TRAIL TRUSTED AI LABS BY DIGITALWALLONIA / SPW-RECHERCHE

Malleable Latent Space for Reliable Time Series Generation and OOD Detection

Full Name of Team Leader	Horacio Tellez
Project Title	Malleable Latent Space for Reliable Time Series Generation and OOD Detection
Profile of the Team Leader(s)	Horacio Tellez is a research scientist in the Fundamental AI team at Multitel [1] led by Emmanuel Jean. Tellez's research focus at the moment is around quality metrics on generative AI and out-of-distribution detection (OOD) methods.
Abstract	The generation of reliable synthetic time series data remains a critical challenge in the application of generative AI to high-stakes domains such as healthcare, finance, and industrial monitoring. Unlike in image or text domains, where visual or semantic fidelity can be easily assessed, temporal data lacks intuitive inspection methods and robust quality metrics. This project proposes a novel approach to time series generation through the construction of a malleable and interpretable latent space using diffusion models. By embedding temporal dynamics into a structured representation space, we aim to make synthetic data generation both controllable and transparent. This latent space will serve as a foundation not only for evaluating the fidelity of generated sequences but also for enhancing the detection of out-of-distribution (OOD) samples, an essential capability for systems operating in uncertain and dynamic environments. Our methodology involves a tailored training regime with new loss constraints to shape the latent topology and leverages known and custom metrics to characterize and differentiate real, synthetic, and anomalous data. Ultimately, this work seeks to advance the reliability, interpretability, and trustworthiness of generative models in temporal contexts.
Project Objectives	The widespread success of generative models in domains such as image, text, and music has transformed the role of AI in both research and daily life [2, 3, 4]. However, the adoption of generative AI in high-stakes applications—particularly in healthcare, finance, and industrial monitoring—remains limited. A major barrier is the lack of robust, interpretable quality metrics for synthetic data, particularly for temporal data such as time series. Time series data presents unique challenges in generation and evaluation. Unlike images, where visual inspection can suffice for quality checks, time series lacks intuitive visualization for non-expert validation. Existing metrics for generated time series often rely on traditional statistical and machine learning approaches, which fail to capture the nuanced differences between real and synthetic data. This project proposes the development of a malleable and interpretable latent space using diffusion models, enabling controlled time series generation and robust out-of-distribution (OOD) detection. To achieve this, we introduce a specialized training regime that shapes the forward diffusion process to enhance both sample synthesis and OOD discriminability. Central to our approach is a targeted noise sampling strategy that progressively

	constrains the latent distribution to a low-dimensional linear subspace. This is achieved through a two-phase noise scaling schedule applied during the forward process: initially, noise is isotropic, but after a threshold—when the latent distribution sufficiently resembles a multimodal Gaussian—a subset of noise dimensions is gradually attenuated. This subspace alignment is further encouraged by an additional regularization term in the loss function, defined as the squared ℓ_2 norm over the selected noise components. Together, these mechanisms guide the generative process toward geometrically structured latent representations, fostering both interpretability and robustness. We hypothesize that the construction of such a latent space will facilitate improvements in both the generation of high-fidelity synthetic data and the detection of anomalies or previously unobserved patterns within novel datasets. Diffusion models were selected for this investigation due not only to their state-of-the art performance in generative artificial intelligence, but also for their inherent capacity to incorporate custom conditioning mechanisms during the training phase such as conditioning [5] and guided diffusion [6]. Our objectives are the following: **Objective 1**: Integrate the proposed custom training regime into the development of a diffusion model tailored for time series generation, with the specific aim of constructing an interpretable latent space. By enforcing selected topological properties within this latent space, we anticipate that its structure will enable both controllable generation and improved interpretability of the model's outputs. **Objective 2**: Choose or devise a metric on this latent space that better captures the difference between generated and real data. A dimensionality reduction of data and distillation of important features through diffusion will allow us to explore several existing metrics and see how they perform in the case of temporal data. **Objective 3**: Investigate how this combination of latent space and metric behaves in the context of out-of-distribution detection. We will use the latent space as a proxy detector for out-of-distribution data. By creating a characterization of common features of real and generated data, we hope anomalies and out-of-distribution data will be easier to detect.
Project Dataset	The dataset utilized in this study is the "NASA Anomaly Detection Dataset SMAP & MSL" [7], initially introduced in [8] for the investigation of anomaly detection in time series data. Since its release, it has been widely adopted for various analyses pertaining to out-of-distribution (OOD) detection [9, 10, 11]. **Dataset Description**: The dataset consists of expert-annotated telemetry anomaly data sourced from two distinct platforms: the Soil Moisture Active Passive (SMAP) satellite and the Mars Science Laboratory (MSL) rover, Curiosity. The captured anomalies reflect operational irregularities encountered during actual spacecraft and rover missions, offering a robust and realistic benchmark for the evaluation of anomaly detection algorithms. Anomalous events previously documented in Incident Surprise Analysis (ISA) reports were comprehensively reviewed. Each ISA involved a thorough examination of all relevant telemetry channels to verify the presence of anomalies within the associated

