SynthAIr D1.1 - Concept Outline

Deliverable ID:	**D1.1**
Project acronym:	**SyntAIr**
Grant:	**101114847**
Call:	**HORIZON-SESAR- 2022-DES-ER-01-WA1-7**
Topic:	**Fundamental Science and Outreach for Artificial Intelligence (AI) for Aviation**
Consortium coordinator:	**SINTEF**
Edition date:	**23 January 2024**
Edition:	**01.01**
Status:	**Official**
Classification:	**PU**

EUROPEAN PARTNERSHIP
Co-funded by the European Union

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Abstract

This deliverable describes the concept outline of SynthAIr project. The main objective of SynthAIr is to explore and define AI-based methods for synthetic data generation in the domain of ATM system due to the limitation of AI-based tools development by the lack of enough data available (e.g., safety-related data) and the problem of generalization of those AI-based models. The project explores data-driven methods for synthetic data generation, since they require 1) less user knowledge expertise (no need to derive the explicit model of the distribution), 2) better generalization capabilities. More in detail, inspired by recent advancement in Computer vision and Language Technology, SynthAIr proposes the concept of Universal Time Series Generator (UTG). A UTG, is a model trained on several different time series, and able to generate a synthetic dataset representing a new dataset, simply conditioned by a compressed representation of it. In aviation domain, this generator can be trained on a certain set of data related, for example to few airports, and be used to generate synthetic data from a new airport. The same principle can be applied to define a Universal Time Series Forecaster (UTF) able to do prediction to a new environment (I.e., data from a new airport) without any new training.

Authoring & approval

Author(s) of the document

Organisation name	Date
SINTEF	16-11-2023
SINTEF	09-01-2024

Reviewed by

Organisation name	Date
DEEP BLUE	08-12-2023
EUROCONTROL	08-12-2023
DEEP BLUE	17-01-2024

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

Organisation name	Date
SINTEF	08-12-2023
SINTEF	18-01-2024

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

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Rejected by[2]

Organisation name	Date
SJU	22-12-23

Document history

Edition	Date	Status	Organisation author	Justification
00.00.01	16-11-23	Draft	SINTEF	Proposed table of content
00.00.02	20-11-23	Draft	SINTEF	Sections drafted
00.00.03	21-11-23	Internal review	SINTEF	Internal peer review #1
00.00.04	21-11-23	Quality check	SINTEF	First quality check
00.00.05	28-11-23	Post 1st review draft	SINTEF, TUD, DBL	Addressed minor changes from internal peer review #1 and first quality check
00.00.06	29-11-23	Internal review	SINTEF	Document reviewed and harmonised.
00.00.07	30-11-23	Internal review	SINTEF	Internal peer review #2
00.00.08	30-11-23	Quality check	SINTEF	Second quality check
00.00.09	01-12-2023	Post 2nd review draft	SINTEF, TUD, DBL	Addressed minor changes from internal peer review #2 and second quality check
00.00.10	06-12-2023	Consolidated version	SINTEF, DBL, ECTRL	Document reviewed and harmonised. Ready for PC approval
00.01.00	08-12-2023	Final release	SINTEF	Deliverable approved for submission to the SJU
00.01.01	17-01-2024	Updated version	SINTEF, TUD, DBL	Addressed SJU comments

[2] Representatives of the beneficiaries involved in the project

00.01.02	18-01-2024	Internal review	SINTEF	Internal peer review #3
00.01.03	18-01-2024	Quality check	SINTEF	Third quality check
00.01.04	18-01-2024	Consolidated version	SINTEF, DBL, ECTRL	Document reviewed and harmonised. Ready for PC approval
01.01.00	23-01-2024	Resubmission - Final release	SINTEF	Deliverable approved for submission to the SJU

Copyright statement

Disclaimer

SynthAIr

Improved ATM automation and simulation through AI-based universal models for synthetic data generation

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114847 under European Union's Horizon Europe research and innovation programme.

Table of contents

List of figures

List of tables

1 Executive summary

This chapter serves as a comprehensive summary of the Concept Outline document, focusing on the key elements evaluated by the "*Improved ATM automation and simulation through AI-based universal models for synthetic data generation*", namely SynthAIr project.

It addresses the pressing challenge of **data scarcity in Air Traffic Management** (ATM) through the **innovative application of AI-driven synthetic data generation**. This approach aims to revolutionise the aviation sector by overcoming the limitations posed by the complex and multidimensional nature of ATM data, leveraging recent advancements in Generative Adversarial Networks (GANs) and recurrent neural networks (RNNs). The project anticipates substantial benefits for ATM if further developed. These include *enhanced predictive analytics capabilities, improved operational efficiency, reduced flight delays, optimised resource allocation, and an overall transformation of the aviation industry through increased automation*.

SynthAIr's focus extends to airport terminal and airside operations, as well as ATM operations. The project is designed to address challenges in these specific operational contexts within the broader aviation ecosystem.

The viability of SynthAIr's concept is substantiated by a **robust capturing exercise, expert group insights, and workshop** reports that collectively provide evidence of the benefits and operational viability of AI-driven synthetic data generation in the ATM domain.

While currently at a low Technology Readiness Level (TRL), SynthAIr does not directly contribute to SESAR solutions. However, the project presents **potential avenues for future application-oriented exploratory research activities**. Key focus areas include the development of adaptable machine learning models, predictive analytics for flight delays and passenger flows, and the creation of aviation-specific datasets.

