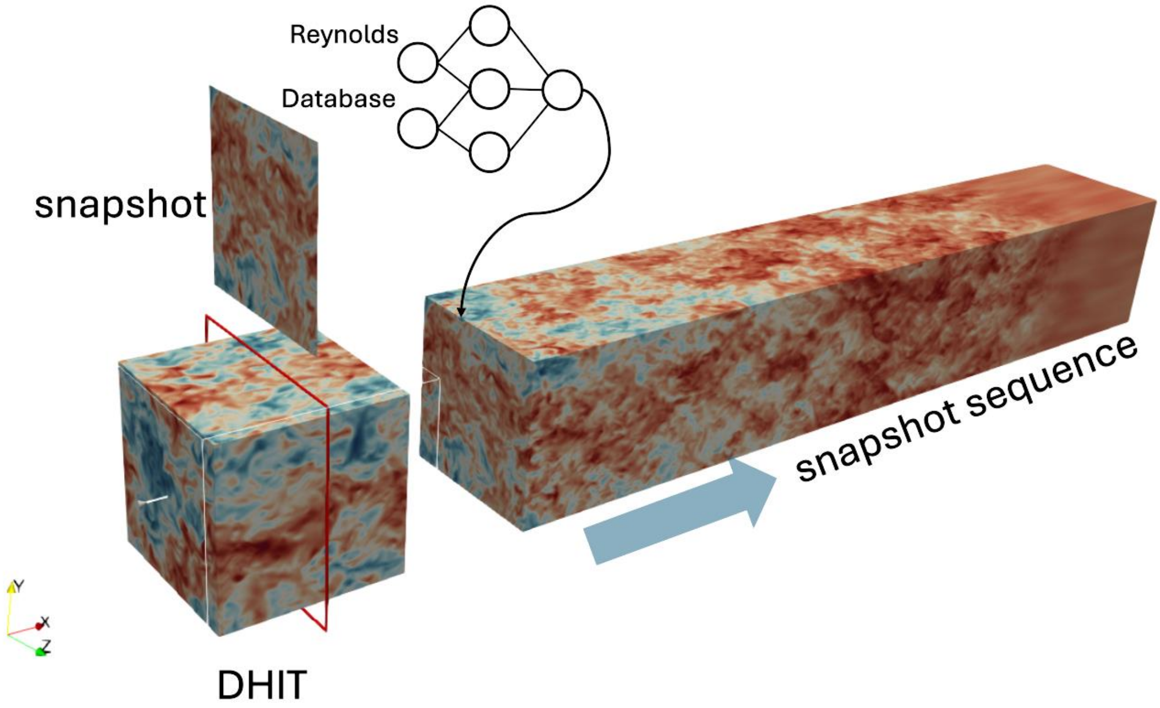
TRAIL Summer Workshop' 25
Project Proposal

Full Name of Team Leader	Dominique Joachim
Project Title	**Generative Learning for Data-Driven Turbulent Inflow**
Profile of the Team Leader(s)	Dominique Joachim – Cenaero – Research Engineer I obtained my Master's degree in Aerospace Engineering from ULiège (Belgium). I then joined the von Karman Institute for Fluid Dynamics, where I completed my PhD in aeroacoustics. My doctoral research focused on the application of data-driven techniques to model wall pressure spectra for Amiet's trailing edge noise model. The PhD was awarded by KU Leuven (Belgium). Following this, I joined Safran Aero Boosters, where I contributed to the optimization of low-pressure compressors. In September 2023, I joined Cenaero as part of the Machine Learning and Optimization research team. My work primarily involves the **application of machine learning to physical problems**, including within the framework of the ARIAC project.
Abstract	Turbulence is a fundamental physical phenomenon that plays a crucial role in understanding and modelling the complexity of real fluid flows. Due to its chaotic nature and multi-scale dynamics, its modelling remains one of the major challenges in Computational Fluid Dynamics (CFD) in particular for the design of industrial applications such as aircraft wings or turbofan blades. The enhancement of computational capabilities and the increased accessibility to large-scale clusters have facilitated the generation of high-fidelity numerical datasets. Simultaneously, the rapid development of machine learning methods, particularly the large diffusion of AI methods for images synthesis and generation, has opened new opportunities to leverage data-driven image generation models for predicting turbulent flow fields. This project focuses on the generation of realistic turbulent flow fields using modern generative techniques, with the goal of improving turbulence injection in numerical solvers. Participants will work with a database of Decaying Homogeneous Isotropic Turbulence (DHIT) simulations obtained at various Reynold numbers, a dimensionless parameter that indicates how turbulent the flow is. The task of the participant is to **develop a generative model** capable of **producing a sequence of 2D turbulent snapshots for a given Reynolds number**. The generated fields must replicate both the statistical properties and dynamic behaviour of the