	time series data. Annotators then manually labeled anomalous time intervals for each affected channel. In cases where multiple anomaly sequences or telemetry channels displayed high similarity, a single representative instance was selected to preserve dataset diversity and mitigate redundancy. Anomalies were categorized into two principal types: point anomalies and contextual anomalies. Point anomalies correspond to isolated deviations that can often be detected using threshold-based or distance-based methods without considering temporal dependencies. Conversely, contextual anomalies require the use of more advanced temporal modeling approaches—such as Long Short-Term Memory (LSTM) networks or Hierarchical Temporal Memory (HTM)—to capture temporal context and identify deviations within structured temporal patterns.
Background Information	**Time Series** **Analysis and Generation.** Time series analysis has evolved from traditional linear statistical models, such as ARIMA and Exponential Smoothing, to advanced deep learning approaches capable of modeling complex, nonlinear, and high-dimensional temporal dependencies. Modern architectures—including RNNs [12], LSTMs [13], GRUs [14], TCNs [15], and Transformer- based models [16] have been specifically developed or adapted to address the sequential and often non-stationary characteristics of time series data. Advances in time series generation leverage GAN based models [17] to synthesize realistic sequences that preserve temporal and distributional properties of the original data. More recent developments include diffusion models such as TimeGrad [18] using DDPM [19] as a basis and its score-based [6] counterpart ScoreGrad [20]. See [21] for an in-depth discussion. Anomaly detection remains crucial for domains such as predictive maintenance, fraud detection and preventive health- care. Deep learning models, including USAD [22], InterFusion [23], and TranAD [24], have demonstrated superior performance over classical methods, particularly in high-dimensional or nonlinear settings. Out-of-distribution (OOD) detection, although less explored in time series than in vision or NLP, is gaining attention [25, 26, 27, 28]. Nevertheless, OOD detection in time series is challenged by data heterogeneity and the absence of standardized benchmarks. **Measuring similarities.** Time series similarity measures are essential for tasks such as classification, clustering, and anomaly detection, with various methods tailored to address alignment, noise, and data complexity. Outside of the immediate Euclidean distance a common metric used is dynamic time warping [29] that effectively handles temporal misalignments. Correlation based metrics, such as Pearson correlation coefficient and cosine similarity prove effective to assess linear or angular relationships between series. These are scale-invariant but may fail to capture shape differences or temporal shifts. Methods like Complexity-Invariant Distance [30] adjust for series complexity. The choice of similarity measure should be guided by data characteristics and task requirements. Due to the intrinsic data heterogeneity in time series no single metric is accepted as universally optimal. **Diffusion Models**

Diffusion models [31, 32, 33, 34, 2] have emerged as a powerful class of likelihood based generative models, demonstrating state-of-the-art performance in high-fidelity image synthesis tasks [31, 19, 35]. These models operate through a two-stage process. In the forward diffusion process, Gaussian noise is incrementally added to the input data over a series of time steps, progressively transforming the data into a noise distribution. In the reverse diffusion process, a neural network is trained to learn the denoising trajectory, effectively reconstructing data by reversing the noise addition step-by-step. Conceptually, diffusion models share a high-level resemblance to Variational Autoencoders (VAEs) [36], in that both involve projecting data into a latent space and reconstructing it. However, unlike VAEs—which directly approximate the data distribution using variational inference—diffusion models learn a Markovian noise corruption process and its time-reversal, enabling a more stable and fine-grained generation process. This hierarchical denoising mechanism allows diffusion models to capture complex data distributions with high precision. For an in-depth mathematical treatment of the underlying principles, readers are referred to [37, 38].