The next Research and Innovation (R&I) phase will involve *generating interest within SESAR and scientific communities, disseminating information to key stakeholders, identifying synergies with other SESAR projects, and laying the groundwork for future development*. Ongoing research and innovation efforts will be guided by **a strategic plan to explore the viability of SynthAIr's concept and its potential impact on the ATM system**.

2 Introduction

2.1 Purpose of the document

The purpose of this document is to elucidate the scope and objectives of the SynthAIr within the context of the HEU SESAR initiative. Each SESAR excellent science and outreach exploratory research project aiming at completing TRL1 shall produce one concept outline document. This document serves as a foundational element, outlining the project's objectives, methodologies, and innovative approaches, thus facilitating a comprehensive understanding of the proposed research within the context of advancing the SESAR initiative.

SynthAIr aims to address the challenges associated with the development of AI-based tools in the Air Traffic Management (ATM) system, specifically arising from the scarcity of certain types of data, such as safety-related information, and the difficulty in generalizing AI models effectively.

The primary focus of SynthAIr is the exploration and definition of AI-based methods for synthetic data generation. This approach is chosen to overcome limitations posed by the insufficient availability of relevant data and to enhance the generalization capabilities of AI models. The project adopts data-driven techniques for synthetic data generation, leveraging their advantages, including reduced dependence on user expertise and improved generalization.

In a more detailed context, SynthAIr introduces the concept of the Universal Time Series Generator (UTG), inspired by recent advancements in Computer Vision and Language Technology. The UTG is a model trained on diverse time series datasets, capable of generating synthetic datasets for new scenarios based on a compressed representation of the target environment. Applied to aviation, the UTG can be trained on specific datasets related to certain airports, enabling it to generate synthetic data for a new airport. Similarly, the concept extends to a Universal Time Series Forecaster (UTF), which can predict outcomes in a novel environment (e.g., data from a new airport) without requiring additional training.

2.2 Intended readership

The primary stakeholders involved in the SynthAIr concept outline deliverable, including project partners, collaborating entities, and external readers, are encouraged to utilise this deliverable for various purposes.

- **Project Partners**: SynthAIr project partners are encouraged to leverage this concept outline deliverable as the foundational cornerstone of our research. This document serves as a comprehensive record, capturing the essence of the work to be conducted and the key outcomes attained by SynthAIr. The insights presented herein are intended to be a fundamental resource, providing valuable guidance and information for shaping and directing future initiatives pertaining to AI-based methods for synthetic data generation within the ATM system domain.
- **SESAR Joint Undertaking**: The SESAR Joint Undertaking is invited to incorporate the insights provided in this deliverable, along with those from other relevant projects in the field, to

formulate a consolidated set of conclusions and recommendations at a programmatic level for advancing AI-based tools in the ATM system.

- **External Readers**: External entities such as CANSO, airports, airlines, ATCOs associations and other HE projects in the field are invited to consider the content of this SynthAIr concept outline deliverable to support their collaborative activities related to AI-based methods within the ATM system domain. The insights derived from this deliverable can contribute to the ongoing dialogue and collaboration in the aviation domain, particularly in addressing challenges associated with AI-based tools development and synthetic data generation.
- **Public Dissemination**: The SynthAIr consortium welcomes the publication of this concept outline deliverable as a public document. This transparency initiative aims to share the findings and concepts developed by SynthAIr with any interested parties involved in the further evolution of AI-based methods within the ATM system domain. The overarching goal is to foster collaboration, innovation, and knowledge-sharing in the field.

By disseminating and utilising the information contained in this concept outline deliverable, the SynthAIr project aims to contribute significantly to the exploration and definition of AI-based methods for synthetic data generation within the ATM system domain, addressing critical challenges and paving the way for future advancements in the field.

2.3 Background

The SynthAIr project draws upon a rich background of previous exploratory research activities both *within and outside the SESAR framework*. Several key projects have contributed valuable insights and methodologies, forming the basis for the development and implementation of SynthAIr.

The following highlights the relevant contributions from other SESAR projects:

- **Artimation** - *Transparent Artificial Intelligence and Automation to Air Traffic Management Systems*: Artimation focused on studying the impact of explainability levels and various Human-AI interfaces on performance and safety. The knowledge generated in Artimation is instrumental for SynthAIr in defining the interaction between proposed AI tools and users, ensuring effective human-AI teaming.
- **MAHALO** - *Modern ATM via Human/Automation Learning Optimization*: MAHALO concentrated on transparency, seeking a balance between optimized AI solutions requiring explainability and conformal AI solutions akin to those familiar to air traffic controllers (ATCOs). SynthAIr benefits from the results of MAHALO in shaping how predictions are presented to users, determining the level of detail, and suggesting actions.
- **HAIKU -** *Human AI Teaming Knowledge and Understanding for Aviation Safety*: HAIKU, a HORIZON.2.5 project initiated in September 2022, delivers AI Digital Assistants for various aviation segments and users. The prototypes developed by HAIKU will contribute valuable insights to SynthAIr, particularly in the context of human-AI teaming for aviation safety.

Additionally, SynthAIr benefits from external projects outside the SESAR framework, contributing valuable insights and methodologies to enrich the project's foundation:

- **Smart Sensing for Aviation**: This national project in the Netherlands focused on improving passenger flow management in airports through sensor data on passenger movement. The

Co-funded by the European Union

knowledge and simulation tools developed in this project will be applied in SynthAIr for use case scenarios and model evaluations.

