ANONYMIZING STRATEGIC DATA WITH AVATAR

White paper coordinated by:

Confiance ai

AIRBUS Air Liquide Confiance ai CIVITEO Ethik-IA Garantie Humaine de l'IA

GICAT MyData-TRUST Nantes Université OCTOPIZE MIMETHIK DATA pwc

Renault Group sopra steria SystemX THALES UNSW SYDNEY VENDÉE LE DÉPARTEMENT

OCTOPIZE THANKS

Our contributors

AIRBUS
Air Liquide
Confiance ai
iA
CIVITEO
Ethik-IA Garantie Humaine de l'IA
GICAT
MyData-TRUST
Nantes Université
pwc
Renault Group
sopra steria
SystemX
THALES
UNSW SYDNEY
VENDÉE LE DÉPARTEMENT

and Confiance.ai partners

Air Liquide
AIRBUS
Atos
cea
Inria
NAVAL GROUP
Renault Group
SAFRAN
SAINT EXUPÉRY
sopra steria
SystemX
THALES
Valeo

FRANCE 2030

This work was supported by the French government as part of the "France 2030" program, within the SystemX Technological Research Institute as part of the Confiance.ai program.

Edited

On behalf of the entire Octopize team, we dedicate this white paper to our dear colleague Rémy, who passed away on January 19. ***Rémy, we will miss you!***

Our deepest thoughts go out to his family and loved ones.

It is with great pride that we share these works with you. I wish you an excellent read.

Olivier Breillacq

Director & Founder
Octopize

Joe Biden's executive order of 02/28/2024 is the most significant measure ever taken by a president to protect the security of Americans' personal data. This decree authorizes the attorney general to prevent large-scale transfers to countries that raise concerns, and provides safeguards for other activities that may give rise to data transfers to other countries. In a US context that is not conducive to such obstacles, this is to say the importance of such an issue. This political choice obliges us to do the same and calls for technical solutions, if possible national, that allow it to be effective. The data avatarization proposed by Octopize addresses this problem, because it meets the specifications of AI (ORIA, for example, essential for our research), ensures the security of French people's personal data and helps reduce our handicap, compared to large American groups.

Philippe Latombe

Member of Parliament for **Vendée**,
Secretary of the **Law Commission**,
Commissioner at the **CNIL**

Summary

INTRODUCTION

A

Confiance.ai[1] is the technological pillar of the Grand Défi "Securing, certifying, and making reliable systems based on artificial intelligence," which was launched by the **Innovation Council**. It is the largest technological research program of the **#AIforHumanity** plan, which aims to make France one of the leading countries in artificial intelligence (AI). It aims to create a **methodological framework** that instills trust for the design and integration of **safe, reliable, and secure AI** in critical systems (e.g., automotive, aeronautics, defense and security, energy, industry). The progress made since 2021 can be consulted through a white paper[2] and scientific publications available in the HAL[3] collection.

In an increasingly **interconnected industrial** environment, data play a key role in **optimizing processes** and **performance**. However, sharing and using these data present new challenges, including data **privacy**. The need to preserve **confidentiality** while allowing its exploitation has led to the **exploration** of desensitization techniques. **Anonymization** refers to a collection of methods used to remove or obscure personally identifiable information, making it impossible to identify individuals, while maintaining a high level of utility.

In the context of **industrial data**, a protection strategy should adapt to the unique challenges of this sector. This white paper presents the **results of work** exploring the prospects of desensitizing industrial data using anonymization methods. This research focuses on **data collected by sensors**, made available by a member of the Confiance.ai program for training anomaly detection models. In addition to this desensitization approach, particular attention is paid to the **anonymization of time series**, a type of data ubiquitous in industrial environments.

Due to its compatibility with several types of data including time series, the **avatar method developed by Octopize** is particularly suited to this study. With this approach, we demonstrate how anonymization can not only **secure data** but also maintain its **usefulness for analysis** and modeling. Thus, this white paper provides valuable insights into combining data protection and industrial innovation, paving the way for the more **secure and efficient management of sensitive information.**

[1] https://www.confiance.ai/
[2] https://www.confiance.ai/contenus-media/
[3] https://hal.science/CONFIANCEAI/search/index/?q=*&rows=30&sort=producedDate_tdate+desc

The contribution of methods for anonymizing strategic data based on avatarization opens up new perspectives: improved protection of privacy, better fairness and proven usefulness for the development of AI systems (machine learning).

Applied to health, this approach facilitates the sharing of sensitive data. In particular, it makes it possible to predict the results of a clinical trial and validate a research project, accelerate the contractualization phases between stakeholders, or even offer possibilities in terms of replacing human beings in interventional research practices.

Octopize takes the step of qualifying and proving their anonymization process to obtain anonymous and statistically relevant summary data. Combined with an AI model supervised by human supervision, avatarization guarantees a robust and ethical analysis, catalyzing innovation and protection.

The mobilization of these avatarization and Human Guarantee tools could offer guarantees of security and AI Act compliance for patients and users, both in terms of data protection and guarantees of trust for health products developed using artificial data.

David Gruson

Home Health Program Director,
La Poste Health & Autonomy
Founder**, ETHIK-IA**

B.

STRATEGIC DATA

B.1 What are we talking about?

Strategic data are specific information considered essential for an organization. Often confidential, strategic data represent a certain value for the organization. It promotes an understanding of the competitive context, market trends, opportunities, or risks. The identification, collection, analysis, and effective use of strategic data are essential to develop and implement successful strategic plans and maintain a competitive advantage in the market.

B.2 Why protect them?

Strategic data may contain **sensitive information** about business objectives, market strategies, innovations under development, financial data, etc. Disclosure of this information to competitors or unauthorized parties may compromise the company's **competitive position**.

Many strategic data are also subject to **strict regulations** regarding confidentiality and data protection, such as the General Data Protection Regulation (**GDPR**), **trade secrets**, or **IG 13100** for classified information. Companies are required to comply with these regulations to avoid being sanctioned.

Beyond the legal framework, protecting strategic data strengthens the **trust** of customers, business partners, or investors in the company, demonstrating its commitment to information security and confidentiality.

Therefore, protecting strategic data is essential to guarantee **competitiveness**, regulatory **compliance**, stakeholder **confidence**, and the **sustainability** of the company, particularly in competitive business environments or intrinsically sensitive areas (e.g., energy, defense).

B.3 What data for what use?

One of the members of the Confiance.ai program owns and operates **production units** equipped with sensors to monitor their proper functioning. As part of the program, this manufacturer sent us the data from these sensors to process a **use case related to anomaly detection**. Given the nature of the sectors concerned, these data are somewhat sensitive, particularly with regard to production processes. Their protection is therefore a strategic issue. This is why the data were previously **de-identified** before being shared.

> *Data anonymization is a central topic for the future of Machine Learning. It is essential to equip ourselves with technologies that protect sensitive data while offering broad exploitation of this data that contains rich information. Accessibility to this data will necessarily strengthen the reliability of Artificial Intelligence systems.*
>
> *The work carried out by Octopize and Sopra Steria described in this white paper helps to remove obstacles to the use of Machine Learning in areas where data confidentiality is essential, in particular health, defense, or energy but also more broadly in all cases where the secrecy of raw data is important. The approach guarantees their security with concrete metrics that are absolutely necessary for a trust-based approach. This last point is one of the great strengths of this work, because not all approaches today are accompanied by such metrics.*

Yves Nicolas

AI group Program Director
Deputy Group CTO
Sopra Steria

ANONYMIZATION, A VECTOR FOR PROTECTING STRATEGIC DATA

C.

C.1 What is anonymization?

According to the **CNIL** definition, anonymization is a process carried out on personal data, which consists of "making impossible, in practice, any identification of the person by any means whatsoever and in an irreversible manner."

The European Data Protection Board (**EDPB**) has defined three criteria that make it possible to ensure that data are truly anonymous.

- **Individualization**: It must not be possible to isolate an individual in the dataset.
- **Correlation** : It must not be possible to link together separate sets of data concerning the same individual.
- **Inference** : It must not be possible to deduce, with near certainty, new information about an individual.

C.2 When should we talk about anonymization?

According to its definition, **anonymization** is specifically applied to personal data. The notion of individual is present in each of the criteria that are individualization, correlation and inference.

Beyond that, anonymization means ensuring that it is **impossible to trace the individual** who created the data. In other words, it is about making it impossible to re-identify the information that generated the data.

Anonymization allows for the broader protection of sensitive, confidential, or strategic data. Thus, anonymization means lowering the sensitivity threshold of the data processed.

C.3 Anonymization adapted to strategic data

Anonymizing strategic data is essential to protect an organization's **confidential information**, while making it possible to use it for **analytical or research purposes**.

Anonymization involves altering data so that it can no longer be directly associated with specific individuals or entities, while still retaining its usefulness for analysis. This may include **aggregating** or **generalizing** data to prevent indirect identification. Ensuring that inference, correlation, and individualization risks are removed helps ensure that privacy risks are reduced.

> *Octopize joined Cyber@StationF, Thales' startup accelerator dedicated to cybersecurity, in 2024. This collaboration represents a major challenge for the processing of strategic and confidential data in Defense. Together, we are exploring the application of their avatar anonymization method to Defense data, in particular for training Machine Learning algorithms. Building on the progress made in the acceleration program, Octopize benefits from the support of Thales technical and business coaches, allowing them to leverage their expertise. They also meet with Thales customers to identify concrete use cases. A proof of concept (POC) is currently underway, and we plan to share the results in a future white paper dedicated to AI and Defense.*

Marine Martinez

Program Lead Cyber@StationF
Thales

In the Defense and Security sector, data remains a central element, both in the planning and conduct of land operations and armament operations, and in the design and manufacture of armed forces equipment by the defense industry.

Beyond that, there are also issues regarding the explosion of data to be processed in theaters of operation, as well as increasing considerations of the use of data from the training of the armed forces, which are increasingly digitalized. Octopize is thus an example of a start-up that has a role to play in the ability of forces and companies to exchange data, among other French technological nuggets supported by the GICAT innovation and start-up accelerator label GENERATE.

Hubert Raymond

Responsible for Innovation and the GENERATE program,
Public Contracts Representative,
GICAT (Group of French land and air-land defense and security industries)

DATA ANONYMIZATION TECHNIQUES

D

D.1 Overview of existing anonymization families

The EDPB defines two main families of anonymization techniques: **randomization** and **generalization**.

