Engage 2

GenAI models for ATM

Final technical report

Engage 2 catalyst fund project

Coordinator: Nommon Solutions and Technologies

Consortium partners: Tecnalia

Thematic challenge: TC3 Disruptive ATM system modernization

Edition date: 07/01/2026

The GenAI models for ATM project has been supported by the SESAR 3 Joint Undertaking and its founding members under the Grant Agreement nr. 101114648. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or SESAR 3 JU. Neither the European Union nor the SESAR 3 JU can be held responsible for them.

SUPPORTED BY sesar JOINT UNDERTAKING

Co-funded by the European Union

Authoring & approval

Author(s) of the document

Organisation name	Date
Nommon	08/10/2025
Tecnalia	08/10/2025

Reviewed by

Organisation name	Date
Nommon	08/10/2025
Engage 2 mentors, University of Trieste; University of Westminster	06/11/2025
Engage 2 Task Leader, University of Westminster	19/11/2025

Approved for submission to the SESAR 3 JU by[1]

Organisation name	Date
Micol Biscotto, Deep Blue srl	07/01/2026

Rejected by[2]

Organisation name	Date

Document history

Edition	Date	Status	Company Author	Justification
01.00	08/10/2025	Submitted to the Engage 2 consortium	David Mocholí	Initial version
01.01	23/12/2025	Revised following mentor feedback	David Mocholí	Updated version
02.00	07/01/2026	Ready for publication	Engage 2	Final version

[1] Representatives of all the beneficiaries involved in the project

[2] Representatives of the beneficiaries involved in the project

Engage 2

THE SESAR 3 KNOWLEDGE TRANSFER NETWORK

Engage 2

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114648 under European Union's Horizon Europe research and innovation programme.

Table of contents

List of figures

List of tables

1 Introduction

1.1 Abstract

Generative Artificial Intelligence (GenAI), has recently emerged as a powerful technology with diverse applications across multiple domains, including natural language processing. In this project, we explored the possible applications of GenAI in the field of air traffic management (ATM), aiming to find areas where GenAI can add value over current solutions. A literature review was conducted, concluding with a catalogue of candidate use cases. From this list a specific use case was selected and further developed as a proof of concept to demonstrate the impact of GenAI technologies. The selected use case is titled "Annotating NOTAMs with a Tag System using GenAI", for which we developed a solution leveraging the bi-directional encoder representations from transformers (BERT) large language model (LLM) architecture. The problem was framed as a multi-label classification problem, allowing multiple tags per NOTAM, and a semi-supervised training scheme, thus leveraging unlabelled data to improve the generalizability and robustness of the model. We tested the model on a test dataset to measure the improvement in accuracy achieved by the proposed framework.

1.2 Executive summary

The European Air Traffic Management (ATM) system is facing three major challenges: rapidly increasing traffic demand, increasing environmental concerns, and the introduction of new airspace users such as drones and low-emission aircraft. The Digital European Sky (DES) vision addresses these demands, while ensuring safe and efficient air traffic. However, the transition towards DES requires the progressive implementation of data-driven, automated, and adaptive airspace management.

Generative Artificial Intelligence (GenAI) has emerged as a key enabler in this transformation. Recent advances in neural architectures and data availability make GenAI particularly suitable for complex, high-dimensional, and language-based ATM tasks. This project explored how GenAI can be strategically applied to ATM through two core objectives: identifying the most promising GenAI use cases in the sector, and developing a proof of concept to demonstrate impact.

In the first phase, a comprehensive literature review and technology assessment [1] was completed, with the aim of identifying areas in which GenAI has the potential to provide added value over existing solutions. This resulted in a catalogue of 10 GenAI-enabled ATM use cases, spanning synthetic data generation, forecasting, anomaly detection, language processing, and content generation. Each use case was evaluated against relevance, GenAI added value, interpretability, and data availability in consultation with experts and industry stakeholders [2,3,4].

Based on this assessment, the use case **"Annotating NOTAMs with a Tag System using GenAI"** was selected for implementation. Unlike earlier classification models that use mutually exclusive categories to annotate NOTAMs, this project framed NOTAM annotation as a **multi-label tagging problem**, allowing each message to receive multiple context-specific labels. A semi-supervised bi-directional encoder representation from transformers (BERT)-based architecture was developed, combining supervised fine-tuning with **MixMatchNL learning** and **focal loss** to leverage both labelled and unlabelled data.

The proposed model achieved **significant performance improvements** over the baseline, with micro F1-score increasing from **0.93 to 0.96**, subset accuracy from **0.78 to 0.88**, and **Hamming loss reduced by nearly half** [5,6]. The approach proved scalable and robust across different message types and lengths. In addition, a **labelled dataset of 8,497 NOTAMs**, the first of its kind for multi-label classification, was created and publicly released [7].

Beyond the proof of concept, the project's findings indicate that while GenAI is often applied to problems where its complexity offers limited benefits, it provides clear advantages in domains with complex, high-dimensional data or natural-language content, where transformer-based models represent the state of the art. These architectures excel at capturing intricate spatio-temporal relationships that traditional machine-learning approaches cannot model effectively.

Looking ahead, future work will focus on refining the NOTAM tagging system with hierarchical label structures, testing larger and more advanced transformer models, and developing a user interface for operational validation on larger datasets. We also plan to pursue SESAR and national research funding to raise the solution's TRL and extend the research towards other promising GenAI applications identified in this study, such as trajectory generation and synthetic data modelling.

2 Overview of catalyst project

2.1 Operational/technical context

Air Traffic Management (ATM) oversees the coordinated control of airspace through services such as Air Traffic Control, Airspace Management, and Flow and Capacity Management, with the primary goal of ensuring safe and efficient air traffic. ATM is a continuously evolving field due to the constant evolution of aircraft and the growing demand for air transportation. The 2025 European Master Plan identifies some key transformation levers such as [1]:

1. **Rising traffic demand:** The steady increase in conventional flights requires ATM to manage higher traffic volumes without compromising safety. SESAR promotes Trajectory-Based Operations (TBO), which rely on continuously shared four-dimensional flight paths and real-time data [8] through System-Wide Information Management (SWIM). Since air traffic behaves unpredictably, the Single European Sky framework must be designed to handle uncertainty and complexity.
2. **Environmental concerns:** Aviation accounts for roughly 4% of EU greenhouse gas emissions, a share likely to rise as other sectors decarbonize. While fossil-fuel dependency limits full elimination of emissions, improving ATM efficiency could reduce aviation emissions by up to 9% [9]. Strategies include optimizing flight routes to minimize contrails and avoiding inefficiencies caused by airspace constraints or airport congestion.
3. **Integration of new airspace users:** Traditional ATM systems focus on commercial jet traffic, but new vehicle types—such as drones, low-altitude zero-emission aircraft, and high-speed supersonic transports—will operate at diverse speeds and altitudes. This demands new frameworks like U-space for drone management and updated procedures to help air traffic controllers manage increasingly heterogeneous traffic.

To respond to all three challenges, the **Digital European Sky** vision proposes more precise, data-driven, and automated trajectory management. This transition will require a shift toward dynamic airspace allocation, higher automation, and evolving roles for human operators. In this context, **Artificial Intelligence (AI) has emerged as a versatile tool**, with the capability to process vast amounts of data rapidly and efficiently. In particular, Generative AI (GenAI) has recently gained attention as an exciting and fast-growing area within the field [10].