Quality Metrics

The advent of Generative Adversarial Networks (GANs) [39] and their capacity to produce high-quality synthetic data have highlighted the need for rigorous and interpretable metrics to evaluate the performance and reliability of generative models. One of the earliest such metrics is the Inception Score (IS) [40], which employs the Inception network [41] to derive the conditional label distribution of generated samples. The IS quantifies sample quality by computing the Kullback–Leibler divergence between the conditional label distribution and the marginal label distribution, aiming to capture both the clarity and diversity of generated outputs. A more robust alternative, the Fréchet Inception Distance (FID) [42], also leverages the Inception network but instead compares the feature distributions of real and generated data. It assumes these distributions are Gaussian and computes the Fréchet distance (also known as the Wasserstein-2 distance) [43] between them. FID has been shown to be more stable and less sensitive to noise than IS. To address the limitations imposed by the Gaussian assumption in FID, the Kernel Inception Distance (KID) [44] was introduced. KID measures the squared Maximum Mean Discrepancy (MMD) between real and synthetic feature embeddings, using a polynomial kernel, and provides unbiased estimations with finite sample sets. More recently, efforts to refine quality assessment have focused on multidimensional evaluations of generative models. Alaa et al. [45] propose a comprehensive framework that separately quantifies three core aspects of generative quality: fidelity, diversity, and generalization. Their methodology relies on an embedding space shared by real and synthetic data, allowing for targeted comparisons using parameterized distance metrics.

Out of Distribution Detection

Out-of-distribution (OOD) detection is crucial for the reliability and safety of machine learning systems, such as alerting autonomous vehicles to unfamiliar situations. Since

	the term introduction in 2017, OOD detection has spurred a range of methods and overlaps with related fields like anomaly detection, novelty detection, open set recognition, and outlier detection—though these areas often develop independently, leading to confusion over definitions. The common denominator of these disciplines is the need to identify samples that are not part (what "being part of" means often being the trait that differentiates between them) of the training distribution. OOD benefits from being a problematic that can be tackled from several angles. Common statistical methods such as the Mahalanobis distance [46], post-hoc methods [47, 48] or a more theoretical insight [49]. We refer the reader to [50] for a more complete overview of the field.
Bibliographic References	[1] Multitel. Multitel – Centre de recherche et d'innovation technologique. Accessed: 2025-04-16. 2025. Url : https://www.multitel.be/. [2] Shekoofeh Azizi et al. "Synthetic Data from Diffusion Models Improves ImageNet Classification". In: arXiv preprint arXiv:2304.08466 (2023). [3] Mert Bulent Sariyildiz et al. "Fake it till you make it: Learning transferable representations from synthetic ImageNet clones". In: CVPR 2023–IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. [4] Hritik Bansal and Aditya Grover. "Leaving Reality to Imagination: Robust Classification via Generated Datasets". In: arXiv preprint arXiv:2302.02503 (2023). [5] Jonathan Ho et al. Cascaded Diffusion Models for High Fidelity Image Generation. 2021. arXiv: 2106.15282 [cs.CV]. [6] Yang Song et al. Score-Based Generative Modeling through Stochastic Differential Equations. 2021. arXiv: 2011. 13456 [cs.LG]. [7] Patrick Fleith. NASA Anomaly Detection Dataset SMAP & MSL. https://www.kaggle.com/datasets/patrickfleith/nasa-anomaly-detection-dataset-smap-msl. Accessed: 2025-04-17. 2022. [8] Kyle Hundman et al. "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding". In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18. London, United Kingdom: Association for Computing Machinery, 2018, pp. 387–395. isbn: 9781450355520. doi: 10.1145/3219819.3219845. [9] Haixu Wu et al. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. 2023. arXiv:2210.02186 [cs.LG]. [10] Lawrence Wong et al. AER: Auto-Encoder with Regression for Time Series Anomaly Detection. 2022. arXiv:2212.13558 [cs.LG]. [11] Alexander Geiger et al. TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks. 2020. arXiv: 2009.07769 [cs.LG]. [12] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. "Recurrent Neural Networks for Time Series Forecasting: Current status and future directions". In: International Journal of Forecasting 37.1 (Jan. 2021), pp. 388–427. issn: 0169-2070. doi: 10.1016/j.ijforecast.2020.06.008. [13] Yaxuan Kong et al. Unlocking the Power of LSTM for Long Term Time Series Forecasting. 2025. arXiv: 2408. 10006 [cs.LG]. [14] N. Benjamin Erichson, Soon Hoe Lim, and Michael W. Mahoney. Gated Recurrent Neural Networks with Weighted Time-Delay Feedback. 2022. arXiv: 2212.00228 [cs.LG].