- **Airport Technology Lab**: Another national project in the Netherlands, the Airport Technology Lab, explored strategies for optimizing airport resource usage and enhancing passenger flows. Factors identified in this project influencing passenger flows will serve as input for one of the use cases in SynthAIr.
- **EUROCONTROL ATM Real-time Simulation**: The synthetic traffic generator developed in SynthAIr will be employed to accelerate and simplify the creation of real-time simulations at EUROCONTROL. This includes training universal policies for air traffic control and flow management using reinforcement learning techniques.
- **EATIN - EUROCONTROL Air Transport Innovation Network Program**: EATIN initiatives, including the curfew collaborative management project and ongoing cooperative projects with Prague and Geneva airports, address turn-around prediction. The results of these projects will contribute as input to one of the use cases in SynthAIr.
- **FADE - Forecast of ATFM Delay Evolution**: FADE, an initiative within the EATIN program, utilized machine learning models to predict flight delays. The models and results developed in FADE will assist SynthAIr in assessing the utility of synthetic data in supervised tasks.
- **RICO – Robust Intelligent Control**: A four-year project funded by the Norwegian Research Council, RICO focuses on robustness in AI, specifically in Robust Time Series Forecasting and Data-Driven Generative Models for Robust Control. Results from RICO directly support and enhance SynthAIr's capabilities.
- **ML4ITS – Machine Learning for Irregular Time Series**:Funded by the Research Council of Norway, ML4ITS addresses challenges in handling irregular time series data. The project's methodologies, including privacy-preserving synthetic data creation, directly contribute to enhancing SynthAIr's capabilities. A structured collaboration is established between the SynthAIr consortium and ML4ITS for a synergistic exchange of knowledge.

This collective body of work provides a solid foundation for SynthAIr, enabling the project to build upon existing knowledge and leverage proven methodologies from diverse initiatives within and beyond the SESAR framework.

2.4 Structure of the document

This document is organized into the following sections, each serving a specific purpose and contributing to a comprehensive understanding of the SynthAIr project:

Abstract: it provides a concise summary of the document, offering a quick overview of the key aspects of the SynthAIr project.

1. **Executive Summary**: it offers a high-level overview of the SynthAIr project, highlighting its objectives, key components, and anticipated outcomes.
2. **Introduction**: it sets the stage for the document by delving into its purpose, intended readership, background, and overall structure. It aims to clarify the objectives and scope, identifying the target audience and offering contextual information about the SynthAIr project. Additionally, this section outlines the document's organization, defines key terms for

clarity in the glossary of terms, and provides a reference for acronyms used throughout the document.

3. **Concept outline**: it centres on establishing the conceptual foundation of SynthAIr. It begins by articulating the problem statement, pinpointing the core issues addressed by the project. Subsequently, the concept description and operational scenarios provide an in-depth exploration of the technical context, stakeholders, key scenarios, and potential limitations. the expected performance outcome sheds light on the anticipated results, while the key assumptions articulate fundamental project hypotheses. together, these elements form a comprehensive understanding of SynthAr's conceptual framework.
4. **Proposed SESAR solutions**: it presents solutions aligned with SESAR initiatives.
5. **Plan for next R&I phase**: it outlines the research and innovation activities planned for the next phase of the SynthAIr project.
6. **References**: it lists applicable and referenced documents.
7. **Annex:** it includes additional information or supplementary material relevant to the project

2.5 Glossary of terms

Term	Definition	Source of the definition
Universal Model	Universal model refers often as “global” models. Global models in the context of time series analysis are models that learn from a large collection of time series data. The term "global" refers to the approach of considering multiple time series as a single dataset, rather than modelling each series individually. A Global Forecasting Model (GFM) is trained on this collective set of series and utilizes a shared set of parameters across all the series. This is in contrast to local models, where each time series would have its own distinct set of parameters.	Montero-Manso P, Hyndman RJ. Principles and algorithms for forecasting groups of time series: locality and globality. Int J Forecast 2021;37:1632–53. T. Januschowski, J. Gasthaus, Y. Wang, D. Salinas, V. Flunkert, M. Bohlke-Schneider, L. Callot Criteria for classifying forecasting methods Int. J. Forecast., 36 (1) (2020), pp. 167-177 Hewamalage, H., Bergmeir, C., & Bandara, K. (2022). Global models for time series forecasting: A Simulation study. Pattern Recogn., 124(C).

Universal Time Series Generator (UTG)	UTG is a model trained on several different time series, and able to generate a synthetic dataset representing a new dataset, simply conditioned by a compressed representation of it.	
Universal Time Series Forecaster (UTF)	Similarly to UTG, UTF is a machine learning model able to do prediction to a new environment (I.e., data from a new airport) without any new training.	

Table 1: Glossary of terms

2.6 List of acronyms

Term	Definition
ATM	Air traffic management
DES	Digital European Sky
GA	Grant agreement
GDPR	General data protection regulation
HE	Horizon Europe
ID	Identifier
SESAR	Single European sky ATM research
SESAR 3 JU	SESAR 3 Joint Undertaking
TRL	Technology readiness level
UTG	Universal Time Series Generator
UTF	Universal Time Series Forecaster

Table 2: List of acronyms

3 Concept outline

3.1 Problem statement

This section serves to articulate the core problem addressed by the SynthAIr project within the European aviation context. The dimensions of the problem and its impact on Air Traffic Management (ATM) are explored, aiming to answer critical questions related to the identification, quantification, and location of the ATM challenges.

The aviation industry, particularly in ATM, is currently undergoing a transformative phase due to increasing operational complexities and the urgent need for enhanced automation. As global airspace congestion rises, there is a pressing demand for sophisticated solutions to manage air traffic more efficiently and safely. Despite the potential of Artificial Intelligence (AI) in addressing these challenges, a significant barrier to its widespread adoption in the ATM domain is the scarcity of adequate training data.