reference dataset so that it can serve as **physically consistent inflow boundary conditions** in Argo, Cenaero's high-fidelity CFD solver.

This project offers a **unique opportunity for ML practitioners** to apply their skills to **model space-time chaotic data with physical constraints**, and complex multi-scale dynamics with **state-of-the-art generative models such as GANs, Diffusion Models or VAEs.**



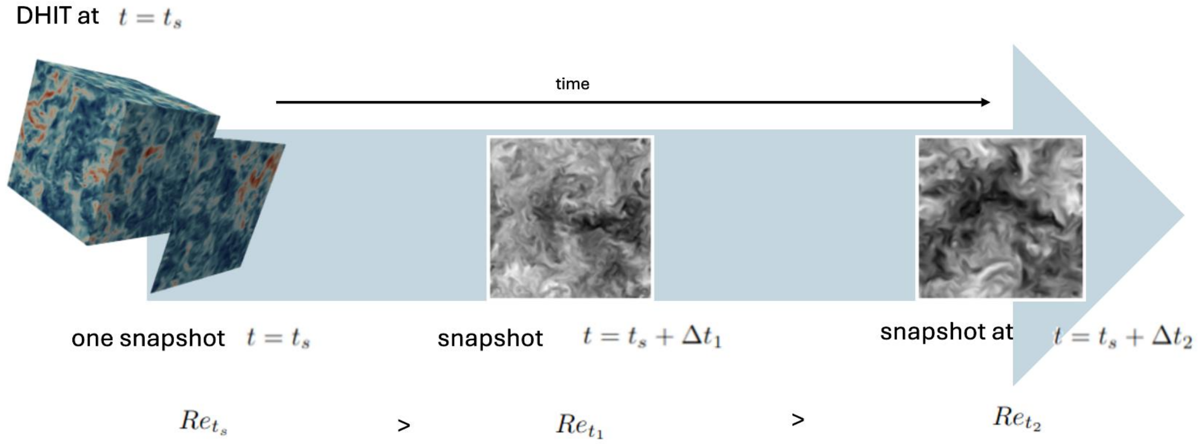
Project Objectives

The objective of this project is to develop and evaluate machine learning-based models for the synthesis of turbulent flow fields images, with a specific focus on their application to turbulence injection. Participants will be provided with a database of turbulent flows and a set of predefined evaluation metrics to guide model development and benchmarking.

To address this task, a wide range of modelling techniques may be considered. These include methods such as Variational Auto-encoders (VAEs), Normalizing Flows, Generative Adversarial Networks (GANs), and Diffusion Models (DMs). In addition, these generative approaches can be combined with temporal models such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRUs), Neural Ordinary Differential Equations (Neural ODEs) and Transformers to account for the dynamic evolution of turbulent structures. The choice of modelling architecture is left open to the participants, encouraging creativity and innovation.

The evaluation of the proposed approaches will be carried out in two stages. First, the models will undergo *a priori* testing, where they are trained and validated against the