◆ (i) **Randomization** is the process of changing attributes in a dataset so that they are less precise, while maintaining the overall distribution. This technique helps protect the dataset from inference risk. Examples of randomization techniques include noise addition, permutation, and differential privacy.

◆ (ii) **Generalization** consists of modifying the scale of the attributes of the datasets or their order of magnitude to ensure that they are common to a set of people. This technique makes it possible to avoid the individualization of a dataset. It also limits the possible correlations of the dataset with others. In generalization techniques, one can, for example, cite aggregation, k-anonymity, l-diversity, or t-proximity.

Each of the anonymization techniques may be appropriate, depending on the circumstance and context, to achieve the desired purpose without compromising the right of the persons concerned to respect their **private life**.

D.2 Octopize avatar method

D.2.1 Principles

The **avatar method** is a unique approach to generating synthetic anonymous data, where the structure and statistical relevance of the original dataset are preserved while maintaining the confidentiality of the data. This technique uses an **individual-centric** approach by creating local simulations based on the individual, which makes the simulation of an avatar unique. The avatar method is designed to meet the three criteria defined by the **EDPB** to assess the robustness of an anonymization process.

Compared to other techniques such as decision trees and Generative Adversarial Networks (**GANs**), the avatar software demonstrates similar utility in preserving the structure and statistical relevance of the original dataset. In addition, the avatar software includes **privacy measures** that allow the protection afforded to anonymized data to be assessed against the three criteria defined by the EDPB.

The avatar method takes original data as input and produces **synthetic and anonymous data** of the same size and nature. For example, numerical data remain numeric, categorical data remain categorical, and so on. The core of the method is illustrated in Figure 1 and described below.

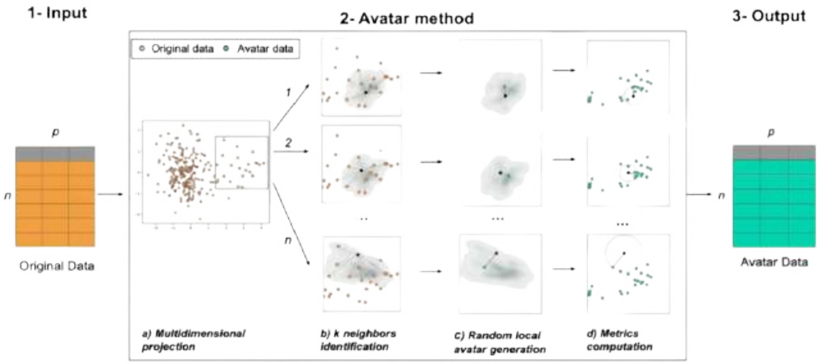


Figure 1: Principles of the avatar method.

[1] Article on the method in Nature Digital Medicine : https://www.nature.com/articles/s41746-023-00771-5

D.2.2 Multidimensional projection

The original data are projected into an appropriate **multidimensional space** using dimension reduction techniques such as factor analysis of mixed data (FAMD), principal component analysis (PCA), or multiple correspondence analysis (MCA). The transformations used must be **reversible**, i.e., there must be an inverse transformation that allows returning to the original representation space.

This step transforms individuals, which are initially described by several numerical and categorical features, into **structured numerical coordinates** that facilitate the calculation of distances between individuals. It also reduces the dimensionality of the dataset in order to highlight the most relevant information.

D.2.3 Calculating k-neighbors

Neighbor distances are then calculated between all points in this space to apply the **k-nearest neighbor (KNN) algorithm**. This creates a local area around each coordinate - each being the projection of an individual from the original data - by defining its nearest neighbors.

D.2.4 Random generation of local avatar data

For each of these local areas, a single simulation is pseudo-randomly drawn, creating a new coordinate within the area, which we call the avatar of the original coordinate. This simulation is influenced by the distance between the original point and each of its neighbors, by a **random weight** following an **exponential distribution** and by a **random contribution** factor for each neighbor.

This allows non-deterministic simulations to be considered an irreversible process, which is a necessary condition for preserving confidentiality.

D.2.5 Reversing the transformation to return to the original encoding

Once a synthetic dataset has been generated for each individual, the avatar coordinates are reversed back to the original encoding, preserving the type of the original attributes (e.g., categorical, numeric). Although it is not possible to recover the original data from the avatar data, **the structure of the dataset is preserved** as illustrated in Figure 2.

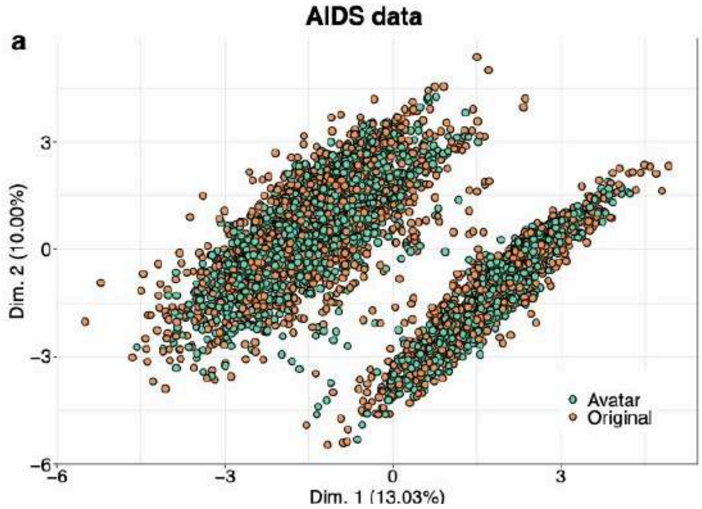


Figure 2: Conservation of the structure of the dataset after anonymization: The avatar points globally covered all of the original points, with the exception of isolated points representing extreme individuals. The two dimensions of the FAMD that explain the largest proportion of variance are represented on the axes.

D.2.6 Calculation of data privacy parameters

The avatar software includes anonymized data **privacy measurement metrics** that are essential to prove the protection provided to the data. The **anonymization report** detailing these metrics, and automatically generated by the software, constitutes a real risk analysis. The metrics aim to cover different attack scenarios and meet the three EDPB criteria: individualization, correlation, and inference.

The first family of metrics aims to evaluate the protection of a dataset against **individualization** attacks.

These attacks can take different forms, requiring different complementary measures. Some individualization measures are model-independent and can therefore be used on any pair of original and processed datasets. Other metrics require temporarily maintaining a link between the original and processed individuals.

Below are three examples of individualization metrics automatically calculated by the **avatar software**:

◆ **Distance To Closest**. To calculate the distance to closest (DTC), the distance between each synthetic individual and its closest original is measured. The median value is kept to have a single representative value associated with this measure. The reasoning behind the DTC is that if each synthetic individual is close to an original, the dataset could be at risk of individualization. However, a low DTC does not necessarily mean that there is a risk; therefore the Closest Distances Ratio (CDR) should be measured to complement it.

◆ **Closest Distances Ratio**. Similar to DTC, the CDR is calculated by first measuring the distance between an avatar and its closest original individual, divided by the distance to its second closest original individual. In other words, the distance between the two closest original individuals is measured. If the ratio is high, the two closest originals are at the same distance and it is therefore impossible to distinguish them with certainty in practice. From the ratios calculated for each processed individual, the median is kept to provide a single CDR value. There is a risk of individualization when both DTC and CDR are low.

◆ **Hidden Rate**. The Hidden Rate is the probability that an attacker will make a mistake in linking an individual to their most similar avatar (synthetic individual). This is where the link between the original and the avatar that has been temporarily kept becomes useful.

◆ **Local Cloaking**. To obtain the Local Cloaking, the number of avatars between an individual and the avatar that the person generated is calculated for each of the individuals. The Local Cloaking is the median value obtained. Note that the Hidden Rate and the Local Cloaking are linked since the Hidden Rate represents the number of individuals for whom the avatar of an individual is the closest avatar of this individual.

The second family of metrics meets the **correlation** criterion. These metrics respond to a common and probable attack scenario.

The attacker has a processed dataset and an external identification database (e.g., a voter register) containing information in common with the processed data (e.g., age, sex, postal code). The more information in common between the two databases, the more effective the attack will be.

The **Correlation Protection Rate** is a metric that measures the percentage of individuals that would not be successfully linked to their synthetic counterpart if the attacker used an external data source. Variables selected as common to both databases are likely to be found in an external data source.

To cover the worst-case scenario, we assume that the same individuals are present in both databases.

In practice, some individuals in the anonymized database are not present in the external data source and vice versa. This metric also relies on the fact that the link between the original and the synthetic is temporarily preserved. This link is used to measure how many matches are incorrect.

Metrics that meet the **inference** criterion correspond to a different type of attack; the attacker seeks to infer additional information about an individual from the available anonymized data.

The inference metric calculates the possibility of inferring, with a significant probability, the original value of a target variable from the values of other processed variables. The inference metric can be used on **numeric and categorical targets**. When the target is numeric, it is called a regression inference metric and the protection is evaluated as the average absolute difference between the value predicted by the attacker and the original numeric value.

> *For Air Liquide Healthcare, the protection of personal data is a major responsibility for the reasoned and compliant use of information. Enabling the investigation potential to be expanded by generating anonymized datasets with avatars opens up immense prospects, particularly in the field of Health. The definition of algorithms for predicting compliance using AI makes it possible to adapt patient support plans in the treatment of their chronic disease or to help with the early diagnosis of the progression of the disease. These are concrete examples for which data anonymization makes these innovations possible for the benefit of the patient.*

Olivier Gruet

Programs Director & Chief Data officer
Air Liquide Healthcare

On the other hand, we talk about classification inference metrics when the target is categorical and the protection level is represented by the prediction accuracy.

The metrics detailed above are just a glimpse of the full set of metrics made available in the **anonymization report automatically generated by the avatar software**. Such a methodology allows the generation of anonymous datasets with a fully explainable model and concrete privacy measures that allow measuring the degree of protection.

D.3 Application of the avatar method for this use case

The avatar method, as described above, is an **anonymization method**. Therefore, it aims to **protect** the individual at the origin of the data, since anonymization applies to personal data. However, and as indicated above, the method can be understood in the sense of **desensitization** of confidential or strategic data. In this case, the individual, basically represented by a line in a dataset, can be assimilated to a machine, a sensor, or even a space-time depending on the nature of the dataset that we wish to anonymize.