In many advanced technological domains, the application of synthetic data has proven effective in overcoming data scarcity challenges. However, in the ATM domain, the potential of synthetic data remains largely untapped, primarily due to the complex nature of ATM data. Unlike other domains, ATM data involves timestamped records of mixed data types, including categorical, continuous, and datetime values. These records often span multiple tables, exhibit time series of unequal lengths, and demonstrate intricate multidimensional dependencies over extended durations.

State of the Art

The evolution of artificial intelligence and machine learning has been revolutionary, particularly in areas such as computer vision and natural language processing. However, when it comes to timestamped data, especially in aviation domains with rigorous data collection requirements, challenges become more pronounced. One primary hurdle is the scarcity of training data, with certain data types requiring specialized domain knowledge for accurate annotations. Timestamped data is critical for predictive analytics in the aviation industry, for example in the areas of flight delay prediction, flight diversion prediction, passenger flow prediction. Here are examples of timestamped data used in machine learning applications for these specific areas:

- Historical flight departure and arrival times.
- Timestamped data on aircraft turnaround activities, including unloading, refuelling, cleaning, and boarding.
- Log of air traffic control advisories and delays. Sequence of weather reports and forecasts for airports and en-route.
- Time-stamped weather updates.
- Historical diversion records, including times of decision-making and actual diversion events.
- Check-in and security checkpoint throughput data with timestamps.
- Boarding gate arrival times and boarding sequences.
- Historical data on flight schedules

- Detailed flight schedules with planned departure and arrival times.

Data sharing, a potential solution to data scarcity, faces challenges such as privacy constraints and data imbalance. The generation of synthetic data for time-series data is notably intricate due to inherent temporal patterns. Recent advancements in synthetic data generation have primarily relied on Generative Adversarial Networks (GANs), especially those employing recurrent neural networks (RNNs).

Objectives and Ambitions

In the evolving landscape of ATM, the challenges posed by increasing intricacies and the need for advanced automation are significant. The primary impediment is the scarcity of data. SynthAIr, as an initiative, aims to address this challenge by emphasizing AI-driven synthetic data generation tailored specifically for the ATM domain. The mission is not only to innovate but to revolutionize the aviation sector, ushering in a transformative phase of ATM automation.

SynthAIr will pursue the following objectives:

- [Obj.1] ENHANCE comprehension of AI-based synthetic data generation methods.
- [O.2] FORMULATE Universal AI techniques for generating authentic synthetic ATM data.
- [Obj.3] ASSESS privacy preservation, fidelity, and diversity in the produced synthetic data.
- [Obj.4] VERIFY the impact of generated synthetic data in operational scenarios, emphasizing generalization, robustness, and resilience.
- [Obj.5] PROVIDE recommendations, principles, and best practices based on experimental results and analyses.
- [Obj.6] EXPEDITE the integration of AI in the ATM system and ESTABLISH collaborative mechanisms.

To achieve these objectives, SynthAIr has formulated two central research questions related to the development of AI-based synthetic data generation methods and optimizing their use to enhance AI in the ATM system:

1. How to develop AI-based synthetic data generation methods for ATM domain?
    - These methods *should produce high fidelity, privacy preserving, and diverse synthetic data.*
2. How to best make use of synthetic data to improve AI-methods in the ATM system?
    - Synthetic method *should be evaluated with respect of increasing robustness, generalizability and performance*.

The diagram depicted below serves as a conceptual map for SynthAIr, providing a condensed overview of the challenges, principal outcomes, and the sequential steps integral to the project's attainment of its objectives.

