS, migration risk is minimized. Simulation results and shadow operations will quantify improvements in separation assurance, throughput, taxi-time, vectoring, and missed-approach rates. The target end state is a system that is at least 10× better on safety/efficiency composites than human controllers, operating in real time with negligible or negative latency deltas, and delivering substantial direct and indirect economic benefits at network scale.

References

Airlines for America. (2024, July 12). *U.S. passenger carrier delay costs*. https://www.airlines.org/dataset/u-s-passenger-carrier-delay-costs/

Civil Air Navigation Services Organisation. (2025). *What the data tells us*. https://canso.org/what-the-data-tells-us/

EUROCONTROL Performance Review Commission. (2024, May). *Performance review report 2023*. https://www.eurocontrol.int/publication/performance-review-report-prr-2023

Federal Aviation Administration. (2022, February 25). *Terminal Automation Modernization and Replacement (TAMR)*. https://www.faa.gov/air_traffic/technology/tamr

Federal Aviation Administration. (2023, February 7). *ADS-B In pilot applications: Automatic dependent surveillance—rebroadcast (ADS-R)*. https://www.faa.gov/air_traffic/technology/adsb/pilot

Federal Aviation Administration. (2023, February 7). *Traffic Information Services–Broadcast (TIS-B)*. https://www.faa.gov/air_traffic/technology/equipadsb/capabilities/ins_outs

Federal Aviation Administration. (2024, August 29). *Digital—Chart Supplement (d-CS)*. https://www.faa.gov/air_traffic/flight_info/aeronav/digital_products/dafd/

Federal Aviation Administration. (2025, January 29). *Terminal Procedures Publication (d-TPP) / Airport Diagrams* (56-day cycle). https://www.faa.gov/air_traffic/flight_info/aeronav/digital_products/dtpp/

Hoekstra, J. M., & Ellerbroek, J. (2016). *BlueSky ATC simulator project: An open-data and open-source approach*. Delft University of Technology. https://pure.tudelft.nl/ws/files/10083831/Hoekstra_BlueSky_project.pdf

National Aeronautics and Space Administration. (2021). *Airspace Concepts Evaluation System (ACES)* (ARC-15068-1). NASA Software Catalog. https://software.nasa.gov/software/ARC-15068-1

OpenAI. (2022, September 21). *Introducing Whisper*. https://openai.com/index/whisper/

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision* (Whisper). OpenAI. https://cdn.openai.com/papers/whisper.pdf

U.S. Bureau of Labor Statistics. (2025). *Air traffic controllers* (Occupational Outlook Handbook). https://www.bls.gov/ooh/transportation-and-material-moving/air-traffic-controllers.htm

U.S. Bureau of Labor Statistics. (2024). *Occupational Employment and Wage Statistics: Air traffic controllers, May 2023*. https://www.bls.gov/oes/2023/may/oes532021.htm

Zuluaga-Gomez, J. (n.d.). *ATCOSIM corpus* [Dataset]. Hugging Face. https://huggingface.co/datasets/Jzuluaga/atcosim_corpus

ATCO2 Consortium. (n.d.). *ATCO2: Automatic collection and processing of ATC communications* (Project overview). https://www.atco2.org/

ATCO2 Consortium. (n.d.). *ATCO2 data: Corpora overview*. https://www.atco2.org/data

Optional supporting/alternative citations (use as needed)

EUROCONTROL. (2024, January 18). *European aviation overview: 2023 review* (CODA/Network performance). https://www.eurocontrol.int/sites/default/files/2024-01/eurocontrol-european-aviation-overview-20240118-2023-review.pdf

FAA. (n.d.). *Standard Terminal Automation Replacement System (STARS): Detailed overview*. https://www.faa.gov/about/office_org/headquarters_offices/ang/offices/tc/library/storyboard/detailedwebpages/stars.html

BlueSky ATC Simulator—project repository. (n.d.). *The open-source air traffic simulator* (TU Delft CNS-ATM). https://github.com/TUDelft-CNS-ATM/bluesky