In a general context, anonymizing data requires having a notion of the individual or entity in the data, since anonymization modifies the data to protect these individuals by hiding them among their respective neighbors. For this industrial use case, the notion of individual was not defined since the data of a variable related to a single machine. The approach chosen to allow the anonymization of the dataset consisted of d**ividing the dataset into time ranges**, each being assimilated to an individual. This choice of segmentation meets a business need. In other contexts, this segmentation can be different or even very often natural depending on what one wishes to protect.

For example, data from a machine shared by multiple users can be naturally divided into segments representing distinct user sessions. Once anonymized, these segments still represent user sessions, but it is impossible to re-identify them or the user to whom they are attached. Another use case may lead us to anonymize data from multiple machines of the same type (e.g., respirators). In this case, the entity to be protected is the user of each machine (whether a company or an individual). Therefore, segmentation can naturally be done by machine identifiers or session identifiers.

The data used for the analyses in this document were from sensor readings during continuous operations in an industrial environment, which refer to measurement sequences, forming a **time series**.

Time series differ from so-called **tabular** or **static data** in that the relationship between successive readings is an integral part of the data. This relationship allows **trends** or **changes** over time to be defined (e.g., pressure or temperature increases). The main principles of the avatar method described so far can be used on time series data. In fact, the steps of projection, neighbor calculation, and generation of synthetic coordinates remain consistent.

However, the **type of projection** differs between tabular data and time series. The **tabular projection** approaches used in the avatar method are PCA and its derivatives. Their use on time series would have the effect of losing all information related to the sequencing of points. It is therefore necessary to use a **projection or transformation method** specific to time series. There are several techniques of this type such as Fourier transforms, discrete cosine transforms, or wavelet decomposition. These approaches have already been used in the context of avatar generation in medical contexts [2].

There is also an adapted version of **PCA** for the functional domain, allowing one to model one variable as a function of another. This functional PCA (FPCA) can be applied to time series, because these data represent a function that translates the evolution of a variable over time. This method is also ideal for dimension reduction, which made it our preferred choice. In addition, similar to classical PCA, FPCA allows an **inverse transformation** to be performed, allowing return from the coordinates to the initial variables, as in the original dataset. For more details on FPCA, we recommend the **article by Wang et al.** [3].

In practice, the data may include several temporal variables, which may have different **sampling frequencies or be periodic**. In addition, these variables are often accompanied by **fixed data**. The avatar method is compatible with this type of context. The different steps allowing it are detailed in Figure 3.

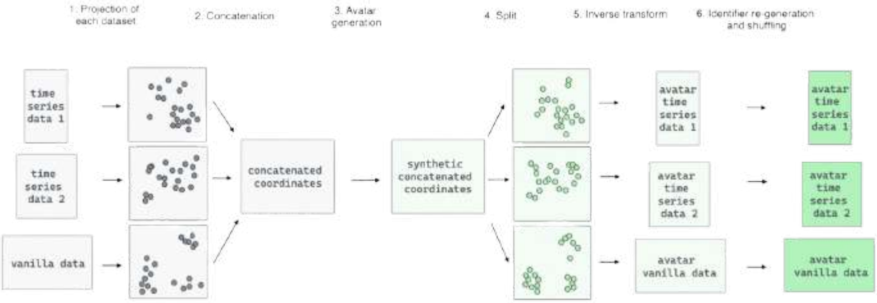

1. Projection of each dataset
2. Concatenation
3. Avatar generation
4. Split
5. Inverse transform
6. Identifier re-generation and shuffling
time series data 1
time series data 2
vanilla data
concatenated coordinates
synthetic concatenated coordinates
avatar time series data 1
avatar time series data 2
avatar vanilla data
avatar time series data 1
avatar time series data 2
avatar vanilla data

Figure 3: Different steps in anonymizing mixed data including time series and static data.

First, each of the datasets is projected into a **digital space** (Step 1). As aforementioned, temporal data are projected with FPCA, while static variables (denoted vanilla in Figure 3) are projected with PCA or its derivatives. Second, to anonymize all of these data in a single step, the coordinates obtained during the different projections are **concatenated** (Step 2). Then the process of **generating synthetic coordinates** is applied (Step 3), and these synthetic coordinates are redivided by **dataset** (Step 4) to allow the **inverse transformation** (Step 5). The sequence of these steps gives as many datasets as input.

The example presented in Figure 4 allows us to visually see the result of anonymizing time series with the avatar method. In particular, we see that the **trends and global characteristics** are preserved but that certain sequences of values specific to a single entity are not. This highlights good conservation of the **signal** as well as a contribution to **privacy**.

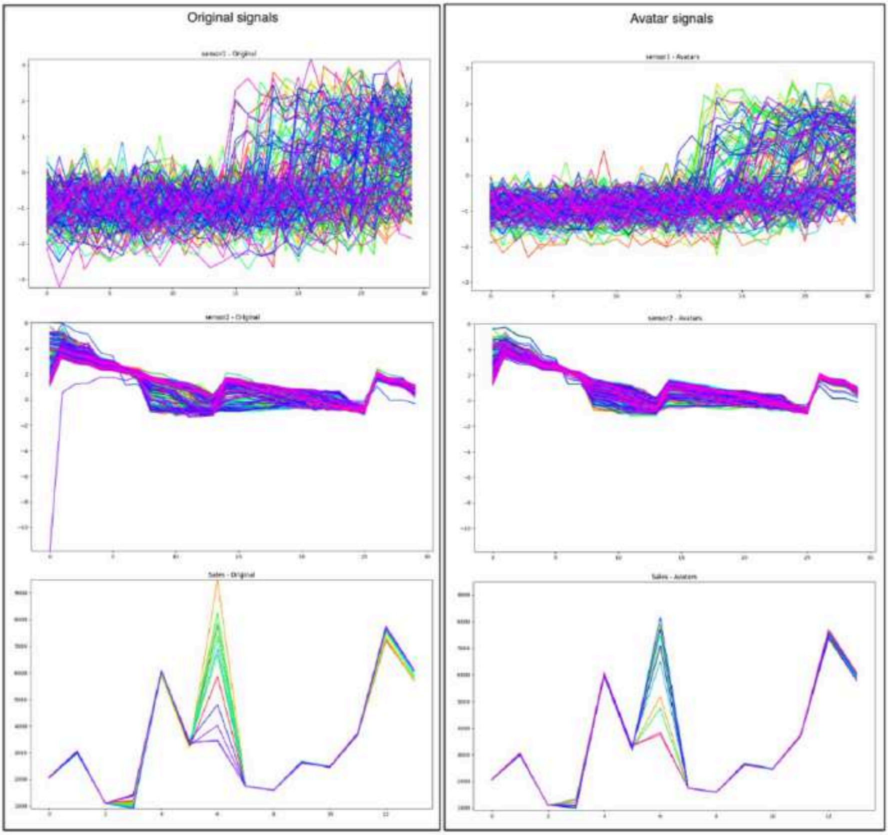


Figure 4: Examples of 3 original time series variables (left) and their avatars (right).

Simply removing directly identifying personal data fields (name, first name, etc.) from a dataset does not always allow anonymization of the dataset. Indeed, the more fields it contains that are filled in at the individual level, the more people will be identifiable by cross-referencing the different characteristics of the different columns, as in the game Who is it?
In this case, the avatarization of the dataset makes sense. This consists of generating a new avatar dataset in which no data point from the initial dataset will be found. The avatar point cloud is designed to preserve the statistical properties of the initial point cloud, while adapting to the specific requirements of the use case. By generating a synthetic dataset, it opens the way to statistical exploitation while minimizing the risk of exposing sensitive data. As such, it can enable a paradigm shift in data management.

Alexis Rouet

Chief Data Officer HR
Renault Group

E.

PRODUCTION OF ANONYMIZED DATASETS

E.1 Choice of perimeter

Although the avatar method can be applied to multiple sensors, this white paper focuses on the **results obtained with the one-dimensional convolutional neural network (1D-CNN)** anomaly detection algorithm, applied to a sensor already analyzed in previous Confiance.ai work. The data from this sensor, previously made non-identifying, were provided by the manufacturer.

The sensor data covered a **period of 1 year**. For training the 1D-CNN model, the data are divided into training and testing sets, and within each set, the data are further divided into smaller segments. Each of these segments represents a sample for training the model.

There are several ways to segment time series. In this work, three approaches were considered.

◆ **Segmentation by fixed and exclusive time window**. With this segmentation, the time series is divided into several series of predefined size (e.g. 7h). Each time step is represented in only one segment (exclusiveness).

◆ **Peak detection segmentation**. For this segmentation, a peak detection step is applied to determine the time steps where the signal has the highest values. Segmentation is done around these peaks so that it is at the center of the segment and to obtain a segment of predefined size (e.g., 7 h). Peak detection segmentation is only recommended on cyclic data. This method allows potential correlations to be preserved between time steps and signal values and is com 19 patible with dimension reduction.

◆ **Sliding window segmentation**. Finally, the third segmentation method uses a sliding window of predefined size that is shifted by a certain number of time steps between two segments. This shift is controlled by a stride parameter. Depending on the stride used, this segmentation can potentially generate a large number of segments (at stride=1, the number of segments is maximum).

Since the target anomaly detection model was an **NN**, large data volume was preferred to ensure good results. For this reason, **sliding window** segmentation was chosen for the present analysis, with stride=1. **Peak detection** segmentation was nevertheless used for some illustrations.

Since the segments were modified during anonymization, it was no longer possible to group them together to return to data representing the evolution of these variables over the entire year. However, these segments were perfectly suited for use in **training models**.

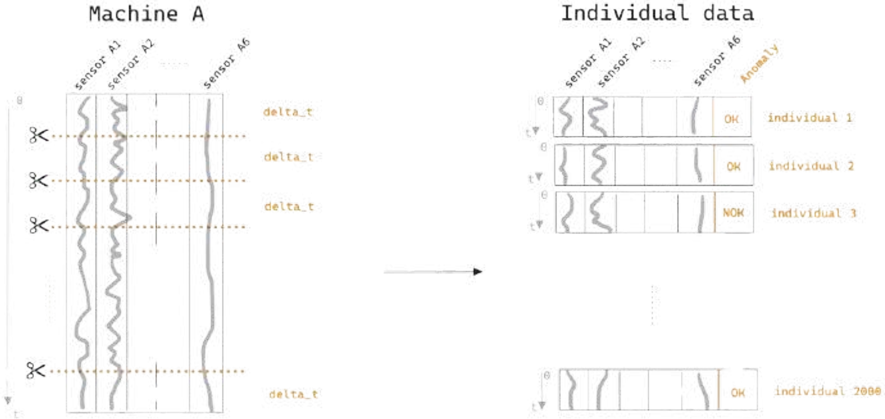


Figure 5: Segmentation of annual data from a machine or system into individual entities (segments). In this example, 1 year of data generated 2000 individuals who could be anonymized.