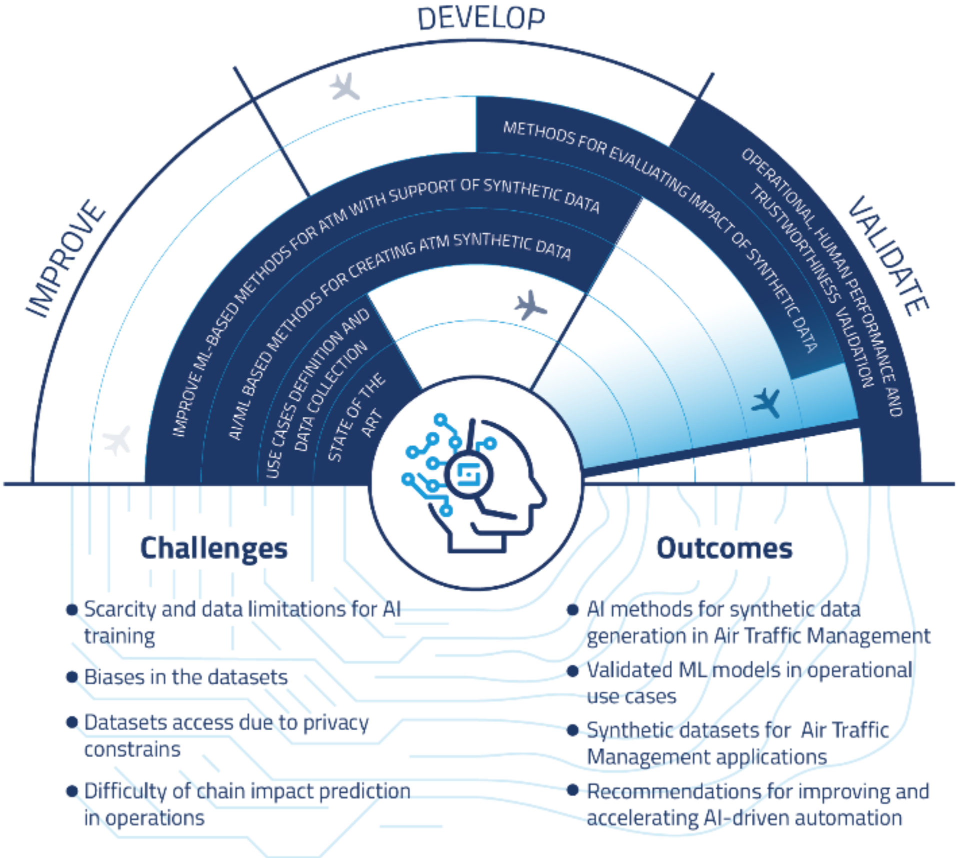

Figure 1: SynthAIr concept map

Related Challenges in the ATM Domain

Synthetic data generation in the ATM domain is complex due to the intricate nature of timestamped, multi-dimensional, and interdependent data sources. Indeed, constraints arise from the **limited availability of training data for AI models**. This limitation stems from factors such as **data accessibility**, **privacy** concerns, and **biases**. While synthetic data has been extensively utilised to overcome such challenges in domains like self-driving cars and autonomous drones, its potential benefits for the ATM domain remain largely untapped. Synthetic data offers the capability to digitally generate required datasets in any desired volume, tailored to precise specifications and needs. A 2021 Garner Study [1] predicts that**, by 2024, 60% of the data used in AI system development will be synthetic rather than real** (refer to Figure 1). This underscores the increasing significance of synthetic data in addressing data-related challenges in AI applications.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

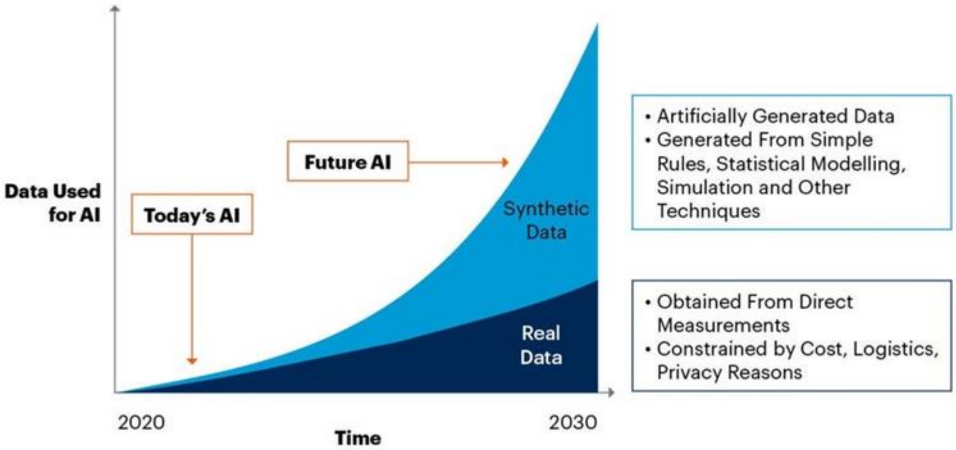

Figure 2: Data trends in artificial intelligence (Gartner)

To overcome these challenges is needed to develop quantitative metrics to evaluate the quality, adequacy, and realism of generated synthetic data, emphasizing the importance of these measures in relation to downstream tasks.

3.2 Concept description and operational scenarios

3.2.1 Operational/Technical context

The use cases considered in the project concern airport terminal and airside operations, as well as ATM operations. At this stage, it is still to be decided based on which particular use cases the project will focus for further elaboration (to be considered in deliverable 2.1).

3.2.2 Stakeholders

The SynthAIr project places a strong emphasis on stakeholder engagement to ensure the assessment, documentation, and alignment of results with industry needs and goals. The following key stakeholders have been identified at the proposal stage and already involved in the initial activities of the project (please refer to Sections 3.2.3 and 3.2.4):

- **SWISS Air and Ryanair**: As active participants in workshops, these airlines contribute real-world insights to validate SynthAIr's concepts, use cases, and final results. Their involvement ensures that the developed solutions align with industry needs and are applicable in practical scenarios.
- **Istanbul Airport, Prague Airport, Malpensa Airport**: These airports and airlines participate in workshops, contributing to the validation of SynthAIr's concepts, use cases, and final results. Their engagement ensures that the solutions are well-aligned with operational requirements and can be effectively integrated into diverse aviation settings.
- **ANACNA - Italian ATCOs Association**: As representatives of Italian Air Traffic Control Officers, ANACNA's participation in workshops ensures the validation of project concepts, use cases,

Co-funded by the European Union

and final results. Their expertise in air traffic management provides valuable feedback to ensure that the solutions meet the specific needs of ATCOs.