	database using the provided metrics. Once the architectures demonstrate sufficient performance, *a posteriori* validation will be conducted by Cenaero at the end of the workshop. In this phase, the models will be used to generate large-scale turbulence fields that serve as inflow boundary conditions in *Argo-DG*, an in-house high-order Discontinuous Galerkin (DG) solver developed at Cenaero for solving the compressible Navier–Stokes equations. This solver is capable of reconstructing fully developed turbulent flows, even from uniform and constant inflow conditions. However, the accuracy of the injected turbulence directly impacts the downstream flow development: the closer the inflow condition is to a physically consistent DHIT field, the shorter the inlet length required to regenerate a fully turbulent flow within the domain, potentially saving precious numerical resources. Consequently, this inlet development length will serve as an additional, physically grounded metric for ranking the performance of the different generative approaches.
Project Dataset	The dataset, generated by Cenaero, is composed of multiple simulations of DHIT, each initialized with a different random seed. This variation in initial conditions leads to distinct realizations of turbulent flow, despite all simulations starting from the same initial energy spectrum. The simulations are performed in a three-dimensional periodic box of size L x L x L, with no imposed mean flow. Each box has an effective resolution of 128 x 128 x 128, allowing for a detailed representation of the fine turbulent structures. Due to the statistical uniformity and directional consistency (isotropy and homogeneity) of the turbulence, the dataset is particularly well-suited for machine learning applications and can be further expanded through data augmentation. While all simulations share the same initial energy spectrum, their different random seeds result in unique flow evolutions. As the simulations evolve in time, turbulence naturally decays, and each time step yields a new 3D fields or *box* of the turbulent flow. Over time, this decay leads to a progressive reduction in the Reynolds number, reflecting the dissipative nature of the flow. At any given moment, it is possible to *freeze* the state of the flow at a particular Reynolds number. These *frozen 3D boxes* are characterized by two primary physical quantities: **Turbulent Kinetic Energy (TKE)** and **Integral Length Scale (Lint)**, which in turn can be combined into an equivalent **Reynolds Number.** From the decaying simulations, a total of 214 such frozen boxes were extracted, each corresponding to a unique pair of TKE and Lint values. These frozen fields can then be used as inflow boundary conditions in numerical simulations with an imposed mean flow, where the frozen turbulence at the prescribed intensity will convected downstream in the numerical domain. This approach aligns with the well-known *frozen turbulence hypothesis* (also referred to as Taylor's hypothesis) widely used in fluid dynamics, which assumes that turbulent structures are advected without significant distortion over short time scales.



Background Information

The enhancement of computational capabilities and the increased accessibility of large-scale clusters have facilitated the use of Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) for real compressible flow configurations. To ensure physically representative simulations, high-quality and realistic boundary conditions are essential. One of the main challenges is reproducing, at the domain inlet, an unsteady turbulent flow compliant with experimental conditions such as turbulence grids. Because such an inflow varies stochastically and continuously in space and time, defining an ideal turbulent inflow remains a complex task. The fluctuations must mimic real turbulence, as they will impact the downstream flow dynamics. The constructed turbulent flow field should satisfy statistical turbulent characteristics both in space and time, be fully developed as quickly as possible after the inlet and should be divergence-free to avoid the injection of spurious pressure waves (i.e., acoustic effects).

A variety of methods have been proposed to obtain high-quality, fully turbulent inflow data, which can be categorized into four groups: (i) transition-inducing methods, (ii) recycling-rescaling methods, (iii) synthetic inflow generators, and (iv) turbulence library-based methods. Each method has its trade-offs between realism, speed, and complexity. The more realistic methods tend to require more computing resources, while the simpler ones may not fully capture the true chaotic nature of turbulence. **The aim of this project is to create a smarter and faster way to add realistic turbulence into fluid simulations through generative Machine Learning approaches** with the short-term objective is to reduce the trial-and-error runs to set up the appropriate turbulence statistics at a given location in the main computational domain as well as the memory consumption of turbulence library-based method while keeping realistic turbulence that has a spectral content similar to that of the actual turbulence.

The fluctuations should be established within a short distance from the inlet. This development distance is a reliable metric for evaluating the efficiency of a method and the validity of the underlying assumptions. The long-term objective is to obtain physical representativeness comparable to library-based and transition-inducing methods while maintaining a low computational cost similar to synthetic methods. **Recent advances in deep learning algorithms, coupled with the increasing power of GPUs and the generation of high-fidelity data, have led to the exploration of novel data-driven**