Although generative models have existed for several decades, newly developed neural network architectures combined with enhanced hardware have resulted in models that are able to generate intricate and high-dimensional synthetic data in the form of text, audio, images, etc. This makes GenAI a valuable tool in a range of disciplines, including ATM. GenAI has been tested by ATM researchers as a solution to a range of challenges in ATM such as trajectory prediction for commercial aircraft and the processing of ground delay reports [10]. GenAI has also been explored as a potential tool to generate synthetic data for ATM [11] which can be used for purposes such as training other algorithms, data anonymisation, balancing unbalanced datasets and performing Monte Carlo simulations. Furthermore, the greater availability of data that is expected from digitalisation and SWIM, should amplify the capabilities of GenAI models, allowing them to fuse different forms of data from a variety of sources to provide unique and transformative solutions that can propel the evolution of the European ATM system.

2.2 Project scope and objectives

This project explored potential applications of Generative Artificial Intelligence (GenAI) in the Air Traffic Management (ATM) domain and developed a proof of concept to demonstrate its transformative potential through a selected use case. The project was structured around two main objectives:

1. **Identification of GenAI opportunities in ATM**: Analyse in detail the challenges within the ATM system that could benefit from GenAI-based solutions (e.g. synthetic data generation, natural language interpretation). Conduct a comprehensive review of applicable GenAI techniques—such as Generative Adversarial Networks (GANs) and Large Language Models (LLMs)—to identify one or more promising use cases for further investigation.
2. **Demonstration of GenAI potential through a proof of concept**: Develop a prototype solution for the selected use case(s), evaluate its effectiveness in addressing the identified challenges, and gather feedback from relevant stakeholders.

The project was divided into two phases: the **first half** addressed the analytical and exploratory work to identify suitable applications, while the **second half** focused on the technical implementation and validation of the selected use case.

Phase 1 – Identification and analysis of GenAI opportunities

The first phase aimed to understand where GenAI could bring tangible benefits to ATM. Its specific sub-objectives were:

- Identify existing ATM problems for which GenAI solutions have already been explored.
- Analyse characteristics of ATM and transport-related problems where GenAI has demonstrated clear advantages over conventional approaches.
- Detect ATM problems that are well suited for GenAI applications but have not yet been investigated.
- Assess the strengths, weaknesses, and limitations of current GenAI techniques.
- Build a **catalogue of potential GenAI applications in ATM**, highlighting their advantages over existing methods, required data sources, and proposing suitable model architectures for each use case.
- Select one use case from the catalogue for demonstration within the project timeframe.

Phase 2 – Proof of Concept: "Annotating NOTAMs with a Tag System using GenAI"

Based on the findings from Phase 1, the use case **"Annotating NOTAMs with a Tag System using GenAI"** was selected for implementation. Previous research had explored the classification of NOTAMs using both traditional machine learning models and LLMs; however, these approaches typically relied on mutually exclusive categories, limiting flexibility and contextual understanding.

The second phase pursued the following sub-objectives:

- Develop a model capable of assigning **multiple tags to a single NOTAM**, reframing the problem as a **multi-label classification** task.
- Implement a **semi-supervised or unsupervised training framework** that outperforms a purely supervised approach using the same LLM baseline.
- Validate the proposed framework on a **test dataset**, demonstrating measurable improvements in model performance.

2.3 Research carried out

This section is structured around the two main objectives of the project. The first subsection describes the research activities conducted during the **first half**, focused on identifying and analysing potential applications of Generative Artificial Intelligence (GenAI) in Air Traffic Management (ATM). The second subsection presents the research carried out during the **second half**, which focused on developing and testing the selected proof of concept.

2.3.1 Phase 1- Identification and analysis of GenAI opportunities

During the first half of the project, we conducted an extensive **desk research** exercise to identify and assess potential GenAI applications in ATM. This process consisted of three main steps:

1. Identifying relevant GenAI application areas in ATM.
2. Assessing GenAI techniques and their applicability to ATM-related problems.
3. Selecting the most promising use cases for further development.

The conclusions from this work are presented in **Section 3.1**.

Identification of Potential GenAI Applications in ATM

For this task, we compiled and analysed relevant literature on how Generative AI has been applied to address problems in ATM. In parallel, we reviewed studies based on traditional methods to assess whether generative models provide a genuine performance advantage over existing approaches. We also identified ATM challenges that have not yet benefited from GenAI solutions but where the technology shows clear potential. In addition, we examined applications of GenAI in other transport domains that could be adapted to ATM contexts.

The review was structured into five main GenAI application areas:

- Data augmentation and synthetic data creation
- Forecasting and prediction
- Anomaly detection and classification
- Language processing
- Content generation

Within each category, the literature was grouped by specific ATM use cases, forming the basis for the development of our **GenAI use case catalogue**.

Assessment of GenAI techniques and applications

We then performed an in-depth review of GenAI methodologies across all key modelling stages: (i) data preparation, (ii) model selection, (iii) training, (iv) evaluation and tuning, and (v) data generation.

The following families of generative models were reviewed in detail:

- Generative Adversarial Networks (GANs) [12]
- Energy-Based Models (EBMs) [13]
- Variational Autoencoders (VAEs) [14]
- Flow-based and diffusion models [15]
- Transformers [16]
- Autoregressive and hybrid architectures

Finally, we included a dedicated section on Explainable Artificial Intelligence (XAI), summarising current techniques that enhance the interpretability and transparency of generative models and their outputs. Further details on this analysis can be found in Deliverable D2.1 [1].

Development of the catalogue of GenAI use cases in ATM

Building upon the literature review and technology assessment, we developed a catalogue of 10 potential GenAI use cases for ATM. Each use case was described using a consistent structure to ensure comparability and clarity:

- Summary: Description of the proposed GenAI tool and its intended ATM application.
- Data requirements: Specification of input, output, and contextual data, as well as any required services or tools (e.g. simulators).
- Added value: Explanation of how GenAI could provide advantages over existing solutions.
- Proposed GenAI algorithm: Identification of the neural network architecture to be employed (existing or newly designed).
- Algorithm added value: Justification of the selected GenAI approach and its expected benefits.
- Related work: References to prior research relevant to the use case.

This catalogue served as the foundation for selecting the use case to be implemented in the second half of the project.

Selection of the GenAI use case

The selection of the use case was performed based on 4 separate criteria:

- Feedback received from stakeholders and experts.
- Interests of the partners for future development.
- Data availability.
- Technical feasibility of completing a successful demo in the available time frame.

To gather feedback from stakeholders and experts an online workshop was organized. This was composed by experts on AI and ATM. During the workshop, the attendees were asked to rate the use cases from 1 to 5 in the following categories:

- Relevance to ATM: how relevant is this use case to current challenges in ATM?
- GenAI added value: how much value does the GenAI solution we propose add over current solutions?
- Interpretability: how important is it that the model used for this use case is interpretable?
- Datasets available: how available is the data required to train and/or evaluate the GenAI model?

The quantitative feedback given by the stakeholders in the workshop is presented on Figure 1 below. Apart from the average stakeholder ratings on each of the four categories, we also compute a "Suitability" rating for each use case. This was calculated as a weighted average of the four categories:

$$Suitability \;=\; 0.4(Relevance \;+\; GenAI \;+\; Data\;) \;-\; 0.2\; Interpretability.$$

	Relevance to ATM	Gen AI Added value	Interpretability	Datasets navailable	Suitability
UC1	4,1	3,7	4,3	2,4	3,22
UC2	4,4	2,6	3,4	1,6	2,76
UC3	2,5	1,8	1,4	3	2,64
UC4	4,5	2,8	3,3	3,8	3,78
UC5	4,6	3,4	3,2	3,8	4,08
UC6	4,8	3	1,5	2,8	3,94
UC7	4	2,7	2	3	3,48
UC8	3,7	4	2	3,7	4,16
UC9	5	3,8	2,3	4,2	4,74
UC10	1,4	4,5	1,6	3,8	3,56

Figure 1. Averaged ratings given by the stakeholders to the use cases at the workshop [4].