E.2 Constraints encountered

Data anonymization is a process based on **data modeling**, just like anomaly detection. Generally speaking, data anonymization must be carried out on **populations of individuals** who are part of the same context and who can therefore be compared with each other. What is valid for individuals is also valid for data from machines.

To fit into the same **context**, temporal data must be defined over the **same time range**. Thus, a **time normalization** step is applied before anonymization. This step redefines each segment over a time range from t=0 (start of the segment) to t=1 (end of the segment). Thus, once normalized, all segments can be **compared** with each other as illustrated below.

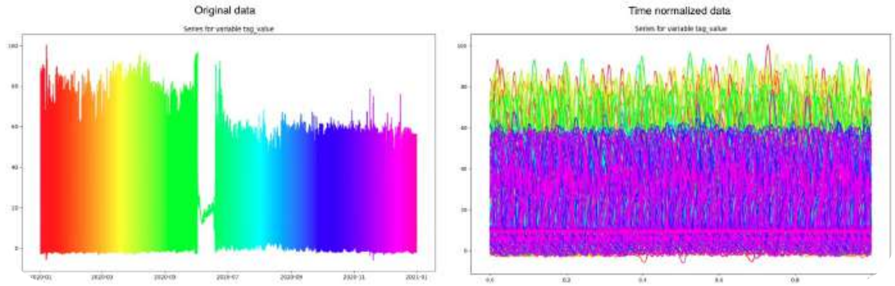


Figure 6: Normalization of time steps for Confiance.ai program member data. Each color represents a separate entity. The original time range covers a full year. Following normalization, all points are in the time interval [0, 1].

A second constraint must be recognized. This is specific to the **FPCA method** used to project time series. FPCA requires that the time series be aligned; that is, that they have **measurements made at the same time steps.** The curves of a variable will therefore have to be resampled if the measurements are not made at the same frequency or generally at the same time steps.

To this end, a number of periodic time steps was defined between t=0 and t=1, and the sensor values were inferred for each. **Linear inference** was used. Note also that the number of time steps cannot exceed the number of distinct entities in the data to be anonymized, another constraint of FPCA. Therefore, **loss of precision** on data containing few entities is expected. However, this was not the case on data segmented by sliding window (and stride=1).

E.3 Assessment of data protection

The avatar method makes modifications to the data in a way that changes each entity distinctly to **maintain maximum utility** while providing privacy. The method is configurable to allow its users to choose the **privacy level** that suits their use case. Because the method is **configurable**, it is necessary to validate the datasets produced by calculating privacy metrics.

Note that the **calculation of privacy metrics** is recommended for the validation of any anonymized dataset, whatever the method, including methods presented as **private by design,** which are also configurable and can therefore result in non-anonymous data.

The need to implement trusted AI solutions is a major challenge for SMEs and, more broadly, for all companies wishing to maintain or increase their productivity and growth. Today, AI is no longer an asset: it is now a necessity.

Trusted AI is all about trusted data. Developing custom solutions or building on existing ones requires understanding the industry challenges, but also ensuring the quality, reliability, and confidentiality of the shared data. This is precisely what the avatar method allows.

This is the key to:

- *reassure business partners in their operationalization projects,*
- *protect citizens by complying with regulatory frameworks and international regulations,*
- *and accelerate the deployment of trusted AI solutions within businesses.*

Only trusted data will ensure the success of AI projects while meeting security, confidentiality and transparency requirements.

Marie-Pierre Habas-Gerard

Director - Industrial Consortium in Industrial Artificial Intelligence
Confiance AI Canada

The **use case** exclusively involved temporal data; only metrics meeting the GDPR **individualization** criterion were evaluated. The state of the art on privacy metrics in the context of time series does not allow for the identification of metrics. However, it was possible to use the individualization metrics presented in D.2.6 because they were based on coordinates in a **projected space**. Since we used a projection (FPCA) to process the time series, it was possible to calculate the **Hidden Rate**, **Local Cloaking**, and any other metric calculated on coordinates.

The privacy metrics obtained on an anonymization run with **k=20** are presented in Table 1. As a reminder, k is the parameter that most controls the **signal preservation/privacy trade-off.** The objectives indicated for each of the metrics are given as an indication and represent a **correct level of anonymization** for most use cases. In practice, privacy objectives must be **set according to the use case being treated**. For example, privacy objectives must be higher when the aim of anonymization is to release data as open data than for a use where the data needs to be shared between countries but internally. The risk and impact associated with a potential data leak is higher in one case than in the other.

Privacy Metrics	Measured value	Average goal
Hidden rate	93.9%	>90%
Local cloaking	10	≥5
Distance to closest	10.0	>0.2
Closest distance ratio	0.83	>0.3
Row direct match protection	99.9%	>90%

Table 1: Privacy metrics measured on 1 anonymization run with the avatar solution from the dataset of a member of the Confiance.ai program with k=20.

The **metrics** presented here respond favorably to the **individualization** criterion of the GDPR.

To further understand the contribution of the avatar method, it is possible to **compare and interpret the data before and after anonymization**. To facilitate this comparison, we used the same sensor data but segmented with a **peak detection method**. This had the effect of aligning the curves with the most prominent peak at t=0.5. It was thus easier to **identify general trends** in the signals but also to **identify rare and potentially re-identifiable signals.**

Figures 7 to 10 represent **four separate anonymizations** (for 4 different months of data). They illustrate in particular the fact that **rare signals were corrected** so as to no longer reveal rare phenomena. In some cases (Figure 9) where several modes were present, we observed that these different modes were **preserved**. This is only possible if enough signals follow these modes of operation, thus making them sufficiently frequent and **removing any risk of re-identification**.

> *For several years, Nantes has been committed to ethical data management, whether in its metropolitan data charter or more recently through its doctrine to establish a regulatory framework for the use of AI. Trust, transparency and control of public data are values that the community strives to translate on a daily basis in its many projects. In this context, being able to reconcile statistical quality of data and the imperatives of protecting users' personal data is a central issue. The work carried out by Octopize on the anonymization of sensitive data with avatars makes it possible to concretely advance this issue by helping to open up new perspectives for developing projects of general interest serving, for example, global health or reducing energy consumption, which are both respectful of people and contribute to trust in these devices.*

Francky Trichet

Vice-President, **Nantes University**
(Responsible Digital and New Uses)
@franckytrichet

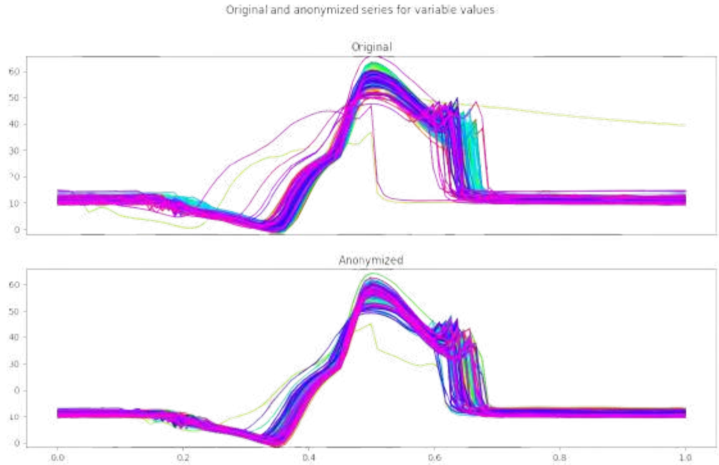


Figure 7: Anonymization of May signals aligned with peak detection: rare signals were no longer discernible in avatars.

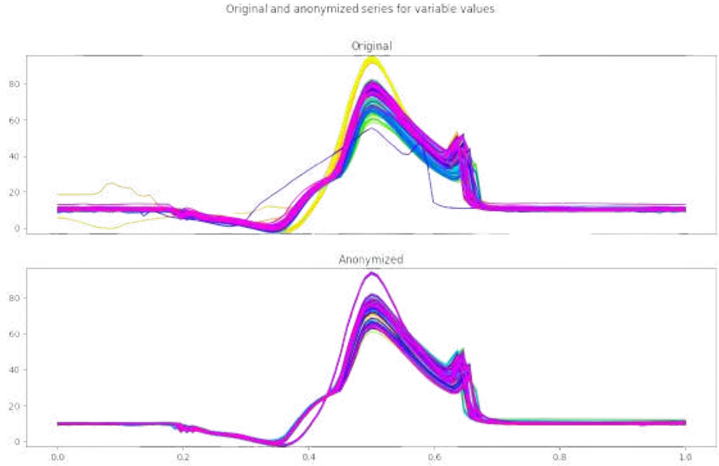


Figure 8: Anonymization of July signals aligned with peak detection: rare signals were no longer discernible in avatars.

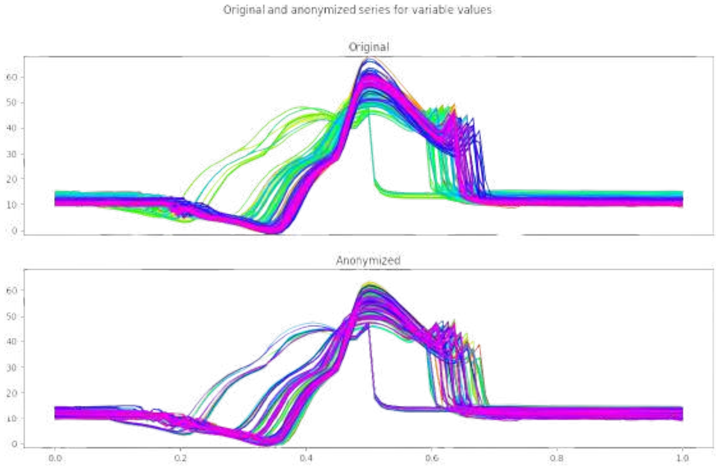


Figure 9: Anonymization of August signals aligned with peak detection: two modes of operation were preserved in the avatars.