	approaches to improve existing turbulence injection methods or to develop new ones. In this vein, *Margaux Boxho et al.* used state-of-the-art score-based Diffusion Models (DM), a class of latent variable models inspired by non-equilibrium thermodynamics, to generate three-dimensional periodic boxes of homogeneous isotropic turbulence. Along with Generative Adversarial Networks (GANs) and Variational Auto-Encoders (VAE), DMs belong to the class of generative models that, by learning the probability distribution of the training data, can produce novel outputs that retain the statistical properties of the original data. All generative models face a fundamental trilemma: balancing sample quality, generation speed, and mode coverage. GANs produce high-quality samples efficiently but often suffer from mode collapse, limiting their ability to capture the full data distribution. In contrast, VAEs and normalizing flows provide better mode coverage but typically generate lower quality samples. Finally, DMs are high-quality image generator that outperforms GANs and achieve strong mode coverage. However, their inference process is relatively slow due to the iterative denoising steps required for generation. Current research in this area aims to overcome these respective limitations: making GAN training more stable and mode-complete, improving the visual fidelity of samples generated by VAEs, and accelerating the sampling process of diffusion models through advanced Stochastic Differential Equation (SDE) solvers. These developments are essential for pushing the boundaries of generative modelling and expanding its practical applications across scientific and engineering domains. Although the current work focuses on homogeneous isotropic freestream turbulence, our goal is to develop a method that can be extended to other types of turbulent inflows, such as anisotropic turbulence, boundary layers, and wakes. Additionally, this approach paves the way for controlling and defining inflow conditions to achieve desired turbulence characteristics at specific locations within the computational domain. A typical scenario involves replicating experimental conditions recorded by a probe in a wind tunnel to match the numerical simulation to the experimental conditions.
Bibliographic References	[1] Fukami, K., Nabae, Y., Kawai, K., & Fukagata, K. (2019). Synthetic turbulent inflow generator using machine learning. *Physical Review Fluids*, *4*(6), 064603.41J. Kim and C. Lee, "Deep unsupervised learning of turbulence for inflow generation at various reynolds numbers," J. Comput. Phys. 406, 109216 (2020). [2] Kim, J., & Lee, C. (2020). Deep unsupervised learning of turbulence for inflow generation at various Reynolds numbers. *Journal of Computational Physics*, *406*, 109216. [3] Kanishk, Nandal, T., Tyagi, P., & Singh, R. K. (2021, November). Generation and Parameterization of Forced Isotropic Turbulent Flow Using Autoencoders and Generative Adversarial Networks. In *ASME International Mechanical Engineering Congress and Exposition* (Vol. 85666, p. V010T10A062). American Society of Mechanical Engineers. [4] Yousif, M. Z., Zhang, M., Yu, L., Vinuesa, R., & Lim, H. (2023). A transformer-based synthetic-inflow generator for spatially developing turbulent boundary layers. *Journal of Fluid Mechanics*, *957*, A6. [5] Liu, X. Y., Parikh, M. H., Fan, X., Du, P., Wang, Q., Chen, Y. F., & Wang, J. X. (2024). CoNFiLD-inlet: Synthetic Turbulence Inflow Using Generative Latent Diffusion Models with Neural Fields. *arXiv preprint arXiv:2411.14378*.

	[6] M. Boxho, J. Dominique, T. Benamara, M. Rasquin, L. Salesses, C. Sainvitu, G. Louppe, and T. Toulorge (2025) Turbulent Injection assisted by Diffusion Models for Scale Resolving Simulations. (In writing)
Detailed Work Plan	In the context of reproducing realistic boundary conditions for high-fidelity numerical simulations of turbulent flows, generating new samples from Decaying Homogeneous Isotropic Turbulence (DHIT) simulations has been identified as an ideal test case due to their spatial homogeneity, isotropic properties and relevance to several industrial applications. A database comprising four-dimensional DHIT fields, frozen at various time steps and corresponding to different Reynolds numbers, has been prepared by Cenaero for use in this study. To address this task, a wide range of modelling techniques may be considered. These include methods such as Variational Auto-Encoders (VAEs), Normalizing Flows, Generative Adversarial Networks (GANs), and Diffusion Models (DMs). In addition, these generative approaches can be combined with temporal models such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRUs), Neural Ordinary Differential Equations (Neural ODEs) and Transformers to account for the dynamic evolution of turbulent structures. **The choice of modelling architecture is left open to the participants, encouraging creativity and innovation.** **The evaluation of the proposed approaches will be carried out in two stages.** First, the models will undergo ***a priori* testing**, where they are trained and validated against the database using a set of metrics provided by Cenaero. Preceding investigations have established that the physical representativeness of the samples can be evaluated using a series of statistical metrics: the energy spectrum, two-point autocorrelation functions, the barycentric map, and the vorticity distribution. These metrics will guide the training and preliminary validation of the proposed machine learning models. Once the architectures demonstrate sufficient performance, ***a posteriori* validation** will be conducted by Cenaero at the end of the workshop. In this phase, the models will be used to generate turbulence fields that serve as inflow boundary conditions in *Argo-DG*. Although Argo-DG is capable of reconstructing fully developed turbulent flows from uniform and constant inflow conditions, the accuracy of the injected turbulence directly impacts the downstream flow development. The closer the inflow condition is to a physically consistent DHIT field, the shorter the inlet length required to regenerate a fully turbulent flow within the domain. Consequently, this inlet development length will serve in the *a posteriori* phase as an additional, physically grounded metric for ranking the performance of the different generative approaches. The full *a posteriori* testing framework, based on the injection of turbulent inflow fields into a free domain using the high-order Argo-DG solver, will not be directly accessible to participants, but they will have the opportunity to submit their generated inflow fields which will be injected by Cenaero into the Argo-DG flow solver. At the conclusion of the workshop, participants will receive feedback on their model performances, specifically in terms of the inlet length required to develop fully turbulent flow downstream.