The weight for interpretability is negative since making generative models interpretable is extremely challenging due to the high number of trainable model parameters. The results of this selection are shown in Section 2.4.1.2.

2.3.2 Phase 2- Proof of concept

The **selected use case** for implementation and demonstration was **"Annotating NOTAMs with a Tag System using GenAI"** (see Section 2.4.1.1). This work addresses several gaps identified in the literature regarding the automated processing of NOTAMs using language models:

- Problem formulation: Unlike previous studies that applied Large Language Models (LLMs) for NOTAM classification [32,36], we reframed the task as a multi-label classification problem, enabling each NOTAM to be assigned multiple context-specific labels. This approach enhances the descriptive power of the model and reduces the risk of misclassification.

- Lack of labelled data: Since multi-label classification of NOTAMs has not been explored before, no labelled datasets were available. To overcome this, we developed an initial labelling system and manually annotated a subset of NOTAMs. The resulting dataset, along with its train–test split, has been made publicly available [7].

- Learning approach: We implemented a semi-supervised learning framework based on a BERT model, combining labelled and unlabelled data. This method allows the model to exploit the structure of unlabelled data while retaining the knowledge captured from labelled examples, improving generalization and robustness.

Methodology

The processing pipeline of the semi-supervised multi-label NOTAM classification model (Figure 2) comprises four main stages [6]:

1. **Preprocessing and Embedding**: Each NOTAM text is first normalized into natural language by expanding all contractions and abbreviations in accordance with ICAO definitions. The text is then tokenized, segmenting it into smaller units (tokens) such as words or subwords. These tokens are converted into contextual embeddings via a BERT encoder, producing vector representations that capture semantic and syntactic information for subsequent classification.
2. **Supervised Pretraining:** The BERT model is initially fine-tuned on the labelled portion of the dataset to learn task-specific linguistic patterns. This phase establishes a solid foundation before introducing unlabelled data.
3. **Semi-Supervised Enhancement with MixMatchNL:** The MixMatchNL approach [37] integrates unlabelled NOTAMs into the training process by combining pseudo-labelling, consistency regularization, and sample mixing. In practice, the model first predicts labels for unlabelled NOTAMs, which are then refined and combined with the manually labelled data to generate additional synthetic training samples. This allows the model to better exploit the structure of the large unlabelled corpus, improving robustness, generalization, and performance on rare or multi-topic NOTAMs. A detailed description of the MixMatchNL formulation and its mathematical implementation can be found in [6].
4. **Fine-Tuning with Focal Loss:** Finally, the model is refined using **Focal Loss**, which mitigates class imbalance by focusing learning on harder, less frequent labels, thereby improving overall accuracy

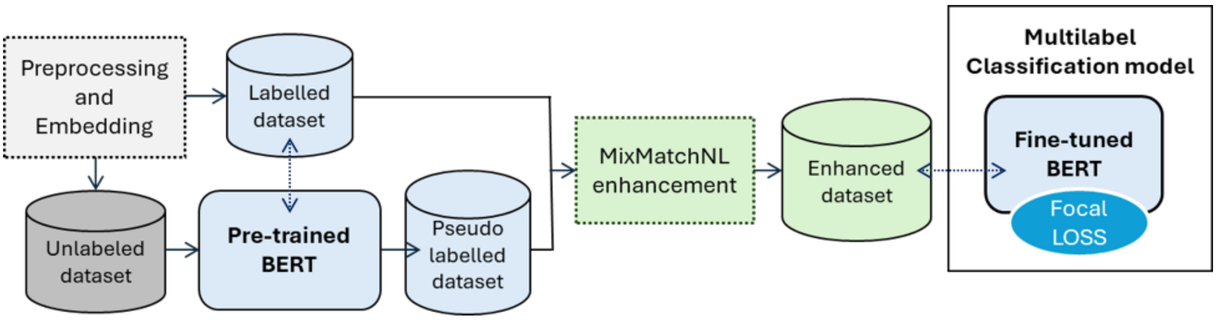

Figure 2. The architecture of the semi-supervised multi-label NOTAM classification model [4].

A total of **107,371 NOTAMs** were collected from public online repositories [32,38]. From this corpus, **8,497 NOTAMs** were manually labelled, of which **1,700** were reserved for testing.

For this first version of the labelling system, we allowed every possible combination of **11 distinct labels**, with no restriction on the number of labels per NOTAM. The labels and their definitions are as follows:

Table 1. Type of labels

Label	Definition
Runway	NOTAMs directly related to a runway.
Taxiway	NOTAMs directly related to a taxiway.
Aerodrome	NOTAMs referring to airports or terminal areas, excluding specific runways or taxiways.
Obstacle	NOTAMs concerning obstacles near aerodromes or relating to obstacle avoidance altitude.
Lighting issues	NOTAMs involving lighting systems (e.g. obstacle, runway, or approach lights).
Hazards	NOTAMs describing potentially dangerous situations (e.g. wildlife, drones, fireworks).
Navigation aid	NOTAMs concerning radio, visual, or electronic navigation aids.
Procedure	NOTAMs giving instructions or procedures for handling specific occurrences.
Airspace	NOTAMs concerning airspace-related disruptions or navigation aid failures.
Closure	Complete unavailability of an infrastructure or service (e.g. runway, aerodrome).
Restriction	Partial or conditional closure of infrastructure, airspace, or procedures.

This dataset represents the **first publicly available resource** for the multi-label classification of NOTAMs and forms the foundation for subsequent model development and benchmarking.

2.4 Results

In this section, we present the project results, once again distinguishing between each phase of the project.

2.4.1 Phase 1- Identification and analysis of GenAI opportunities

Using the methodology outlined in Section 2.3.1 and based on the literature review, we proposed 10 specific potential use cases of GenAI in the domain of ATM. Each of these cases is outlined below. For more details on the data requirements and the advantages of the proposed GenAI algorithms, please see D2.1 [1].

2.4.1.1 Catalogue of use cases

<u>1. Trajectory generation and conflict resolution in TMA</u>

Summary: based on expected arrival times of aircraft to the TMA and their access point, use a generative technique to generate a conflict free set of trajectories by imitating successful and efficient resolutions. Include physical restrictions and a check to ensure that separation regulations are obeyed.

If Air Traffic Control (ATC) messages are also included in the dataset, one can also predict/ generate the actions that the Air Traffic Controller (ATCO) has to take to ensure that each aircraft lands safely.

Added value: Trajectory generation approaches in literature typically process only a single trajectory at a time. This can be useful to make stochastic simulations, but it is not as helpful as a tool for decision making. This model would provide solutions to predicted conflicts in TMAs including the ATCO actions that are required and when they are required. Furthermore, through the inclusion of contextual data sources, such as weather data and flight plans, the generated trajectories can better resemble real data.

Proposed GenAI algorithm: Conditional Tabular Generative Adversarial Networks (CTGAN) [17]and for prediction of the ATCO reactions Variational Auto-encoders (VAE) [14].