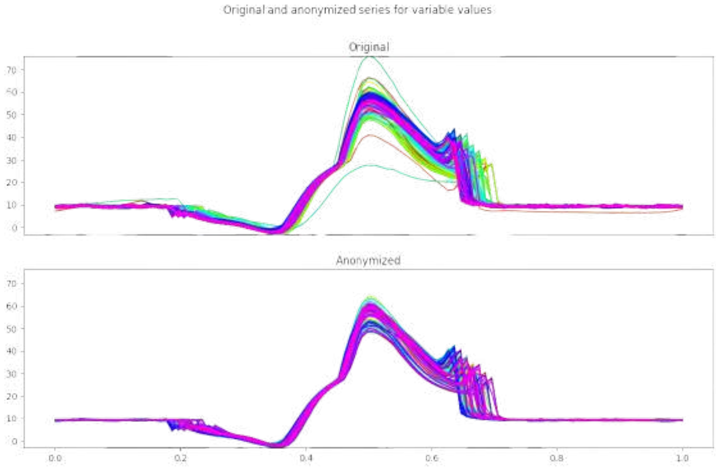


Figure 10: Anonymization of December signals aligned with peak detection: only one mode of operation was kept in the avatars.

Anonymization is a key issue in the large-scale manipulation of initially personal data. Indeed, as a process that makes it almost impossible to re-identify the persons concerned, anonymization makes it possible to overcome many constraints imposed by the GDPR while ensuring a certain secrecy of the initial data set. In terms of Machine Learning in particular, this makes it possible to remove many obstacles in the creation of training data sets for AI models. Indeed, there can be many constraints in order to be able to lawfully use sets of personal data, whether sensitive or not. However, failure to comply with these constraints can have multiple harmful consequences that can go as far as making the illegally trained AI model illegal.

The work carried out by Octopize and Sopra Steria described in this white paper aims to demonstrate how anonymization through the avatar method developed by Octopize "can not only secure data but also maintain its usefulness for analysis and modeling". These parameters are obviously essential to the development of a trusted AI that complies with regulatory requirements.

Valérie Aumage

Head of IP/IT/Data Privacy
Lawyer
PwC

E.4 Generic assessment of the maintenance of the statistical value of data

In addition to the **privacy** assessment associated with the anonymized dataset, **generic utility metrics** can be calculated. Since they are generic, these metrics do not require knowing or simulating the future use of the data, which is potentially costly in terms of computation time (e.g. use in machine learning). As a result, they allow for rapid iteration when **setting up** the avatar solution.

The table below summarizes the utility metrics calculated on the transmitted data. The metrics were divided into two families: **global metrics**, calculated on all points, all series combined; and **metrics calculated at the individual level**, i.e., on each series.

Utility Metrics	Measured value	Objective
Global Metrics		
Pointwise Hellinger distance	0.01	<0.2

Individual metrics (% difference)		
Series mean	0.00%	<5%
Series minimum	0.01%	<10%
Series maximum	0.00%	<10%
Series sum of values	0.00%	<5%
Series Entropy (20 bins)	0.00%	<5%
Autocorrelation (10)	0.00%	<5%

Table 2: Utility metrics measured on one anonymization run with the avatar solution from the dataset of a member of the Confiance.ai program with k=20.

The **Hellinger distance** is a distance between two distributions. The larger its value, the more different the distributions compared. For a **pointwise** Hellinger distance, the distribution of the values measured is compared with its equivalent constructed with the anonymized values. Although given for information purposes only, a Hellinger distance less than 0.2 is considered an indicator of **good signal conservation**.

Individual metrics were calculated from **features** generated on each series. The values were **averaged** and the **relative difference** between the original and avatar values was expressed as a percentage. The indicators extracted from the data were: the **mean** of a series (series mean); its **minimum** (series minimum) and its **maximum** (series maximum); the sum of its values (series sum of values); its **entropy** (series entropy, which can be interpreted as representing the complexity of the series); and its **autocorrelation** (correlation between points of the same series).

The indicative **objectives** for the differences observed on these indicators were **achieved**. Finally, a **visual interpretation** of the generated data confirmed the **conservation of the trends** present in the original data as shown in Figures 11 and 12.

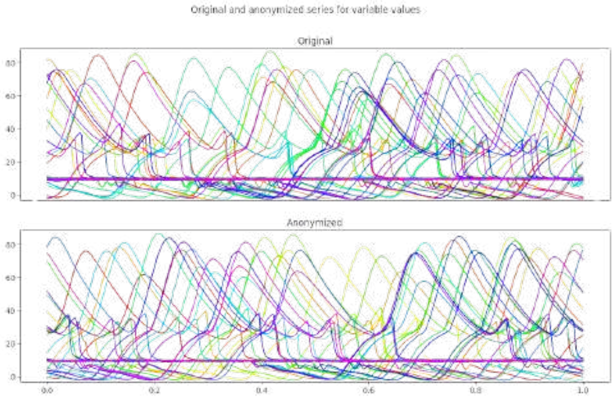


Figure 11: Set of original curves (top) and avatars (bottom) for the data of a member of the Confiance.ai program. Fifty curves are represented for readability reasons.

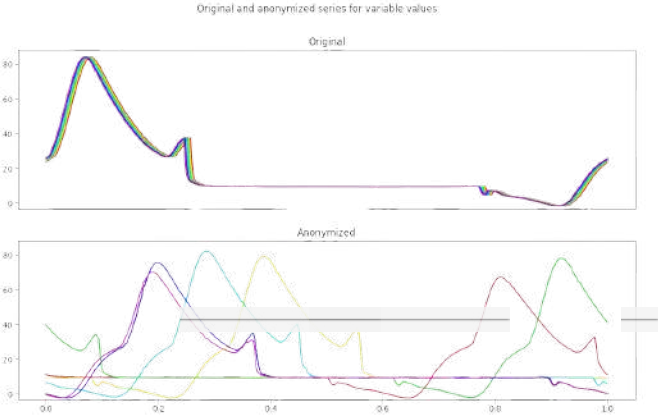


Figure 12: Selection of the first six original curves (top) and avatars (bottom) for the data of a member of the Confiance.ai program.

Generic metrics help **identify potential utility losses** in the avatar solution configuration phase and provide an initial estimate of signal retention. However, it was recommended to continue **evaluating the utility** of the anonymized data by ensuring that it was usable for the targeted use case. Therefore, the following section focuses on **learning anomaly detection models** from this data.

The development of new drugs and treatments very often involves the reuse of data initially collected in the context of healthcare or previous research projects. Access to this data is subject to a strict regulatory framework, compliance with which justifies the use of technologies that strengthen the protection of personal data.

Synthetic data generation solutions including the "Avatar" method developed by Octopize aim to protect data integrity while guaranteeing data confidentiality.

This method represents significant progress, making it possible to move outside the scope of the General Data Protection Regulation (GDPR), to lift the boundaries of data transfer outside the European Union while reducing operational costs and the time associated with obtaining repeated consents.

Gregory Collet

Early Stage Success Manager, certified DPO
My Data-Trust

As the lead of the Climate Links Initiative, a growing global consortium dedicated to connecting local municipalities with SDG-relevant technologies, I recognize the immense value of the anonymization techniques presented in this white paper. Our project relies on the ability to analyze large datasets on municipal needs, sustainable technology offerings, and policy contexts. The "avatar" method, with its rigorous approach to generating synthetic data while preserving statistical relevance, offers a powerful solution. Anonymization will allow Climate Links to build and analyze its knowledge graph without compromising stakeholder privacy or revealing sensitive business information. By adopting these techniques, we can build trust among our partners, accelerate the matching of sustainable solutions to local needs, and ultimately contribute to the effective implementation of the United Nations Sustainable Development Goals. The Climate Links consortium looks forward to integrating this technology to connect local actors and global suppliers in a secure, ethical, and efficient manner.

Newton H. Campbell Jr.

Senior Adjunct Lecturer,
UNSW (University of New South Wales), Sydney

VALIDATION OF LEARNING MODELS ON ANONYMIZED DATA

F.

F.1 Anomaly detection model chosen

The method used to assess the relevance of anonymized data was an **unsupervised anomaly detection approach** from the Confiance.ai program. It operates on regularly sampled **univariate time series**. It is based on a two-step method using deep **1D-CNN architectures**.

◆ **The representation learning stage**: Using pretext tasks to learn a representation of so-called "normal" data samples, i.e., without anomalies, in a self-supervised manner. In this stage, the model learns to reconstruct the data.

◆ **Anomaly detection**: An anomaly is detected whenever the anomaly score of the tested data sample is greater than a threshold, i.e., when the signal reconstruction is of insufficient quality. Anomaly scores were calculated for each point of a data sample. To have the most robust score possible, for a given point, the anomaly score was calculated as the average of all reconstruction scores associated with the windowed data samples containing this point. This way of calculating the score also made it possible to propose a temporal location of the anomaly in the window considered.

This mode of operation was chosen because it is considered one of the **most mature approaches** of the Confiance.ai program in terms of unsupervised anomaly detection at the time of the experiment. The works [4], [5] and [6] mention similar approaches.

There are a few important points to note in applying the method:

◆ Training should be performed on **data that is assumed not to contain anomalies**,

◆ The method is particularly effective in the case where the signal is periodic and the window size is chosen equal to the **period of the signal**,

◆ As with many signal processing learning methods, it is recommended to perform a **sequence cutting** with overlap to increase the learning dataset in size but also in diversity,

◆ Input data are **standardized**.

The objective is to obtain two models: a model trained on an **original dataset** (which we will call the Original model) and a model trained on **avatar data** built from the same original dataset (which we will call the avatar model). This last point is important since it is then necessary to **anticipate the signal segmentation step** with overlap, which must be carried out before anonymization.

CIVITEO teams support many local administrations in their data management. All of them (municipalities, inter-municipalities, departments or regions) use increasingly massive data to fulfill their public service missions on a daily basis. But all are attentive to the protection of privacy as well as the carbon footprint of their digital tools. The use of avatars, anonymous synthetic data, clearly represents a fantastic prospect, in particular for training systems that will use machine learning in areas as varied as energy or water management, waste management, travel but also social action, education or even health prevention.