	Resources needed • A laptop equipped with a Python environment (>= 3.9) and a ML framework is needed. • TRAIL researchers can ask access to Lucia, the Tier-1 supercomputer hosted by Cenaero to train the ML models. Requests for getting access to Lucia for a researcher of one of the Walloon universities are managed by the CÉCI and hence must be filed through the corresponding request system. However, the full CFD-based simulation chain will not be directly accessible to the participants. **Resources provided** • A database containing 214 DHIT boxes, each sized 128x128x128, at various Reynolds numbers. • A custom-made PyTorch dataloader script including data augmentation procedures. • A set of *a priori* metrics to evaluate the generated fields. • Access to a private GitHub repository. **Project Management** A brainstorming session will be organized at the beginning of the workshop to introduce the project in detail, to exchange ideas, and to organize the team (depending on the number of participants, a split-off into groups of 2-3 people should be done). Daily scrum meetings with the whole team will be held. A private GitHub repository will be used to collaborate on the code developed during the workshop. If necessary, the database can already be shared before the start of the workshop. **Benefits of the research** This project aims to advance turbulence injection methodologies by leveraging machine learning to generate physically realistic turbulent inflow conditions **at a reduced computational cost and memory footprint.** By addressing the trade-off between simulation fidelity and resource efficiency, **this work contributes to the development of next-generation hybrid turbulent modelling strategies.** The tangible outputs / deliverables : • A demo code for each generative approach investigated on the provided DHIT data. • Python codes on GitHub to reproduce the generation on inflow turbulence. When mature enough, the codes will be released on the TRAIL Factory. The dataset could also be stored with the code on the TRAIL Factory. • Final slides and reports presenting the methodology and the results (with appropriate visualization plots). Depending on the results, the writing of a scientific paper could be considered after the end of the workshop.
Other Remarks	This project is linked to **Grand Challenge 1, "Hybrid Modelling Methods towards an Augmented Engineering,"** led by Cenaero. Similar to the use case of AI-enabled additive

manufacturing, the main purpose is to investigate how machine learning can accelerate numerical simulations framework, such as Computational Fluid Dynamics (CFD), making them faster and more accessible for industrial application. In this specific proposal, we tackle the theme of physics-based models and generative AI to propose a novel framework for turbulence injection approaches. The long-term objective is to reduce the computational cost and memory footprint of such models, paving the way toward controlling inflow turbulence and streamlining the turbulence injection process of ARGO-DG.

Potential team members: (to be completed)

- Joahim Dominique – Cenaero (Team Leader)
- Lionel Salesses – Cenaero
- Caroline Sainvitu – Cenaero
- Margaux Boxho – Cenaero (remote – for the *a posteriori* validation)

For this project, we are particularly interested in participants who have **expertise or a strong interest in generative modelling techniques** (such as diffusion models, VAEs or GANs) and are keen to explore the application of ML to areas such as fluid dynamics, even if they have no prior expertise in physics or computational fluid dynamics. If you are passionate about applying generative models to new and challenging domains, we encourage you to contribute to this project.