2. Synthetic Generation of Loss of Separation Events

Summary: Loss of Separation (LoS) events occur when the distance between two aircraft in the airspace in the horizontal or vertical direction goes below a safety threshold. This safety threshold is set to avoid mid-air collisions. These events are rare and thus there are considerably fewer trajectories with LoS events than trajectories without them. The aim of the tool developed in this project would be to synthetically generate LoS events in a specific airspace to obtain more balanced datasets.

Added value: Loss of Separation events are not very frequent; this means that most trajectory datasets have very few of these types of occurrences. Training a LoS detection model on these unbalanced datasets would not result in a great performance. A mitigation for these is balancing the dataset using synthetic data with minority over sampling techniques.

Proposed GenAI algorithm: LoS-GAN (Loss of Separation Generative Adversarial Network)

- GAN Architecture [13], generates new trajectories and evaluates the realism of them
- LoS event Injection: Adjusts generated trajectories to create scenarios where the distance between aircraft falls below the safety threshold. Essentially, it introduces potential conflict situations (LoS events) into the synthetic data, providing a diverse set of scenarios not present in real data.

3. Airport passenger inflow forecasting

Summary: predict how many passengers are going to be arriving at the airport on each security checkpoint based on arrivals and departures, and any other available data. The prediction should be continuously updated for a time window of 5 min to 1 day in the future. Based on this, one can decide in advance how many security booths to use for the check in advance and make access to airports quicker.

Added value: This topic has been explored in literature but the only GenAI technique applied so far are Long Short Term Memory networks (LSTM) [18], although other GenAI methods have been applied to similar problems in other industries [19]. Given a sufficient number of outputs to be predicted (i.e. big airport with many access points), and a large prediction window GenAI could provide significantly more accurate predictions compared to simpler techniques. The extent to which GenAI can provide an improvement over simpler algorithms would have to be tested.

Proposed GenAI algorithm: Stacked autoencoder (SAE) combined with other forecasting methods, i.e., DNN or LSTM [20].

4. Prediction of Loitering Aircraft in TMA before landing

Summary: One of the main causes of flight inefficiencies is loitering in TMAs, since the aircraft has to fly additional miles, which results in extra cost and emissions, due to non-optimal scheduling. Predicting this kind of events in advance would allow ATCOs to devise actions that minimise the final collective impact on delays and emissions by changing the aircraft trajectories without resorting to loitering. The tool would aid the extended arrival manager (E-AMAN), a position proposed by SESAR which consists in planning the arrival streams at a further distance from the airport compared to the standard TMA.

Added value: although trajectory prediction and trajectory generation are topics which are being thoroughly explored, often loitering / rerouting trajectories are excluded from the data. However, loitering represents a significant inefficiency in flights and should be eliminated when TBO is implemented. The latent space representation that is part of GenAI models such as VAEs or GANs could facilitate the classification task if the problem is framed similarly to an anomaly detection problem.

Proposed GenAI algorithm: Montecarlo Simulation (MC) of artificially generated trajectories by means of a VAE algorithm [21,22,23].

5. Generative Adversarial Imitation Learning for Multi Agent Taxi Speed prediction and Conflict resolution

Summary: The aim of this tool would be to learn taxi speed model policies using imitation learning from real data. This would require representing the taxiways in an airport as a graph network as is done by Pham *et al.* [24]. The agents would be trained using a decentralised algorithm, with multiple agents taking part in a single simulation. A possible extension is to Incorporate a centralised conflict resolution algorithm that imitates an ATCO.

Added value: One of the strategic development objectives in the SESAR 2025 Master Plan is "alerts for reduction of collision risks on taxiways and runways". The development of this tool would aid in predicting this type of events in runways as it would allow for simulations to be performed in real time to predict possible conflicts. This project would also allow us to explore how simulations can be coupled effectively to generative models.

Proposed GenAI algorithm: GAIL with supervised learning

6. Long Term Daily Origin-Destination Matrix Prediction for Airports

Summary: Origin-Destination (OD) matrices concisely describe the mobility patterns in a country or region, thus allowing the authorities to better organise the transportation resources available. Hence, the topic of OD-matrix prediction has been explored in literature by several authors, with solutions ranging from analytical / intuition models to machine learning models [25,26]. GenAI approaches have also been tested, such as the ODformer by Huang *et al.* [26]. Nonetheless, GenAI approaches have not yet been applied to OD-matrices for air traffic, with the latest contribution coming from Sudakov [27] who used fuzzy logic to predict how changing the available routes would change the OD-matrix.

Added Value: to our knowledge, although GenAI prediction approaches have been used for OD matrices for other transportation modes, this would be the first attempt to use it to predict air traffic. We think this tool could help a variety of stakeholders, ranging from airlines, airports (by using the OD matrix to estimate the inflow / outflow at the airport) and air traffic managers as it could also give a long term estimate of the airspace demand.

Proposed GenAI Algorithm: Conditional Wasserstein Generative Adversarial Network with Gradient Penalty (CWGAN-GP) [28].

7. An Audio Enhancement Approach to Automatic Speech Recognition (ASR) for Air Traffic Control

Summary: Denoising ATC audio using GenAI techniques, such as diffusion models to improve automatic speech recognition scores. This has been done by Yu *et al.* [29] for Mandarin Chinese. However, in this study they only focus on the SE part and use a loss function to account for the ASR side. A more complete approach would be to integrate a denoising architecture into the ASR algorithm, thus being able to use the ASR loss function.

Added value: although the latest approaches such as whisper-ATC [30] achieve an impressive score of Word Error Rate (WER) below 2% for clean speech audios, the WER increases up to 13.5% for real ATC speech audios. We think that adding a denoising network to an existing ASR algorithm could boost its performance by a significant margin given the results shown by the separate SE and ASR approach of ROSE in Chinese audios [29].

Proposed algorithm: A denoising Autoencoder integrated in an ASR model

1. Preprocessing/ Noise Reduction: Use a diffusion model (GenAI) to reduce background noise in the ATC audio.
2. Feature Extraction: Mel-Frequency Cepstral Coefficients (MFCCs): Extract MFCCs, which are a representation of the short-term power spectrum of the sound. These features are commonly used in speech recognition because they capture the important characteristics of the audio signal. Spectrograms: Generate spectrograms to visualize the frequency content of the audio signal over time. This helps in capturing the time-frequency representation of the audio, which is useful for the ASR model.
3. Denoising Architecture Integration: Denoising Autoencoder: Integrate a denoising autoencoder within the ASR model to enhance the audio quality before transcription. Multi-Objective Learning: Use a multi-objective loss function that combines denoising and ASR objectives to optimize both tasks simultaneously.
4. ASR Model: Recurrent Neural Network (RNN) [18]: Use an RNN-based ASR model, such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU), to transcribe the denoised audio. Attention Mechanism: Incorporate an attention mechanism to focus on important parts of the audio signal.

8. Annotating NOTAMs with a Tag System using GenAI

Summary: NOTAMs are documents used to convey abnormal situations to pilots in a concise and efficient manner. There exist studies that have studied adding context to NOTAMs through LLMs [31,32]. The model used by Szeto and Das [32], based on BERT, achieves a high accuracy in a classification task of 99%. Nonetheless, this approach uses only 5 different categories which are obtained by grouping 13 separate categories, thus meaning that the problem is simplified considerably.