[15] Jens Schreiber, Stephan Vogt, and Bernhard Sick. "Task Embedding Temporal Convolution Networks for Transfer Learning Problems in Renewable Power Time Series Forecast". In: Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track. Springer International Publishing, 2021, pp. 118–134. isbn:9783030865146. doi: 10.1007/978-3-030-86514-6_8.
[16] Qingsong Wen et al. Transformers in Time Series: A Survey. 2023. arXiv: 2202.07125 [cs.LG].
[17] Eoin Brophy et al. Generative adversarial networks in time series: A survey and taxonomy. 2021. arXiv: 2107. 11098 [cs.LG].
[18] Kashif Rasul et al. "Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting". In: (2021). arXiv: 2101.12072 [cs.LG].
[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models". In: Advances in Neural Information Processing Systems 33 (2020), pp. 6840-6851.
[20] Tijin Yan et al. ScoreGrad: Multivariate Probabilistic Time Series Forecasting with Continuous Energy-based Generative Models. 2021. arXiv: 2106.10121 [cs.LG].
[21] Caspar Meijer and Lydia Y. Chen. The Rise of Diffusion Models in Time-Series Forecasting. 2024. arXiv: 2401.03006 [cs.LG].
[22] Julien Audibert et al. "USAD: UnSupervised Anomaly Detection on Multivariate Time Series". In: KDD '20. Virtual Event, CA, USA: Association for Computing Machinery, 2020, pp. 3395–3404. isbn: 9781450379984. doi: 10.1145/3394486.3403392.
[23] Zhihan Li et al. "Multivariate Time Series Anomaly Detection and Interpretation using Hierarchical Inter-Metric and Temporal Embedding". In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. KDD '21. Virtual Event, Singapore: Association for Computing Machinery, 2021, pp. 3220–3230. isbn: 9781450383325. doi: 10.1145/3447548.3467075.
[24] Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data. 2022. arXiv: 2201.07284 [cs.LG].
[25] Angus Dempster, Francois Petitjean, and Geoffrey I. Webb. "ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels". In: Data Mining and Knowledge Discovery 34.5 (July 2020), pp. 1454–1495. issn: 1573-756X. doi: 10.1007/s10618-020-00701-z.
[26] Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. "MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification". In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. KDD '21. ACM, Aug. 2021, pp. 248–257. doi: 10.1145/3447548.3467231.
[27] Kenny Schlegel, Peer Neubert, and Peter Protzel. HDC-MiniROCKET: Explicit Time Encoding in Time Series Classification with Hyperdimensional Computing. 2022. arXiv: 2202.08055 [cs.LG].
[28] Emadeldeen Eldele et al. Time-Series Representation Learning via Temporal and Contextual Contrasting. 2021. arXiv: 2106.14112 [cs.LG].
[29] "Dynamic Time Warping". In: Information Retrieval for Music and Motion. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 69–84. isbn: 978-3-540-74048-3. doi: 10.1007/978-3-540-74048-3_4.
[30] Gustavo E.A.P.A. Batista, Xiaoyue Wang, and Eamonn J. Keogh. "A Complexity Invariant Distance Measure for Time Series". In: Proceedings of the 2011 SIAM