- **CANSO - Civil Air Navigation Services Organization**: As an industry association representing air navigation service providers, CANSO's participation in workshops contributes to validating project concepts and results. Their involvement aids in disseminating SynthAIr's findings across the global air traffic management community and among industry leaders.
- **HAIKU - Human AI Teaming Knowledge and Understanding for Aviation Safety**: This European research project actively participates in workshops, providing valuable insights, commenting on SynthAIr's achievements, and aligning their own work with SynthAIr's outcomes. Their feedback contributes to cross-project knowledge sharing and the identification of best practices.
- **TRUSTY - Trustworthy intelligent system for remote digital tower:** The overall goal is to enhance the trustworthiness of AI-powered decisions in the context of remote digital towers on specific tasks, such as runway and taxiway monitoring. The project will apply information visualisation techniques like visual analytics, data-driven storytelling, and immersive analytics in human-machine interactions (HMI). As an AI for aviation flagship project SynthAIr aims to interact with this project, possibly with topics related to UC2 Flight delay prediction and in terms of ML and methodologies used to deal with data .
- **ASTRA - AI-enabled tactical FMP hotspot prediction and resolution:** SESAR Exploratory research project with the aim to predict and resolve hotspots much earlier than today using an AI-enabled tool. The overall aim is to optimise capacity while enabling aircraft to adhere to more efficient and green trajectories. As an AI for aviation project SynthAIr aims to interact with this project, possibly with topics related to UC5 Flight diversion prediction.
- **JARVIS - Just a rather very intelligent system**: The project will develop three AI based solutions, an Airborne digital assistant to support crew and single pilot operations, an ATC digital assistant to support more efficient and green tower operations; and an Airport digital assistant will increase the level of automation in airports, enhancing safety and security for intrusion detection scenarios. As an AI for aviation project SynthAIr aims to interact and have synergies with this project, especially with the airport digital assistant team, due to topics related to UC1 Turnaround time prediction and UC3 Passenger Flow Prediction.

Additionally, other stakeholders with complementary expertise have been identified and formally invited to join the current advisory board. Invitations were extended through an email and formal invitation letter (please refer to Annex 7.1) to ensure a diverse and knowledgeable group that aligns with the goals and objectives of the SynthAIr project. Through this collaborative engagement with stakeholders, SynthAIr establishes a robust feedback loop, ensuring that the developed solutions are not only relevant and effective but also capable of addressing the complex challenges in the aviation domain. This approach is designed to accelerate the adoption of AI in air traffic management, foster knowledge exchange, and promote innovation within the aviation industry. The project's commitment to stakeholder involvement ensures that their feedback is documented in project deliverables, and their demonstrated interest in the proposed concepts and technologies further validates the project's objectives.

3.2.3 Key scenarios

This sub-section is dedicated to outlining the key relevant scenarios, providing a comprehensive understanding of how the SynthAIr project's concept, technology, capability, or research element will function within the specified operational and technical context described in section 3.2.1. It elucidates the involvement of stakeholders, as detailed in sub-section3.2.2, and explicates the interactions between them. The scenarios, which are detailed within the use cases, are crafted to capture a wide array of standard and challenging environments. This approach anticipates the formulation of design parameters for future Research and Innovation (R&I) activities at higher Technology Readiness Levels (TRLs), ensuring that selections of Use Cases (UCs) or scenarios are systematically based on criteria outlined in a defined matrix (Table 4). The matrix will guide the selection process, ensuring that each scenario is relevant and contributes meaningfully to the project's objectives. The impact on key performance areas and the estimated performance contributions of these scenarios are also meticulously considered, providing a clear path from current research to practical application.

3.2.3.1 UC.1 Turnaround Time Prediction

The initial operational use case focuses on the precise prediction of turnaround times at airports. Turnaround time, a critical metric influenced by various factors, significantly impacts flight delays. Recognizing this, the European ATM Master Plan underscores the need for early warning indicators during the turnaround process. This use case aims to develop models capable of generating synthetic data for the turnaround process, improving the accuracy of predictions. By modelling stochastic processes involved in diverse turnaround steps, the goal is to create a universal model applicable to various scopes and granularities across different airport categories and turnaround processes.

3.2.3.2 UC.2 Flight Delay Prediction

The second use case centres on predicting flight delays, a pivotal indicator in the aviation industry. A robust flight delay prediction model is crucial for mitigating congestion and disruptions. This study seeks to augment existing models by incorporating synthetic data to improve prediction accuracy, especially in scenarios with limited real data. The focus is on identifying and addressing the most uncertain types of delays, with a particular emphasis on the challenging task of estimating the target off-block time (TOBT).

3.2.3.3 UC.3 Passenger Flow Prediction

The third use case addresses passenger flow prediction as a means of managing uncertainties caused by passengers in air traffic management. Over 50% of flight delays are attributed to passenger-related issues. The study aims to enhance predictive models for passenger flows within airport terminals by leveraging synthetic data augmentation. The goal is to create a universal model applicable across various scenarios with limited data availability, optimizing resource allocation, reducing waiting times, and enhancing passenger experiences.

3.2.3.4 UC.4 Synthetic Traffic Generator for Fast and Real-Time Simulations

The fourth use case focuses on developing a synthetic traffic generator to enhance real-time and fast-time simulations. Real-time simulations often suffer from limited variability, predictability, and bias in results. This use case aims to counteract these effects by generating representative and varied traffic conditions. For fast-time simulations, a synthetic traffic generator could provide the necessary

Co-funded by the European Union

variability for robust policy generalization, thereby enhancing decision-making and policy development.

3.2.3.5 UC.5 Flight Diversion Prediction

The fifth use case homes in on predicting flight diversions, a crucial operational concern influenced by adverse weather conditions and unforeseen events. With limited datasets available due to the infrequency of diversions, the study aims to improve the prediction model by incorporating synthetic positive observations. Through synthetic data augmentation, this use case seeks to enhance the training dataset, leading to more accurate predictions of flight diversions and improved resource allocation and operational planning.

Collectively, these operational use cases contribute to SynthAIr's overarching goals, exploring the applicability of synthetic data generation across various facets of the ATM domain. The use cases address challenges related to prediction accuracy, simulation improvement, and operational planning, aligning with the project's mission to enhance AI-driven automation and decision support in the aviation industry.