Added value: due to increasing air traffic, the amount of delivered NOTAMs is also bound to increase. Overlooked or misunderstood NOTAMs could potentially result in dangerous situations that should be avoided. Adding context to the NOTAM through annotations has the potential to minimise these types of events. Previous studies propose a classification task with mutually exclusive classes. We instead propose the use of a tagging system where the classes are non-mutually exclusive. We think that rather than using a classification system a tagging system could be more effective, allowing a NOTAM to belong to multiple categories. Hence, when a NOTAM could be perceived as belonging to two categories the model is not forced to choose in which one to add it.

Proposed GenAI algorithm: Pre-trained BERT model fine-tuned for multi-label classification tasks

9. Large Language Models to aid Air Traffic Managers in Strategic Decisions

Summary: Specialise a LLM using ATM and aviation information, including text books, regulations, reports, etc. The goal would be to aid strategic decision making for Air Traffic Flow Management by passively responding to queries from a user, similarly to ChatGPT or DeepSeek.

Added Value: by retraining a LLM on a large amount of documents related to ATM and aviation, the model will become more specialised on this field, becoming knowledgeable on acronyms, procedure and regulations. Hence, it should be able to save time for traffic managers, who would otherwise have to search through a large number of documents to find answers for doubts regarding regulations or how similar problems were solved in the past.

Proposed GenAI Algorithm: Retrieval Augmented Generation (RAG) [33].

10. Virtual Reality with AI Generated Video

Summary: a potential use of virtual reality in aviation is drone control and visualisation. An issue that exists is that with a growing number of users coupled with higher definition images and other signals such as smell, noise, etc., the use of the available bandwidth will have to be more efficient, requiring a reduction in the bitrate of the transmissions.

Added value: traditional media casting already uses an encoding method [34], the additional value from this method is the first step in which a lower quality version of the images is obtained. Eliminating unnecessary details will mean that the encoding can be smaller, thus achieving a lower required bitrate. One could also achieve this by creating an encoder which can separate its latent space representation into two: one for the shapes of objects and another section for the details in objects.

Proposed GenAI algorithm: semantic communication data pipeline proposed by Sehad *et al.* [35]. A large Language Model generates a description of an image which is then used to regenerate the image.

2.4.1.2 Selection of GenAI use case

Based on the 4 criteria mentioned in Section 2.3.1, we reduced the options down to use cases 1,6,8 and 9. It was finally decided to proceed with **use case 8 "Annotating NOTAMs with a Tag System using GenAI"**. Although use case 9 was very highly rated, it was decided that as a research exercise it would not provide as much value as the other options. Use cases 1 and 6 were identified as candidates to explore in larger projects with a larger budget to purchase data and a longer time frame to develop and test algorithms.

2.4.2 Phase 2- Proof of concept

In this section, we present the results of the proof of concept for the selected use case, which was developed during the second phase of the project. Specifically, we summarize the performance of the supervised model and the MixMatchNL BERT model on the test dataset. Table 2 below provides a comparative overview of both models using a selected subset of evaluation metrics [6].

- Precision: $TP\ /\ (TP\ +\ FP)$
- Recall: $TP\ /\ (TP\ +\ FN)$
- F1-score: $2\ \times Precision\ \times Recall\ /\ (Precision\ +\ Recall)$
- Hamming loss: fraction of all labels that are predicted incorrectly
- Subset accuracy: fraction of NOTAMs with all labels correctly predicted

Where *TP* stands for true positive, *FP* for false positive and *FN* for false negative. In micro averaging the F1 score is computed for all labels simultaneously. In macro averaging, the F1 score is calculated for each label separately, and then the F1-score for each label is averaged.

Table 2. Comparison of evaluation metrics between the pre-trained BERT and the MixMatchNL BERT with Focal Loss on the test dataset [6].

Metric	Pre-trained BERT	MixMatchNL BERT
F1-score (Micro-Averaged)	0.93	0.96
F1-score (Macro-Averaged)	0.93	0.96
Hamming Loss	0.029	0.016
Subset Accuracy	0.78	0.88

The results demonstrate that combining semi-supervised learning (MixMatchNL) with focal loss significantly enhances classification performance compared to the baseline supervised BERT model. All evaluation metrics show consistent improvements, with particularly notable gains in F1-scores and a nearly 50% reduction in Hamming loss. Multi-label performance also increases substantially, reflected in a 10% improvement in subset accuracy.

Figure 3 presents the F1-score for each individual label, allowing us to assess whether the model faces greater difficulty in detecting specific NOTAM categories. Significant improvements are observed for labels such as Closure, Obstacle, and Taxiway, with the latter two achieving near-perfect F1-scores. Other categories, including Lighting issues, Navigation aid, and Runway, also exhibit notable gains.

Even for the more challenging classes, such as Procedure and Restriction, the retrained model achieves a better balance between precision and recall, resulting in higher overall F1-scores. Among all labels, Procedure remains the weakest in the retrained model, with an F1-score of 0.91. Although this still represents strong performance, the relatively lower value suggests residual difficulty in consistently identifying this category—likely due to higher linguistic variability within its associated text features.

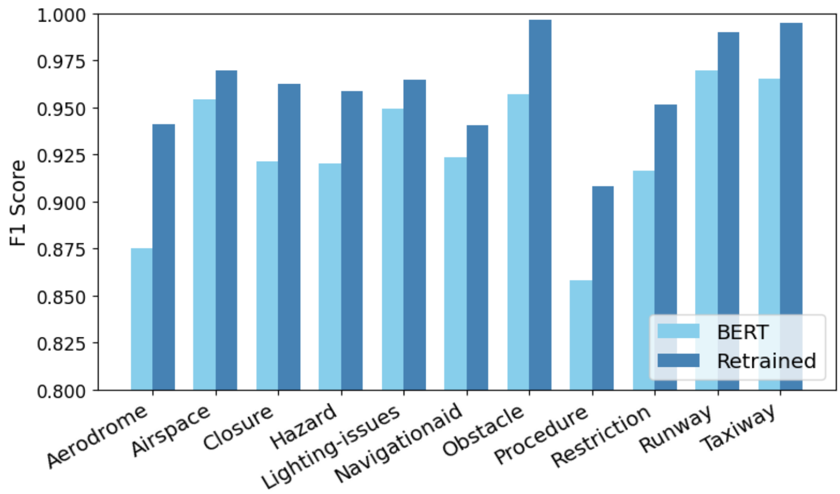

Figure 3. Comparison of F1-score per label between the pre-trained BERT and the MixMatchNL BERT with Focal Loss on the test dataset [6].