Jacques Priol

President, data & AI expert,
Author and speaker,
Cabinet CIVITEO

F.2 Results of comparative model training tests

The evaluation of avatar data is done by training the **1D-CNN model** once on original data and a second time on anonymized data. Figure 13 summarizes the procedure.

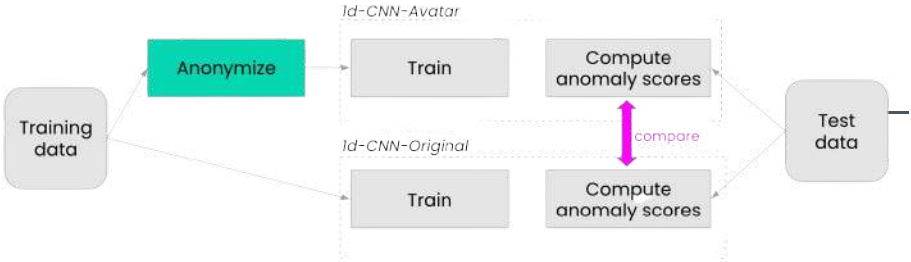


Figure 13: Training and comparison of anomaly scores produced by 1D-CNN trained on original data and anonymized data, respectively.

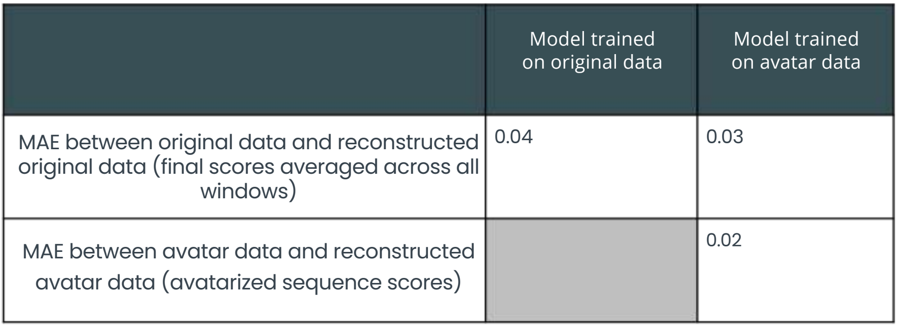
A **training** data set is chosen from the available dataset. This dataset is anonymized and used for training the model, just like its non-anonymized version. The 1D-CNN **parameterization** is the same for both trainings.

Data distinct from the training data were chosen as **test data** to evaluate the quality of the models. Note that these test data were **de-identified** (or pseudonymized). Thus, the result of the evaluation allows the **use of anonymous data** to train models subsequently applied to real data.

This is made possible by the fact that the **avatar method preserves the statistical properties** of the dataset, which is essential for an NN. Each of the models is used to calculate **anomaly scores** for the test data.

Since the training of the 1D-CNN is non-deterministic, 10 training runs were performed to produce **confidence intervals** (CIs).

The measure used to evaluate the effectiveness of the reconstruction by the 1D-CNN model is the **Mean Squared Error** (MAE). The training was performed on 1 month of data and launched 10 times. The scores presented in Table 3 are averaged over the 10 launches.



	Model trained on original data	Model trained on avatar data
MAE between original data and reconstructed original data (final scores averaged across all windows)	0.04	0.03
MAE between avatar data and reconstructed avatar data (avatarized sequence scores)		0.02

Table 3: Reconstruction error (MAE) of 1D-CNN trained on original data and avatars.

We then noted the following:

- The two models **converged**,
- The reconstruction of avatars by the avatar model was a **little better** than the reconstruction of the original data by the original model, and
- The reconstruction of the original data by the avatar model was **slightly better** than the reconstruction of the original data by the original model.

These last two observations could be interpreted using the **anonymization process** itself. As partially explained above, anonymization seeks to **model the modes of the original distribution** and generate individuals who correspond to these modes. In fact, the process should tend to **erase marginal behaviors** and somewhere already achieve a form of **modeling of normality**. This could facilitate the work of learning.

We present in Figure 14 the **anomaly scores** for the two models on all data of the period covered. The results show the 95% CIs for these scores, while Figure 15 shows these results for the **month of May** only, which had the most anomalous points.

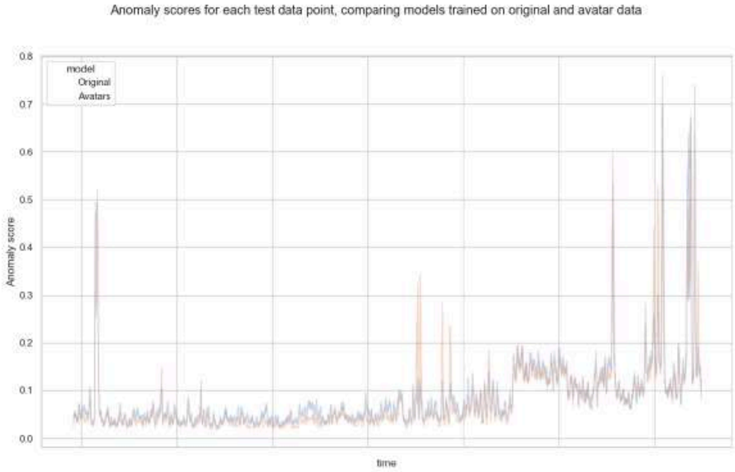


Figure 14: Comparison of anomaly scores obtained over 10 runs for the entire data set.

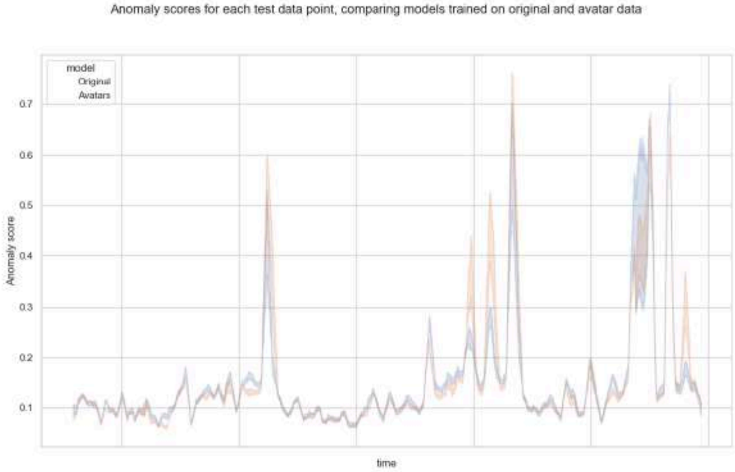


Figure 15: Comparison of anomaly scores for the month of May obtained over 10 runs.

Finally, Figures 16 and 17 present the **scores** obtained over one iteration as well as the **signal** around two known and visible anomalies in **January and May**, respectively. These two anomalies, confirmed by business experts, were identified in a relatively **similar** manner.

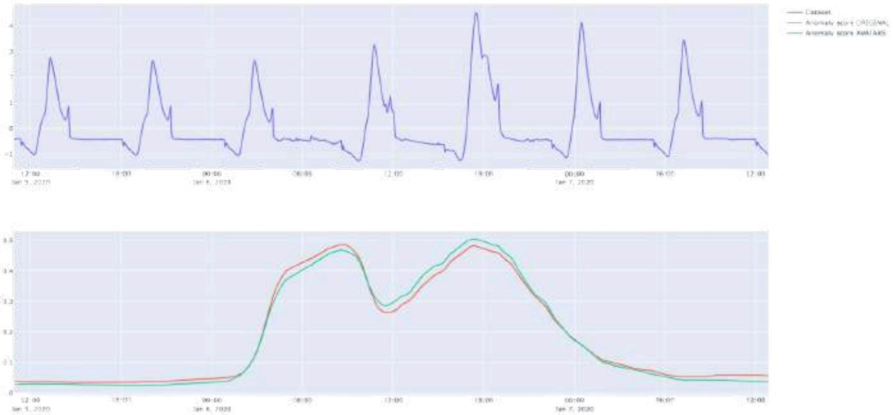


Figure 16: Anomaly scores around a known anomaly from January: the anomaly resulted in high original and avatar scores.

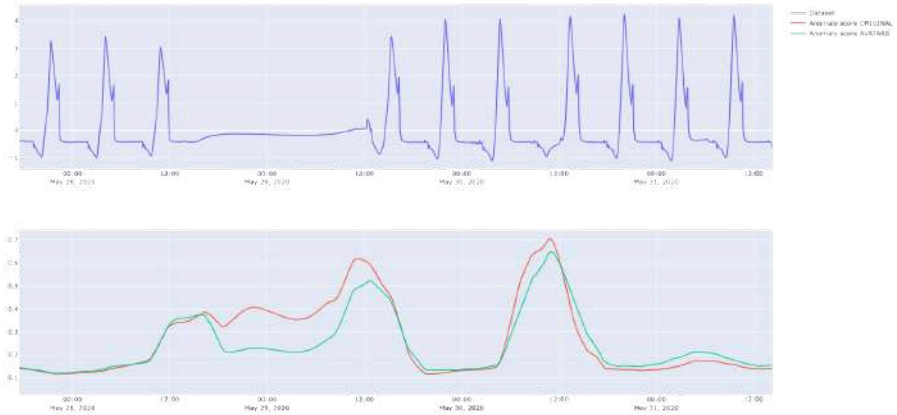
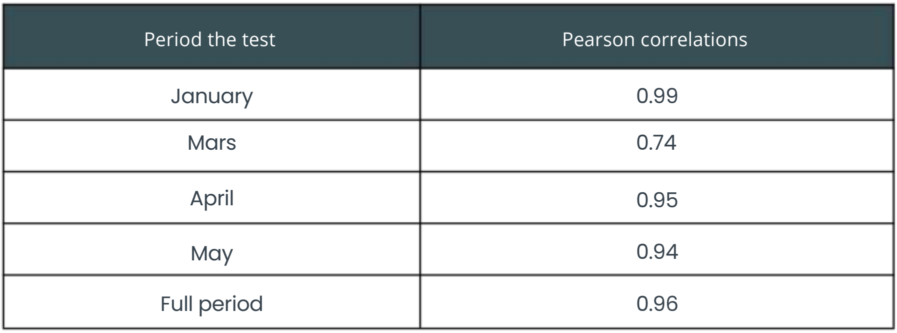


Figure 17: Anomaly scores around a known anomaly from May: the anomaly resulted in high original and avatar scores.