3.2.4 Potential limitations, weaknesses and constraints

In the SynthAIr project's research and innovation (R&I) phase, we have defined a set of criteria to analyse the Use Cases (UCs) based on their relevance to the project's goals and potential impact. This sub-section will discuss the framework used to assess UCs, considering the inherent limitations, weaknesses, and constraints identified during the R&I phase.

The criteria, as detailed in Table 3, guide us in selecting UCs by their significance to the project's objectives and the severity of data scarcity they address. Our methodology among other criteria, focuses on prioritizing UCs where synthetic data can play a transformative role, particularly in areas where the lack of data hinders advancement. This strategic approach ensures that our efforts are concentrated on UCs poised to benefit most from synthetic data solutions.

To navigate these challenges, we've established a selective process that will determine the UCs we proceed with. This process is data-driven, relying heavily on the insights gained from our state-of-the-art review (please refer to D2.1 for more datils), comprehensive 1:1 interviews, and feedback from Advisory Board (AB) meetings, especially the initial one which is scheduled for February 16th from 9:30 to 13:00 CET.

We have conducted a series of interviews with relevant stakeholders, and the table below provides an overview of these interactions.

Who	SESAR project	When	Scope	Additional notes
Westminster university	MultimodX	12-12-23	• Explore potential **synergies** between SynthAIr and MultimodX • **Strengthen collaboration** with exploratory projects & universities.	

Athens International Airport	JARVIS	04-01-24	• Collect perspectives on use cases from the **airport** standpoint • Foster collaboration with **SESAR Industrial projects** • Expand our **Advisory Board**	New member of SynthAIr AB
Austrian Airlines	In different initiatives/projects	16-01-24	• Collect perspectives on use cases from the **airline** standpoint • **Disseminate** project information to **pertinent stakeholders** • Expand our **Advisory Board**	To be invited in SynthAIr AB
Swiss International Air Lines	In different initiatives/projects	25-01-24	• Collect perspectives on use cases from the **airline** standpoint • **Disseminate** project information to **pertinent stakeholders** • Confirm their participation to our **Advisory Board**	
Civil Air Navigation Services Organization	In different initiatives/projects	25-01-24	• Collect perspectives on use cases from the **air traffic management industry** • **Disseminate** project information to **pertinent stakeholders** • Confirm their participation to our **Advisory Board**	

Table 3: Schedule of the interviews with relevant stakeholders

The key findings from these interviews will be documented in D2.1 and will offer insights to guide the selection of the most suitable use cases based on the requirements outlined in the following matrix.

Indeed, it outlines the criteria that will inform our UC prioritization. These criteria will be weighted differently, signifying their varying impact on the final selection. The criteria encompass factors such as data availability, alignment with project objectives, stakeholder interest, and the potential for meaningful outcomes.

The decision-making framework we put forth here emphasizes the project's commitment to delivering realistic and impactful results. It is a testament to our strategic approach: to channel our resources and expertise into the most viable UCs, thereby ensuring that the progress we make is both significant and sustainable.

In due course, once the UCs are finalized, we will reintroduce the comprehensive table to transparently convey the decisions made and the rationale behind them, maintaining a clear and accountable project trajectory.

	Req ID	Requirements / Notations description	UC1_Turnaround Time Prediction	UC2_Flight Delay Prediction	UC3_Passenger Flow Prediction	UC4_Synthetic Traffic Generator for Fast and Real-Time Simulations	UC5_Flight Diversion Prediction
Data	**Req_01**	**Availability**					
	Req_02	**Confidentiality**					
	Req_03	**Types and format** - Determine the types of data and the format in which data will be collected and stored, whether structured (e.g., databases), semi-structured (e.g., JSON), or unstructured (e.g., text documents)					
	Req_04	**Quality** - Define the level of data accuracy, completeness, and reliability required (consider data cleansing and validation processes)					
	Req_05	**Privacy and security** - Ensure that data requirements align with privacy and security regulations, and establish measures to protect sensitive information					

EUROPEAN PARTNERSHIP
Co-funded by the European Union

	Req_06	**Frequency** - how often data should be collected or updated, whether in real-time, daily, weekly, or at some other interval					
	Req_07	**Documentation** - Updated documentation that describes the data's source, structure, and usage to facilitate understanding and collaboration					
Analysis and modelling	Req_08	**Analytic capability** of performance variability qualitatively and quantitatively (i.e. amount of uncertainty in prediction tasks)					
	Req_09	**Complexity** of modelling/prediction					
	Req_10	Formal **analysis capability**					
	Req_11	Level of **abstraction** (for modelling)					
	Req_12	Level of **Generalization**					
	Req_13	Time **horizon** for predictive/forecasting models					
	Req_14	Expected quality of the model					
Integration	Req_15	**Integration** with other AI/methods and modelling techniques ensuring compatibility					
Stakeholders	Req_16	**Interest**					

EUROPEAN PARTNERSHIP
Co-funded by the European Union

	Req_17	Relevance						
	Req_18	Priority						
	Req_19	Novelty						
Validation	Req_20	Validation difficulty						
	Req_21							
	Req_22							

Table 4: SynthAIr matrix

EUROPEAN PARTNERSHIP
Co-funded by the European Union

This matrix serves as a comprehensive tool for discussing the feasibility of each use case with stakeholders, involving them in the early stages of the project. The identified requirements will help qualitatively assess potential limitations, weaknesses, and constraints, steering the project toward effective research activities and solutions.