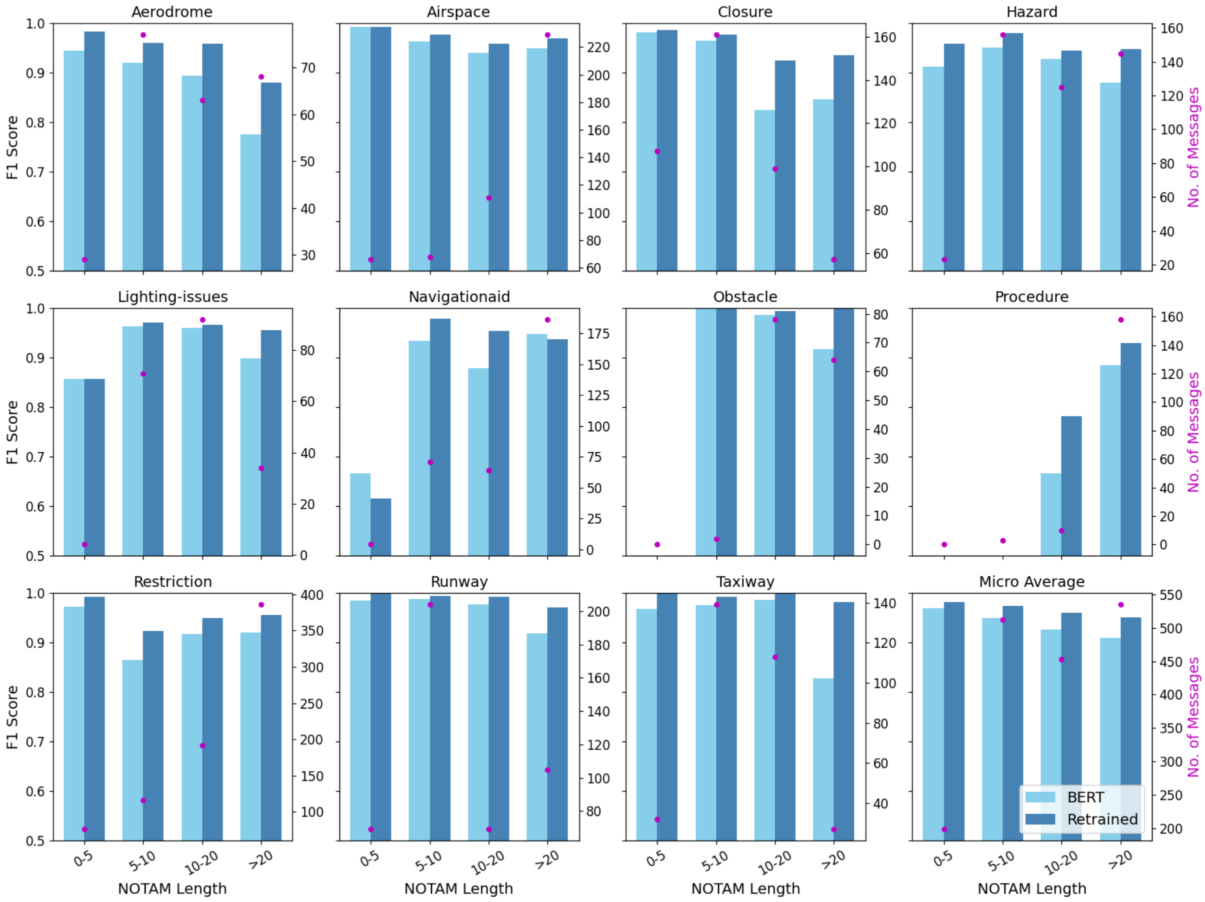

Figure 4. Per-label F1-scores for the BERT model and the retrained model using MixMatchNL and focal loss on the test dataset for different NOTAM lengths [6].

Figure 4 presents the F1-scores for each label across NOTAMs of different lengths. For labels with sufficient sample sizes (more than five NOTAMs), the retrained model consistently achieves F1-scores above 0.75, with the majority exceeding 0.9. In contrast, the baseline BERT model without MixMatch shows lower performance across all length categories.

The performance gap is particularly evident for longer NOTAMs, as shown in the bottom-right plot displaying micro-averaged F1-scores. While MixMatch retraining with focal loss provides only modest gains for shorter NOTAMs, its advantage becomes increasingly significant with message length. For NOTAMs exceeding 20 words, the F1-score rises from 0.91 to 0.96. Nevertheless, even after retraining, the model still exhibits a slight performance decline for the longest messages, indicating that message length remains a factor influencing classification accuracy.

3 Conclusions, next steps and lessons learned

3.1 Conclusions

Literature review conclusions

Our review shows that Generative AI (GenAI) is already being tested across a wide spectrum of ATM applications, ranging from traditional areas such as trajectory prediction to emerging roles like LLM-based assistant tools for air traffic operations. A detailed breakdown of references by application domain and GenAI methodology is provided in Deliverable D2.1 [1]. The main conclusions from the literature review are summarised below:

- Appropriateness and benefits: While GenAI is not always justified for simpler problems due to its limited transparency, it offers substantial advantages for tasks involving complex spatio-temporal dynamics, high-dimensional datasets (e.g. TMA trajectories), or natural language processing. State-of-the-art architectures such as self-attention mechanisms and autoencoders enable these models to detect intricate patterns that conventional approaches often miss.
- Trajectory prediction: This remains one of the most explored topics. However, most studies model trajectories in isolation, overlooking interactions between aircraft and controllers — a limitation that reduces real-world fidelity. Moreover, critical but infrequent events such as holding patterns or go-arounds are often excluded from training datasets, resulting in less robust models. Incorporating such edge cases is essential to capture the true operational variability of the ATM system.
- Contextual data integration: Across the literature, contextual information, including flight plans, weather conditions, and schedules, consistently proves crucial for achieving superior performance compared to traditional models such as regressions or Gaussian mixtures. The main challenge lies in accessing, harmonising, and fusing these heterogeneous datasets into a unified modelling framework. Progress toward more integrated airspace data and centralised data-sharing mechanisms would significantly enhance GenAI's impact in this domain.
- Transformer-based models: These models have shown particular strength in language-related tasks, where they clearly outperform conventional techniques. Although training transformers from scratch is computationally expensive, fine-tuning pre-trained models on small aviation-specific datasets has proven effective. Nevertheless, their current accuracy levels remain insufficient for safety-critical deployment without human oversight.
- Anomaly detection: In this area, unsupervised autoencoders have delivered mixed or mediocre results, especially in trajectory-related applications. Semi-supervised or supervised approaches tend to be more reliable and accurate, as they incorporate expert knowledge through small but valuable labelled datasets.

Overall, the literature indicates that GenAI is often applied to problems where the underlying complexity does not fully warrant its use, as the performance gains are sometimes marginal relative to the loss of interpretability. However, in domains involving highly complex, high-dimensional, or language-based data, GenAI has demonstrated clear value. Architectures leveraging self-attention or autoencoding mechanisms can capture intricate spatio-temporal relationships that traditional models fail to reproduce.

The literature review thus serves as a valuable reference for future research on GenAI in ATM, providing a comprehensive overview of existing applications and key sources. Its relevance was reinforced during the stakeholder workshop, where participants discussed additional use-case opportunities and rated the proposed ones positively — with average scores of 3.9/5 for relevance and 3.2/5 for GenAI's added value. The strong alignment among attendees suggests that the identified research directions are consistent with current industry needs and hold significant potential for advancing ATM innovation.

NOTAM annotation tool

We proposed a semi-supervised multi-label method for classifying NOTAMs using a BERT-based architecture. A portion of the dataset is manually labelled to fine-tune BERT, which then generates pseudo-labels for the unlabelled data. These are combined with labelled samples using the MixMatch algorithm and retrained with focal loss to address class imbalance. Results show significant improvements over the baseline: the micro F1-score rises from 0.93 to 0.96, Hamming loss is nearly halved, and even the weakest class ("procedure") improves. The model performs best on short NOTAMs but remains highly accurate (>0.95 F1) for longer texts. Adding a semi-supervised phase with MixMatch (α=0.25) and focal loss further boosts performance, particularly for rare or complex labels. Overall, this two-phase strategy enhances accuracy, robustness, and scalability for NOTAM classification.