We observed the following.

- Regardless of the model, there were **different operating modes** over time. Thus, the scores obtained in April and May were significantly higher than those in January or February. This should be taken into consideration when interpreting the results.

- A **correlation between original scores and avatar scores** was confirmed by calculating Pearson coefficients (see table and figure below). The closer the correlation is to 1, the more the scores tend to evolve in the same way. In this case, all correlations were greater than 0.7, the threshold beyond which a correlation is generally considered to be strong. Apart from the month of March alone, the correlation coefficient was above 0.9, reflecting a very strong correlation.



Period the test	Pearson correlations
January	0.99
Mars	0.74
April	0.95
May	0.94
Full period	0.96

Table 4: Correlations between original anomaly scores and avatars over different periods: the coefficients were above 0.7, which represented a strong correlation.

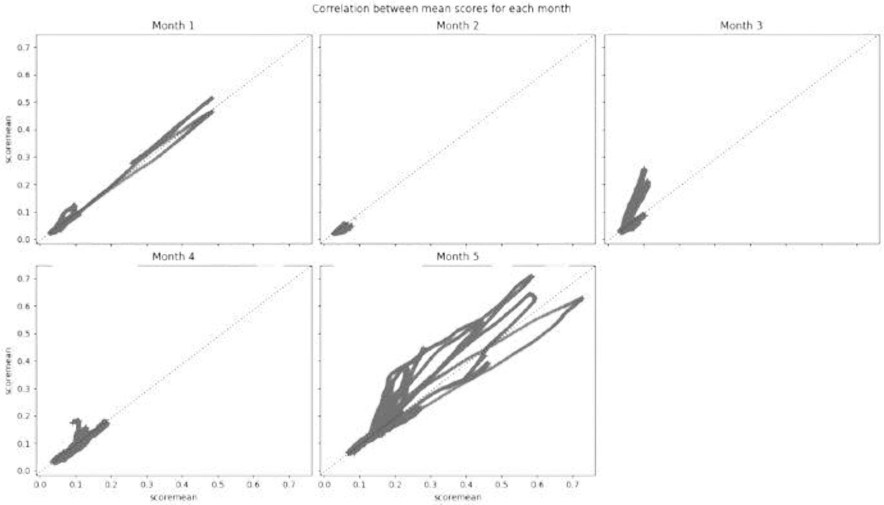


Figure 18: Correlations between original scores and avatar scores: the points are concentrated around the y=x axis, illustrating a strong correlation.

To further compare the scores, we focused on the **time steps** where the scores produced by the avatar model fell **outside the CI** associated with the original model. For these time steps, we measured the **maximum difference** and reported it as a **positive** difference if the avatar score was higher than the original score and as a **negative** difference in the opposite case. To take into account the magnitude of the score in the calculation of the differences, they were divided by the score. Thus, the differences could be **compared** between the time steps.

These values could be divided into three categories.

◆ **Differences close to 0** (e.g., difference in the interval [−0.5,0.5]). These differences can be related to several factors such as model uncertainty. Differences on these time steps have very little impact on anomaly detection in practice.

◆ **Largely positive differences** (e.g., difference >0.5). These differences tend to produce false positives, i.e., create invalid alerts or assign very large scores to existing alerts. In the context of this industrial data, the goal of anomaly detection is to report suspicious time steps to experts.

False positives therefore have the effect of increasing the mobilization of human expertise but do not represent a danger as false negatives could be. Regarding the anomaly scores that are much higher on real anomalies, this does not represent a particular risk. The management of these false positives can be done by adapting the threshold used to raise an alert.

◆ **Largely negative differences** (e.g., difference < 0.5). These differences tend to generate false negatives, i.e., not to create alerts when suspicious events occur. In an anomaly detection context, false negatives are more impactful than false positives.

Figure 19 shows the **differences obtained** over the entire period. Largely positive differences existed. Compared with the anomaly scores (Figure 20), it turned out that these were only observed on **already high scores** and therefore on potential anomalies. Note that no large negative differences were observed, illustrating that training the 1D-CNN on avatar data **does not increase** the risk of false negatives.

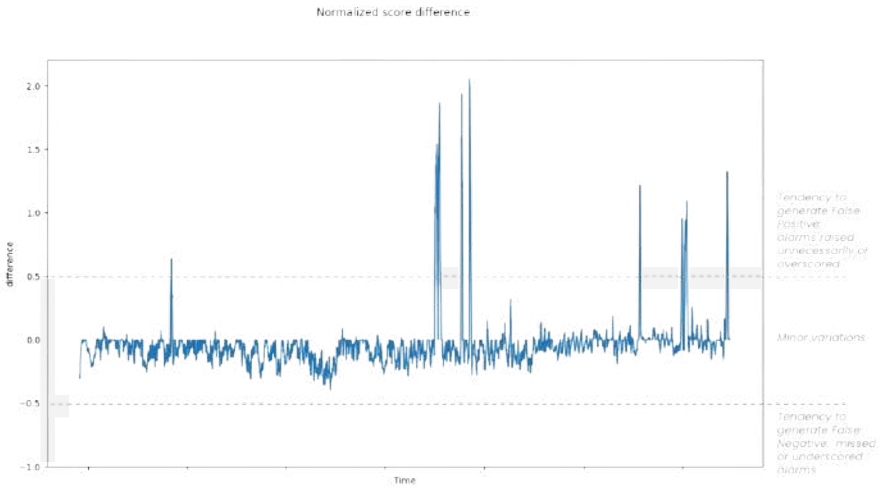


Figure 19: Classification of normalized differences obtained on the use case of a member of the Confiance.ai program.

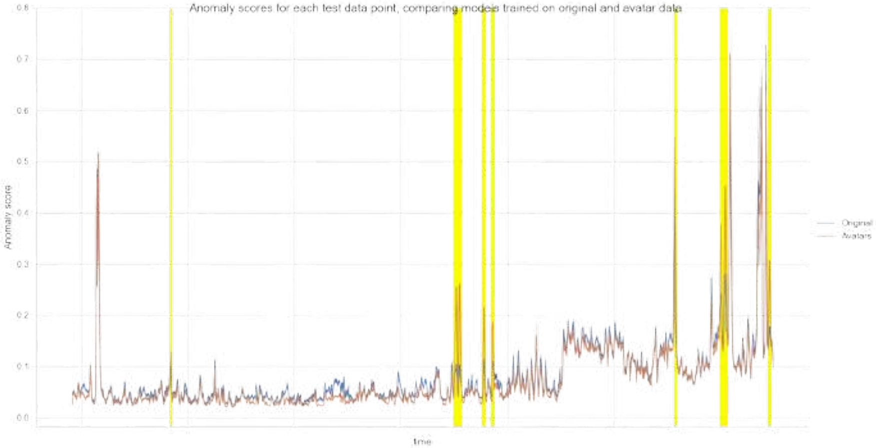
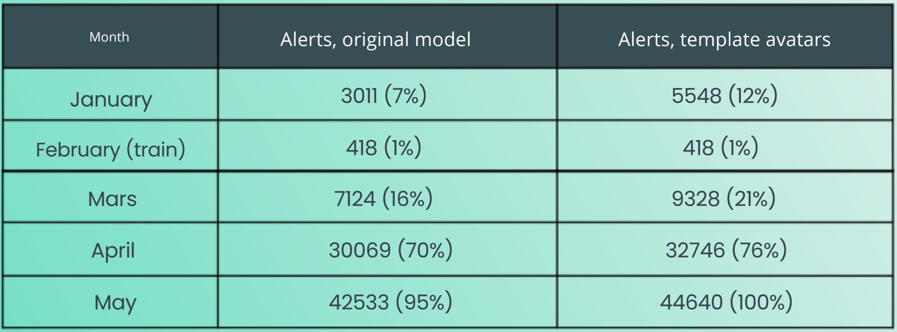


Figure 20: Anomaly scores obtained by the two models and normalized differences greater than 0.5: the largest differences were observed on time steps with already high original scores.

In practice, an anomaly detection model is often associated with a **threshold beyond which an alarm can be triggered**. For this use case, this alarm was intended to report **suspicious behavior** to experts for in-depth analysis. In previous work on 1D-CNN, the **recommended threshold** was determined by taking the 99th percentile of the anomaly scores obtained on the training data. Thus, the thresholds of **0.085** and **0.061** were obtained for the original and avatar models, respectively.

By applying these thresholds to the scores obtained over the test months, it was possible to calculate the **number** and **percentage of alerts** that would be raised by these models. We present these results below.



Month	Alerts, original model	Alerts, template avatars
January	3011 (7%)	5548 (12%)
February (train)	418 (1%)	418 (1%)
Mars	7124 (16%)	9328 (21%)
April	30069 (70%)	32746 (76%)
May	42533 (95%)	44640 (100%)

Table 5: Number of alerts created by each model.

Because the nominal operation of the system studied **evolves over time**, it was important to consider these results month by month. The difference in the number of alerts between the two models was relatively **constant**. The model trained on avatars generated 5% more alerts than its equivalent trained on the original data.

This phenomenon can mainly be explained by the fact that anonymization tends to **refocus individuals** and in particular the most extreme ones. Although this is beneficial from a **privacy** point of view and **reduction of the sensitivity** of the data, differences in the tails of distributions are likely.

A **threshold calculation** based on the same assumptions (taking for example the 99th decile) can produce differences such as those observed in the analysis of the industrial data. The strong growth in the number of alerts according to the month, whatever the model, is explained by the **profile of the data** with a nominal regime at the end of the period, which differs from that of February used for training as illustrated previously.