More in details, assessing the quality of synthetic data generated within the SynthAIr project entails a multifaceted approach, incorporating both **qualitative** and **quantitative** measures. Qualitatively, the generative models are evaluated based on three distinct criteria: **Fidelity**, which gauges the realism of synthetic samples as compared to actual data; **Diversity**, assessing the range of variation in the synthetic data against the variability in real samples; and **Generalization**, which examines the model's tendency to overfit to the training data [Reference]. These dimensions create a comprehensive framework for scrutinizing individual data points and overall dataset integrity.

Quantitatively, the evaluation involves a meticulous comparison of feature distributions between synthetic and original datasets using histogram plots. This method is augmented by calculating the differences in these distributions to provide a numerical measure of the data's authenticity. Additionally, the examination of correlation matrices through heat map visualizations offers insights into the interdependencies within the data features, comparing them against those found in the real data.

For the quantitative assessment, domain experts play a pivotal role in the assessment process, where their expertise is utilized in subjective evaluations of the synthetic data points. This expert analysis is instrumental in determining whether synthetic data can be distinguished from real data. The effectiveness of this approach can be quantified using classification performance metrics, where lower performance indicates greater difficulty in differentiating between synthetic and real data, thus higher quality of the synthetic data. This comprehensive evaluation strategy ensures that the synthetic data generated by the SynthAIr project stands up to rigorous standards of quality and applicability.

SynthAIr's Contributions to the ATM System

This section aims to address the potential strengths and benefits of the SynthAIr concept within the context of the Digital European Sky programme and the broader aviation infrastructure. The information presented here will guide future validation activities.

Resilience in ATM Automation: SynthAIr aims to introduce adaptable machine learning (ML) models capable of swift adjustment to new environments, unforeseen events, or varied configurations, particularly in situations where data might be scarce.

Managing Abnormal Situations: Through the utilization of AI-generated synthetic data, SynthAIr will develop models to enhance predictions related to flight delays and passenger flow. This will assist airport operators in decision-making and contribute to simulating new airport configurations, forecasting operations, and identifying and understanding unexpected situations.

AI-Driven Automation: The goal is to enhance the performance of predictive models using synthetic data, enabling the development of advanced AI tools. This advancement is expected to optimize performance and capacity through improved prediction mechanisms.

Dataset Creation for AI Applications: SynthAIr recognizes the significance of datasets in AI development. The project will innovate in generating aviation-specific datasets from diverse communication sources, democratizing data access and supporting ML research in the ATM domain.

Enhancing Passenger Experience: Improving passenger flow predictions using extensive synthetic datasets is a key focus. This initiative aims to optimize airport resource allocation, reducing waiting times, and overall enhancing the passenger experience. Additionally, better airside delay predictions contribute to a smoother passenger journey.

Optimizing Human Performance: SynthAIr anticipates a reduction in workload for controllers and flight crews through the integration of air and ground automation systems. As aviation becomes more data-driven, the application of ML and AI is expected to lead to a smarter and safer ATM system, optimizing human performance in the process.

Alignment with Digital European Sky Programme

SynthAIr's objectives align with the vision of the Digital European Sky programme, contributing to the creation of a scalable, economically viable, and environmentally efficient aviation infrastructure.

TRL Consideration

It's essential to note that SynthAIr is a Research Innovation Action (RIA), and its Technology Readiness Level (TRL) is currently low. While the envisioned strengths and benefits presented earlier hold substantial potential, they remain within the realm of possibilities for future development and validation. As a project in the early stages of research, SynthAIr aims to realize these strengths and benefits through ongoing efforts and advancements, recognizing the need for further exploration and validation to fully manifest its potential impact on the ATM system.

3.3 Key assumptions

The key assumptions depend on the choice of use cases to be performed in Deliverable D2.1.

4 Proposed SESAR solutions

Given the exploratory nature of the SynthAIr project and its current low Technology Readiness Level (TRL), it is essential to clarify that SynthAIr does not directly contribute to any SESAR (Single European Sky ATM Research) solutions at this stage. The project primarily focuses on experimental exploration and research, and its immediate impact on SESAR-specific implementations is limited. It is crucial to underscore that the potential benefits for the Air Traffic Management (ATM) system resulting from SynthAIr have been elucidated in the preceding chapter. As the project progresses and matures, further alignment with SESAR solutions may be explored in subsequent phases.

Co-funded by the European Union

5 Plan for next R&I phase

Given the inherently exploratory nature of the SynthAIr project and its current low Technology Readiness Level (TRL), research and innovation (R&I) efforts are paramount at this initial stage. The project has recently commenced, and our primary focus is on

- generating interest within SESAR and relevant scientific communities
- disseminating pertinent information to key stakeholders
- identify synergies with other SESAR projects
- lay a solid foundation for future development

As we navigate this early phase of development, dedicated research and innovative exploration will guide us in shaping the project's trajectory, fostering collaboration, and garnering support from relevant stakeholders. The emphasis on research and innovation aligns with the project's objective of pushing the boundaries in pursuit of cutting-edge advancements, thereby establishing a robust foundation for subsequent phases of development (for further details please refer to D6.1 CDE activities plan and report).

EUROPEAN PARTNERSHIP
Co-funded by the European Union

6 References

6.1 Reference documents

[1] Sara Castellanos. (2021, July 23). Fake It to Make It: Companies Beef Up AI Models With Synthetic Data. The Wall Street Journal. https://www.wsj.com/articles/fake-it-to-make-it-companies-beef-up-ai-models-with-synthetic-data-11627032601

7 Annex

7.1 Annex 1