For future improvements, the labelling system should be refined. Currently, the labels mix different types of information—some refer to infrastructure (e.g., runway, taxiway), while others describe the issue (e.g., hazard, restriction). Although the current model performs well, expanding the label set to improve descriptiveness could make the output too chaotic and unstructured. To address this, future work should explore hierarchical text classification by organizing labels into structured subsets and designing a model (or an ensemble of models) capable of handling these groups within a unified framework. This would help constrain outputs to logical combinations, simplifying both training and expert annotation.

The model itself could be further enhanced by testing stronger transformer architectures like RoBERTa or Longformer, which are better suited for longer sequences. MixUp could also be improved by using adaptive rather than fixed interpolation strategies. Domain-specific pretraining on aviation data could yield more meaningful embeddings, and integrating active learning could help prioritize the most informative samples for manual labelling, reducing effort while improving performance.

3.2 Next steps

As immediate steps to follow up on the development of the NOTAM annotation model we have the following:

- We have submitted a paper titled "A semi-supervised approach to multi-label classification of NOTAMs using BERT" on the work we completed on the second half of the project. We will present it at the SESAR Innovation Days (SIDs) if it is accepted.

- We will continue contacting experts and stakeholders who may be interested in the solution we developed to gather feedback on the current solution.

Further in the future we are planning on completing the following steps:

- We plan to improve the NOTAM annotation solution by including the recommendations that we outline at the end of Section 3.1.
- Once these improvements have been made, tested and validated we want to encapsulate the model in a tool with a visual interface, with the aim of reaching a high TRL (5-7).
- For this we plan to participate in future SESAR Calls for either exploratory or industrial research depending on the TRL level of the tool at the time that the proposal is submitted.
- Furthermore, we will also look for funding opportunities for the tool at a national level.

Regarding the literature review and use case catalogue that we made in the first part of the project, we have planned the following next steps:

- We will be using the use cases that are outlined in the catalogue, or parts of them, as part of future SESAR project proposals.
- The conclusions and use cases related to trajectory generation that were extracted from the literature review have already been included as part of the GAIA-ATM project proposal which was submitted on the SESAR 3 Exploratory Research 3 Call. In case it is accepted we will implement some of the concepts that are outlined in outer literature review for improving the state-of-the-art of trajectory generation.

3.3 Lessons learned

- In retrospect, we think for a project with such a short time frame it would be best to limit the scope of the project more at the beginning, with a clear use case already in mind. This would allow for the development of the tool to a much further level, and the inclusion of stakeholder feedback on the specific tool early on in the development.
- We found that data limitations were too often a limitation we found when selecting a use case, and is something we should have considered earlier in the project when defining the use cases rather than afterwards.
- For a 1 year project it would be best to already have the data needed for the project from the beginning.
- We think it could be very useful to have an environment to contact the mentors in a more casual manner, such as Microsoft Teams, as it would make communication quicker.
- Finally, we think the group of people invited to the workshop should have been better targeted, and it would have been best to hold the workshop in person, even if it meant holding it later. The participation in the online workshop was quite limited despite a large number of people signing up to it.
- Furthermore, we think that for areas where there is commercial competition occurring, such as NOTAM processing, it is best to engage experts separately as they are not open to sharing insights when there are competitors in the workshop. We found the bilateral meetings that we held during the project very productive.

- We do think that the 1 year time frame is sufficient to explore and open questions, and that having projects with the purpose of finding research directions that can be explored in larger projects could be beneficial.
- The figure of the mentors was quite positive for us, as they provided clear feedback as well as a contrast of different opinions that helped us in making decisions.
- Spending time in defining the structure of the literature review was quite helpful in ensuring the quality of the final product was high, and in making the process of collecting literature and reviewing it efficient.

4 Dissemination

Table 3 shows the different dissemination actions carried out during the project.

Table 3. Dissemination actions

Dissemination action	Description
Workshop	A project workshop was organised to collect feedback on the catalogue of identified GenAI use cases in ATM. Participants were asked to comment on aspects such as missing topics, potential technical challenges (e.g. computational requirements, data availability), relevance and added value for the ATM sector, and possible risks of GenAI implementation. Based on this feedback, the most promising use cases were refined, and one was selected for further development. The presentation slides and workshop minutes (including results) were shared with participants.
Bilateral meetings	Throughout the project, several bilateral meetings were held with stakeholders and experts in the ATM and AI domains. Meetings with Lufthansa Systems were key to validate the relevance of the proposed use cases from an operational perspective, refine the NOTAM labelling scheme, and confirm the practical value of multi-label classification for airline and operational workflows. In addition, a final exchange with EUROCONTROL (@Data Science meeting) took place. During this meeting, the proposed semi-supervised BERT-based approach for NOTAM annotation was presented and discussed. Feedback focused on the scalability of the solution, the importance of handling label imbalance and multi-topic NOTAMs, and the potential of semi-supervised learning to reduce annotation costs. This exchange further confirmed the relevance of the approach for future data-driven ATM applications and research activities within EUROCONTROL.
Internal deliverables	Deliverable “D2.1 Identification of potential GenAI applications in ATM and assessment of GenAI techniques” was completed and shared publicly on Zenodo. This document serves as a key reference for researchers seeking to explore GenAI applications in the ATM domain. Deliverables “D3.1 GenAI-ATM model(s): case studies” and “D4.1 Dissemination and exploitation report” are not ready for public release yet, but will be uploaded to a repository in Zenodo.

Dissemination action	Description
Results	The datasets and results developed for the NOTAM classification use case were published on Zenodo. Since the team created a custom labelling system to demonstrate the framework's potential, both the raw data and the experiment outcomes were made openly accessible to facilitate reuse and further research.
Scientific papers	A paper entitled "A semi-supervised approach to multi-label classification of NOTAMs using BERT" was submitted to SIDs 2025 (pending acceptance). The paper presents the work conducted during the second half of the project. It has also been made publicly available through Zenodo to ensure open access.

5 References

5.1 Project outputs

[1] A. Montoya Santamaría, O. Rey Orozco, D. Mocholí González, L. M. Viso Domínguez, and O. García Cantú-Ros, *D2.1 Identification of potential GenAI applications in ATM and assessment of GenAI techniques*, Zenodo, Oct. 1, 2025. DOI: 10.5281/zenodo.17302024.

[2] D. Mocholí González, 20250305 GenAI models for ATM workshop minutes, 2025.

[3] A. Montoya Santamaría 20250305 GenAI models for ATM workshop presentation, 2025.

[4] A. Montoya Santamaría, O. Rey Orozco, D. Mocholí González, L. M. Viso Domínguez, and O. García Cantú-Ros, *D4.1 Dissemination and exploitation report*, Zenodo, Oct. 1, 2025. Available at: https://zenodo.org/communities/sesarengage2notambert/records?q=&l=list&p=1&s=10&sort=newest.

[5] A. Montoya Santamaría, O. Rey Orozco, D. Mocholí González, L. M. Viso Domínguez, and O. García Cantú-Ros, *D3.1 GenAI-ATM model(s): case studies*, Zenodo, Oct. 1, 2025. Available at: https://zenodo.org/communities/sesarengage2notambert/records?q=&l=list&p=1&s=10&sort=newest.

[6] A. Montoya Santamaría, O. Rey Orozco, D. Mocholí González, L. M. Viso Domínguez, and O. García Cantú-Ros, *A semi-supervised approach to multi-label classification of NOTAMs using BERT*, Zenodo, Oct. 2, 2025. DOI: 10.5281/zenodo.17251034.