> *Data in the specific field of occupational health have become a tool for both research on occupational risks and disinsertion. More recently, indicators have been developed to monitor actions carried out by organizations responsible for monitoring employees and agents, such as inter-company or autonomous occupational health and prevention services, and services assigned to public functions. These data are also a means of social dialogue between stakeholders in this specific field. However, due to the extreme sensitivity of these data, only reports with aggregated data have been possible in practice until now.*
>
> *It is in this context that the Avatar-type solution presented by Octopize without possible re-identification opens a path of reflection and collaboration between all the actors, including the employees/agents themselves, on these questions of research, indicators and dialogue of course at a macro level (sector, prevention service, etc.). It is possible to imagine real sharing of data, without risk for the employee or agent intended for actors in the world of work and prevention.*

Pr Alexis Descatha

- Clinical Professor of Occupational Medicine, Epidemiology and Prevention, **Donald and Barbara Zucker School of Medicine**, **Hofstra/Northwell, USA**
- Inserm, Irset UMR1085 Ester Team, University of Angers, France
- Grand Ouest Poison Control and Toxicovigilance Center, Prevention, Angers University Hospital, France

F.3 Validation of results on a second analysis

To validate the conclusions, the analysis was **replicated** on **another time period** for the same sensor. In this second analysis, the month of August was used to train the 1D-CNN model and the data from June to November were used as a test. Note that over this period, the data from June reflected a **very different** operating mode from the other months, thus resulting in **high anomaly scores**. This phenomenon is shown in Figure 6. The results were consistent with those obtained on the first analysis. In particular:

- Highly **correlated** original and avatar scores (Figures 21 and 22)
- Score differences not generating **false negatives** (Figure 23)

Anomaly scores for each test data point, comparing models trained on original and avatar data

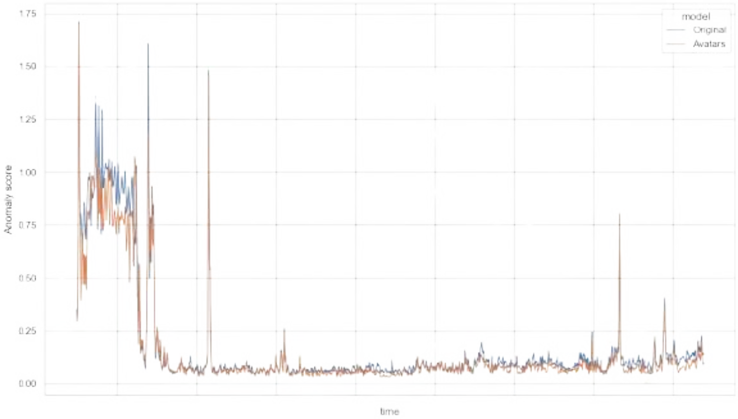


Figure 21: Anomaly scores for the two models obtained in the second analysis.

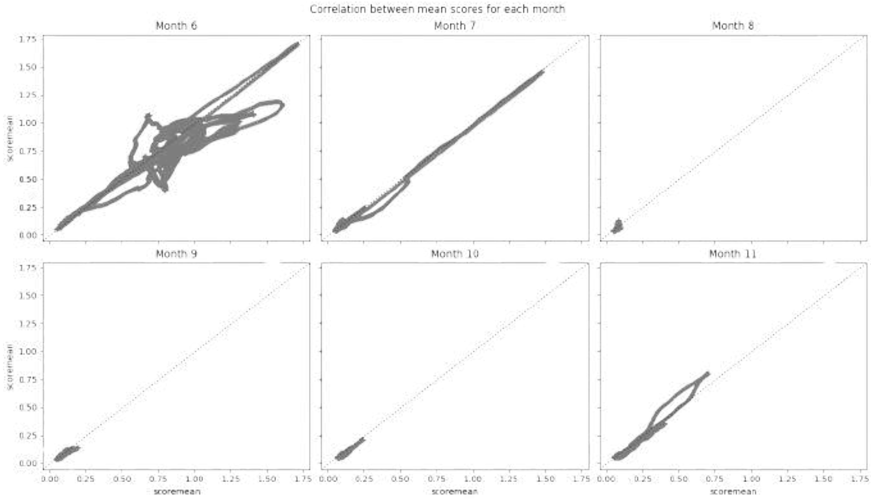


Figure 22: Correlation between original scores and avatar scores in the second analysis.

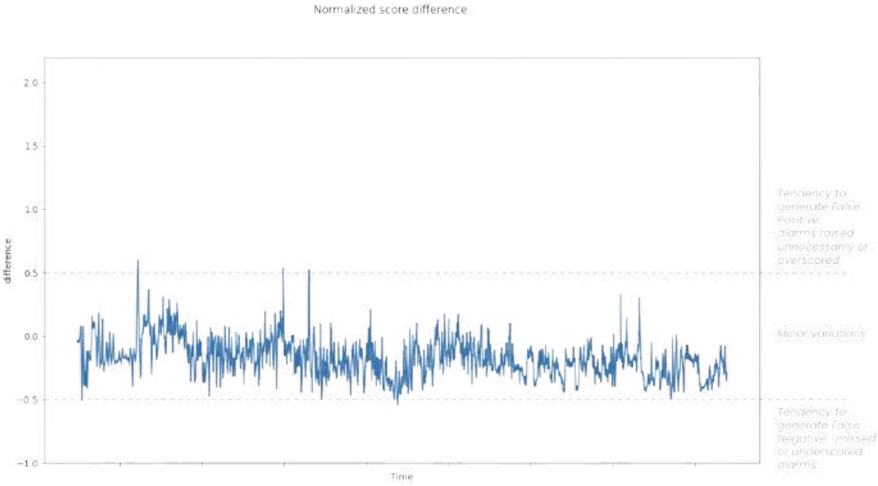


Figure 23: Classification of the normalized differences obtained in the second analysis.

Reducing the "time to market" of systems integrating Artificial Intelligence (AI) is a strategic challenge for organizations. While the rise of Generative AI has accelerated the development of proofs of concept, the transition to production remains a major obstacle, with many projects still frozen at the experimental stage. The data protection strategy is, among other things, an important barrier to overcome in an industrial context.

The Trusted AI community for industry has been structured around the Confiance.ai program to address the many challenges associated with engineering trusted AI for critical systems and brings together more than 50 partners: manufacturers from various sectors, leading research centers and innovative startups such as "Octopize".

Recognized internationally, this community is now opening up and organizing itself into a European-wide foundation, "The European Trustworthy AI foundation".

Paul Labrogère

Managing Director / CEO

SystemX Technological Research Institute

CONCLUSION

G.

This exploratory project, carried out with the support of the teams of **Confiance.ai**, one of its members, and Octopize, demonstrated the feasibility of **desensitizing industrial data** through their **anonymization**. It responded to a crucial challenge: enabling the secure sharing of sensitive data between partners. Our objective was to **share statistical information** on industrial data without disclosing their strategic value, such as production processes.

This study validated the **effectiveness of the avatar method** for this use case, by allowing the anonymization of data from industrial sensors. It also demonstrated the method's ability to process **complex data** such as time series, which are ubiquitous in industry. For data to be fully exploitable, it must strike a **balance between utility and privacy**. In this study, we found that the avatar data offered both of these guarantees: highly **correlated** anomaly scores and proof of **privacy** provided by privacy metrics.

Furthermore, we illustrated the interest of anonymous synthetic data for training **Machine Learning models** in an unsupervised context. Our results showed that models trained on avatar data displayed **equivalent performances** as those trained with original data.

Avatarization could even prove beneficial for **unsupervised learning**, by removing **unrepresentative features** from datasets. This helps model normality before training anomaly detection models.

By combining **privacy** and **performance**, avatarization opens up new perspectives for **innovation** in the industry.

BIBLIOGRAPHY

H

1 Guillaudeux, M., Rousseau, O., Petot, J. *et al. Patient-centric synthetic data generation, no reason to risk re-identification in the analysis of biomedical pseudonymized data.* Nature Digital Medicine. 2023

2 Barreteau A-F, Regnier-Coudert, Le Carpentier E, Moussaoui S. *Génération de signaux anonymes à partir de données non anonymes par modèle de mélange linéaire local*. 2023. Colloque Francophone de Traitement du Signal et des Images GRETSI

3 Wang J-L., Chiou J-M., Muller H-G. *Functional data analysis*. Annual Review of Statistics and its application. (3)-1 2016

4 Bailly, R., Malfante, M., Allier, C., Ghenim, L., and Mars, J. (2021). *Self-supervised learning for anomaly detection on time series: application to cellular data*. Conférence sur l'Apprentissage automatique (CAp2021)

5 Bailly, R., Malfante, M., Allier, C., Ghenim, L., and Mars, J. (2021). *Deep anomaly detection using self-supervised learning: application to time series of cellular data*. In ASPAI 2021 - 3rd International Conference on Advances in Signal Processing and Artificial Intelligence, Porto, Portugal.

6 Li, D., Zhang, J., Zhang, Q., and Wei, X. (2017). *Classification of ecg signals based on 1d convolution neural network*. 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), pages 1–6.

ABSTRACT

Anonymization of strategic data with avatar

This white paper, written with the support of the Confiance.ai program, is a project aimed at developing trusted technologies for artificial intelligence, particularly in the field of sensitive data protection.

In an increasingly connected industrial environment, data plays a vital role in optimizing processes. However, sharing and using it exposes us to increased privacy risks. This white paper explores anonymization as a solution to protect critical data while maintaining its usefulness.

The avatar method developed by Octopize allows the generation of synthetic data preserving both the statistical structure of the original data and their confidentiality. It is based on multidimensional projection techniques to create data avatars and meets the strict anonymization criteria defined by the CNIL and the European Data Protection Committee, guaranteeing the non-reidentification of the individuals or entities concerned.

The work presented here focuses on a use case provided by a consortium member, where time series data collected by sensors were anonymized to be used for training anomaly detection models based on NNs.

The results show that anonymization protects data without compromising its usefulness for analytical applications. Anomaly detection models trained on avatar-anonymized data show similar performance to those trained on original data. This white paper thus demonstrates the feasibility of protecting sensitive data in an industrial setting, while enabling technological innovation through secure data analysis.

Authors

Olivier Regnier-Coudert

PhD, Data Scientist
Octopize

Amélie Bosca

Data Scientist
Sopra Steria & IRT System X

Mathieu Bleunven

Business Manager, Defence & Mobility Expert
Octopize

Marie Berthon

Business Manager, Defence Expert
Octopize

Gabrielle Crolard

Communication & Marketing Manager
Octopize

About

This white paper, produced as part of the Confiance.ai program, presents the work of Octopize and a consortium member on the anonymization of industrial data. It explores the “avatar” method, an innovative technique for generating synthetic data that guarantees both confidentiality and utility for applications such as anomaly detection.

The document highlights the challenges of protecting sensitive data in the industry while respecting the requirements of European regulations.

To learn more, contact us at **contact@octopize.io** or visit www.octopize.io

OCTOPIZE
MIMETHIK DATA