International Conference on Data Mining (SDM), pp. 699–710. doi: 10.1137/1.9781611972818.60.
[31] Jascha Sohl-Dickstein et al. "Deep unsupervised learning using nonequilibrium thermodynamics". In: International Conference on Machine Learning. PMLR. 2015, pp. 2256–2265.
[32] Lilian Weng. "What are diffusion models?" In: lilianweng.github.io (July 2021).
[33] Jko Hee. "Awesome Diffusion Models". In: (2023). url: https://github.com/heejkoo/Awesome-Diffusion-Models.
[34] Won Andrew. "Stable Diffusion Samplers: A Comprehensive Guide". In: (Apr. 2023).
[35] Prafulla Dhariwal and Alexander Nichol. "Diffusion models beat gans on image synthesis". In: Advances in Neural Information Processing Systems 34 (2021), pp. 8780–8794.
[36] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. 2022. arXiv: 1312.6114 [stat.ML].
[37] Sergios Karagiannakos and Nikolas Adaloglou. "How diffusion models work: the math from scratch". In: (Sept. 2022). URL: https://theaisummer.com/diffusion-models/.
[38] Calvin Luo. "Understanding diffusion models: A unified perspective". In: arXiv preprint arXiv:2208.11970 (2022).
[39] Ian J. Goodfellow et al. Generative Adversarial Networks. 2014. arXiv: 1406.2661 [stat.ML].
[40] Tim Salimans et al. Improved Techniques for Training GANs. 2016. arXiv: 1606.03498 [cs.LG].
[41] Christian Szegedy et al. Rethinking the Inception Architecture for Computer Vision. 2015. arXiv: 1512.00567 [cs.CV].
[42] Martin Heusel et al. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. 2018. arXiv: 1706.08500 [cs.LG].
[43] Maurice Fréchet. "Sur la distance de deux lois de probabilité". In : Annales de l'ISUP VI.3 (1957), pp. 183–198.
[44] Mikolaj Binkowski et al. Demystifying MMD GANs. 2021. arXiv: 1801.01401 [stat.ML].
[45] Ahmed Alaa et al. "How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models". In: International Conference on Machine Learning. PMLR. 2022, pp. 290–306.
[46] Jie Ren et al. "A Simple Fix to Mahalanobis Distance for Improving Near-OOD Detection". In: CoRR abs/2106.09022 (2021). arXiv: 2106.09022.
[47] Shiyu Liang, Yixuan Li, and R. Srikant. "Enhancing The Reliability of Out-of distribution Image Detection in Neural Networks". In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net, 2018.
[48] Weitang Liu et al. Energy-based Out-of-distribution Detection. 2021. arXiv: 2010.03759 [cs.LG].
[49] Peyman Morteza and Yixuan Li. Provable Guarantees for Understanding Out-of distribution Detection. 2021. arXiv: 2112.00787 [cs.LG].
[50] Jingkang Yang et al. Generalized Out-of-Distribution Detection: A Survey. 2024. arXiv: 2110.11334 [cs.CV].
[51] R. L. Winkler et al. "Scoring rules and the evaluation of probabilities". In: Test 5.1 (June 1996), pp. 1–60. issn: 1863-8260. doi: 10.1007/BF02562681.

	[52] TRusted AI Labs (TRAIL). TRAIL – Connecting the AI Ecosystem to You. Accessed: 2025-04-16. 2025. url: https://trail.ac/en/.
Detailed Work Plan	While the overall structure of the project pipeline follows conventional practices, we provide here an informal overview of its anticipated progression. The principal objective is the successful implementation of a customized training regime for a diffusion-based generative model. Initial efforts will focus on established metrics and out-of distribution (OOD) detection methods within the latent space, with novel approaches considered only upon time availability. The project is structured in the following stages: • Model training: A diffusion model will be trained, comprising both a forward process M_f and a backward process M_b, with the custom training regime applied throughout. • Synthetic data generation: The backward model M_b will be employed to generate multiple synthetic datasets, denoted $\{D_i\}$. • Latent space projection: Each synthetic dataset D_i will be mapped to the latent space via the forward model M_f. • Metric evaluation: A variety of metrics will be applied in the latent space to quantify the divergence between real and synthetic features. Metrics that most effectively capture these distinctions will be selected. • Model refinement: A cloned instance of the backward model M_b will be fine-tuned to improve the placement of synthetic samples in the latent space in accordance with the selected metrics. • Final evaluation: Improvements in synthetic sample quality and the effectiveness of OOD detection will be assessed as downstream outcomes of the preceding steps. Below are details pertaining to specific components of the project that may be of particular interest for further specification. **Forward process schedule and loss function:** The training protocol is outlined as follows. During the forward diffusion process, we implement a targeted noise sampling strategy. Let $\mathbf{Z} = \left(\mathbf{Z_1}, \ldots, (\mathbf{Z_n}) \sim \mathbf{N}(\mu, \Sigma)\right)$ denote a multivariate Gaussian distribution. It is a well-established property that any sub-vector $\left(\mathbf{Z}_{\mathbf{i_1}}, \ldots, \mathbf{Z}_{\mathbf{i_k}}\right)$ of $\mathbf{Z}$ also follows a Gaussian distribution. We leverage this property to direct noise sampling along specific linear subspaces, thereby enhancing the separability and visibility of anomalous or alien features. The technical details may be refined as the project progresses. Let n be the dimension of samples in the dataset and of the sampled noise and T the last step on the diffusion process. To operationalize this, we define a noise scaling schedule $\{\gamma_{\mathbf{t},\mathbf{i}}\}$, where $\mathbf{t}$ denotes the diffusion time step and i indexes the i-th component of the noise vector. This schedule operates in two distinct phases. In the initial phase, noise is isotropic across all components: $\forall t, \gamma_{t,i} = 1$. Once the forward diffusion distribution $q_t \sim q(x_t, x_0)$ sufficiently approximates a multimodal Gaussian (quantified by a threshold),