[7] A. Montoya Santamaría, O. Rey Orozco, *Datasets and results presented in the paper "A semi-supervised approach to multi-label classification of NOTAMs using BERT"*, Zenodo, Sep. 26, 2025. DOI: 10.5281/zenodo.17208970.

5.2 Other

[8] R.Lyons, Complexity analysis of the next gen air traffic management system: trajectory based operations. Work, 41(Supplement 1), 4514-4522, 2012.

[9] "Performance Review Report (PRR) 2023," EUROCONTROL, Jun. 10, 2024. [Online]. Available: https://www.eurocontrol.int/publication/performance-review-report-prr-2023.

[10] E. C. Pinto Neto, D. M. Baum, J. R. D. Almeida Jr, J. B. Camargo Jr, and P. S. Cugnasca, "Deep learning in air traffic management (ATM): a survey on applications, opportunities, and open challenges," *Aerospace*, vol. 10, no. 4, art. 358, 2023.

[11] "SynthAIr: Transforming ATM Through AI-Powered Synthetic Data Generation," SynthAIr, [Online]. Available: https://synthair.github.io/. Accessed: Oct. 2, 2025.

[12]I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets. *Advances in neural information processing systems*, *27,* 2014.

[13]Y. Lecun, S. Chopra, and R. Hadsell, "A tutorial on energy-based learning," 2006. *(No journal/conference provided).*

[14]D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint* arXiv:1312.6114, 2013.

[15]C. F. Higham, D. J. Higham, and P. Grindrod, "Diffusion models for generative artificial intelligence: An introduction for applied mathematicians," *arXiv preprint* arXiv:2312.14977, 2023.

[16]A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[17]H. Zhang and Z. Liu, "Four-Dimensional Aircraft Trajectory Prediction Based on Generative Deep Learning," *Journal of Aerospace Information Systems*, vol. 21, no. 7, pp. 554–567, 2024.

[18]S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[19]C. J. Peng, Y. J. Chan, C. Yu, S. S. Wang, Y. Tsao, and T. S. Chi, "Attention-based multi-task learning for speech-enhancement and speaker-identification in multi-speaker dialogue scenario," in *2021 IEEE Int. Symp. Circuits and Systems (ISCAS)*, May 2021, pp. 1–5, IEEE.

[20]L. Liu and R. C. Chen, "A novel passenger flow prediction model using deep learning methods," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 74–91, 2017.

[21]Z. Islam, M. Abdel-Aty, Q. Cai, and J. Yuan, "Crash data augmentation using variational autoencoder," *Accident Analysis & Prevention*, vol. 151, art. 105950, 2021.

[22]T. Krauth, A. Lafage, J. Morio, X. Olive, and M. Waltert, "Deep generative modelling of aircraft trajectories in terminal maneuvering areas," *Machine Learning with Applications*, vol. 11, art. 100446, 2023.

[23]T. Krauth, B. Figuet, X. Olive, and J. Morio, "Collision risk assessment in terminal manoeuvring areas based on trajectory generation methods," in *Proc. 15th Air Traffic Management Research and Development Seminar*, Savannah, USA, Jun. 2023.

[24]D.-T. Pham, T.-N. Tran, S. Alam, and V. N. Duong, "A generative adversarial imitation learning approach for realistic aircraft taxi-speed modeling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2509–2522, 2022.

[25]B. Kodanda and A. Verma, "Estimation of Origin-Destination air traffic flows from air transport movement data: An application to Indian domestic airport network system," in *World Conf. Air Transport Research Society (ATRS 2011)*, Sydney, Australia, Jun. 2011.

[26]B. Huang *et al.*, "ODformer: spatial–temporal transformers for long sequence Origin–Destination matrix forecasting against cross application scenario," *Expert Systems with Applications*, vol. 222, art. 119835, 2023.

[27]V. Sudakov, "Improving air transportation by using the fuzzy origin–destination matrix," *Mathematics*, vol. 9, no. 11, art. 1236, 2021.

[28]C. Li, L. Zheng, and N. Jia, "Network-wide ride-sourcing passenger demand origin-destination matrix prediction with a generative adversarial network," *Transportmetrica A: Transport Science*, vol. 20, no. 1, 2022. DOI: 10.1080/23249935.2022.2109774.

[29]X. Yu, D. Guo, J. Zhang, and Y. Lin, "ROSE: A Recognition-Oriented Speech Enhancement Framework in Air Traffic Control Using Multi-Objective Learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[30]J. van Doorn, J. Sun, J. M. Hoekstra, P. Jonk, and V. de Vries, "Whisper-ATC: Open Models for Air Traffic Control Automatic Speech Recognition with Accuracy," in *Int. Conf. Res. Air Transp. (ICRAT 2024)*, 2024.

[31]S. S. Clarke *et al.*, "Natural Language Processing Analysis of Notices To Airmen for Air Traffic Management Optimization," in *AIAA Aviation Forum*, Aug. 2021.

[32]A. Szeto and A. N. Das, "Classification of Notices to Airmen using Natural Language Processing," in *AIAA SCITECH 2024 Forum*, 2024, p. 2585.

[33]P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W-T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.

[34]S. Ahn, H. J. Yim, Y. Lee, and S. I. Park, "Dynamic and Super-Personalized Media Ecosystem Driven by GenAI: Unpredictable Plays Never Repeating the Same," *IEEE Transactions on Broadcasting*, 2024.

[35]N. Sehad, L. Bariah, W. Hamidouche, H. Hellaoui, R. Jantti, and M. Debbah, "GenAI for immersive communication: the next frontier in internet-of-senses through 6G," *IEEE Communications Magazine*, 2024.

[36]I. Rosa, "Leveraging machine learning and language models to assess the importance of notices to air missions (notams)," Master's thesis, NTNU, 2024.

[37]H. Zou and Z. Wang, "A semi-supervised short text sentiment classification method based on improved bert model from unlabelled data," Journal of Big Data, vol. 10, no. 1, p. 35, 2023.

[38] J. Coupon, "Notam-classifier (capstone project)," https://github.com/jcoupon/NOTAM-classifier, 2020, accessed: 2025-08-29.

6 List of acronyms

Acronym	Description
AI	Artificial Intelligence
ASR	Automatic Speech Recognition
ATC	Air Traffic Control
ATCO	Air Traffic Controller
ATM	Air Traffic Management
BERT	Bi-directional Encoder Representations from Transformers
CTGAN	Conditional Tabular Generative Adversarial Network
CWGAN	Conditional Wasserstein Generative Adversarial Network
DES	Digital European Sky
DNN	Deep Neural Network
E-AMAN	Extended Arrival Manager
EBM	Energy Based Model
FAA	Federal Aviation Administration
GAIL	Generative Adversarial Imitation Learning
GAN	Generative Adversarial Network
GenAI	Generative Artificial Intelligence
GRU	Gated Recurrent Unit
ICAO	International Civil Aviation Organization
LLM	Large Language Model
LoS	Loss of Separation
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficient
NOTAM	Notice to Air Missions
RAG	Retrieval Augmented Generation
RNN	Recurrent Neural Network
ROSE	Recognition-Oriented Speech Enhancement
SAE	Stacked Autoencoder
SESAR	Single European Sky ATM Research Programme

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Acronym	Description
SWIM	System Wide Information Management
TBO	Trajectory Based Operations
TMA	Terminal Manoeuvring Area
TRL	Technology Readiness Level
VAE	Variational Autoencoder
WER	Word Error Rate
XAI	eXplainable Artificial Intelligence