we initiate the second phase. At this stage, a subset of components $\{i_k\} \subseteq \{1, \dots, n\}$, selected based on dataset complexity, are progressively attenuated. Specifically, the scaling coefficients $\gamma_{t,i}$ decay according to: $\gamma_{t,i} = \left(\frac{1}{n}\right)^t$ $if\ i \in \{i_k\}$, $else$ 1 such that, asymptotically, the noise vector is confined to a lower-dimensional linear subspace of the latent space. To further reinforce this structural bias in the latent space, we augment the loss function with an additional regularization term. This term penalizes deviations from the desired subspace alignment and introduces an inductive bias toward the intended representation geometry. For simplicity and computational efficiency, we currently define the regularization as the squared ℓ_2 norm over the selected components: $\mathcal{L}_{reg} = \sum_{i \in \{i_k\}} x_{T,i}^2$. This formulation encourages concentration of the latent representation within the designated subspace, thereby promoting more interpretable and structured generative dynamics.

Data format: With temporal data there is always the key choice of the representation space. We are going to use both the time-domain and frequency-domain representation. We hypothesize that models could benefit from having both these representations given in an explicit way.

Model selection and architecture: We will start by using two common approaches for diffusion models: DDPM [19] and its generalization using score-based SDEs [6]. Both approaches are flexible enough to allow for a multitude of architectures. Currently, U-Nets with attentions layers have proven to be the most surefire approach for diffusion models. The models will be trained with a standard agenda for diffusion models.

Gen AI metrics and OOD detection: Under the assumption that the latent space captures the essence of the data, we will use known methods as Mahalanobis distance, Kernel Density Estimation and Laplacian eigenmaps to find three types of characteristics in the data:

- features that are common to both synthetic and real data;
- features unique to the synthetic data;
- features only particular to the real data;

We can then use these to both identify synthetic data and to find outliers in the real data by identifying which features are most prominent in each sample. This method can be used in an iterative refinement process to improve the generative model by either eliminating features only particular to the synthetic data or by making them alike to those of the real data.

Evaluation: For OOD benchmarking we will use standard metrics such as AUROC, F1 score, FPS and accuracy on our dataset. For generative AI quality we will compare how our method performs against the standard metrics such as FID, IS and KID when applied to temporal data. We will also use common metrics for time series: Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Continuous Ranked Probability Score (CRPS) [51].

	Project Management A preliminary literature review and initial discussions among researchers will take place prior to the start of the workshop. The workshop itself will focus on designing and implementing the framework, conducting experiments, and analyzing results. A Kanban board (using tools like Notion or Trello) will help break down the work into manageable tasks for team members. Code development will be managed through a Gitlab served at the TRAIL Factory, enabling real-time collaboration. Communication will be facilitated via Slack for easy exchange of ideas and file sharing. For remote participants, a Microsoft Teams meeting will remain open during working hours to ensure continuous engagement. Tasks related to writing the publication will be scheduled after the workshop. Once final teams are confirmed, a detailed task breakdown and Gantt chart will be shared. **Tools and Resources** • Hardware: GPU for training Deep Learning models, hopefully at least 24Gb of VRAM on minimum a 4090 RTX; • Logistics: internet access, electrical outlets and/or extensions for team members, room with tables and chairs for the team members, white board or a similar writing surface, projector to share our computer screens to present results, etc; • Programming Language: Python; • Frameworks: PyTorch; • Development environment: git + docker
Other Remarks	This project is directly related to the “Great Challenge 5” (Weakly Supervised Machine learning, Towards a More General AI) of the ARIAC/TRAIL initiative [52]. **Expected Outcomes** The expected outcomes of this work include the development of a novel, interpretable diffusion-based latent space specifically designed for time series data. We aim to introduce new metrics for evaluating the quality of generated time series, alongside a validated methodology for out-of-distribution (OOD) detection grounded in the structure of latent embeddings. To support transparency and reproducibility, we will provide an open-source implementation and a comprehensive benchmarking framework. If results show sufficient innovation and impact, we will of course publish our findings.