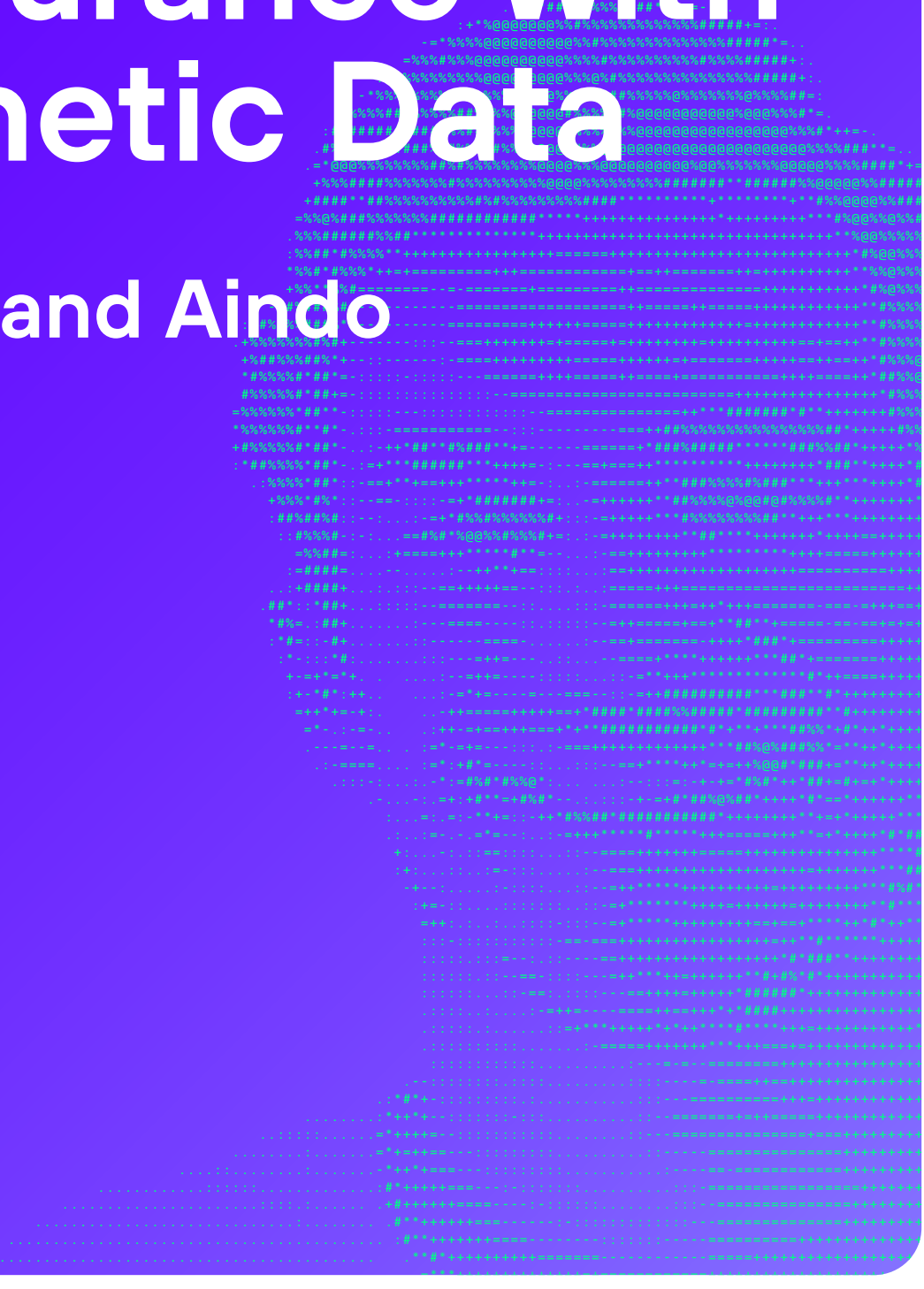
aindo

BE Rebel

Case Study:
Enhancing AI in Insurance with Synthetic Data

BeRebel and Aindo



Index

Introduction

The insurance sector can greatly benefit from AI models that optimize processes and enhance customer experience. The models' effectiveness would increase even further if trained on pooled data from multiple insurance providers, resulting in more efficient and accurate AI solutions.

However, the sensitive nature of client information in the insurance sector makes data sharing between organizations challenging, with stringent privacy regulations often creating additional barriers.

Synthetic data offers a transformative solution by enabling secure data pooling without compromising data security or compliance. This technology replicates the statistical properties of real data while removing identifiable information, allowing organizations to share insights and train AI models safely.

Synthetic data enables organizations to get insights and develop AI models safely.

To validate the feasibility of synthetic data for industrial use, car insurer BeRebel partnered with synthetic data provider Aindo in a regulatory sandbox initiative organized by the Italian Institute for the Supervision of Insurance (IVASS). This experimental initiative is divided into two phases:

- **Phase 1**, completed in December 2024, focused on assessing the performance of AI models trained on synthetic data compared to those trained on real data.
- **Phase 2**, scheduled for completion by mid-2025, will evaluate the efficiency of synthetic data for pooling insights from multiple providers to create a larger, privacy-preserving database for advanced analytics and AI model training.

This leaflet highlights the key outcomes of Phase 1 of the project.

Project overview and objectives

Regulatory sandbox

BeRebel and Aindo collaborated within a regulatory sandbox initiative organized by the Italian Institute for the Supervision of Insurance (IVASS). This sandbox provided a controlled environment to test AI innovations while ensuring full compliance with GDPR and sector-specific regulations. For more details about the sandbox, visit: IVASS Sandbox Overview.

BeRebel replicated their AI models by training them on synthetic datasets.

BeRebel's AI models

BeRebel, an innovative Italian car insurance provider, specializes in offering personalized monthly insurance plans. To enhance its products and services, BeRebel relies on AI-driven models. Prior to the project, BeRebel had developed two key AI models:

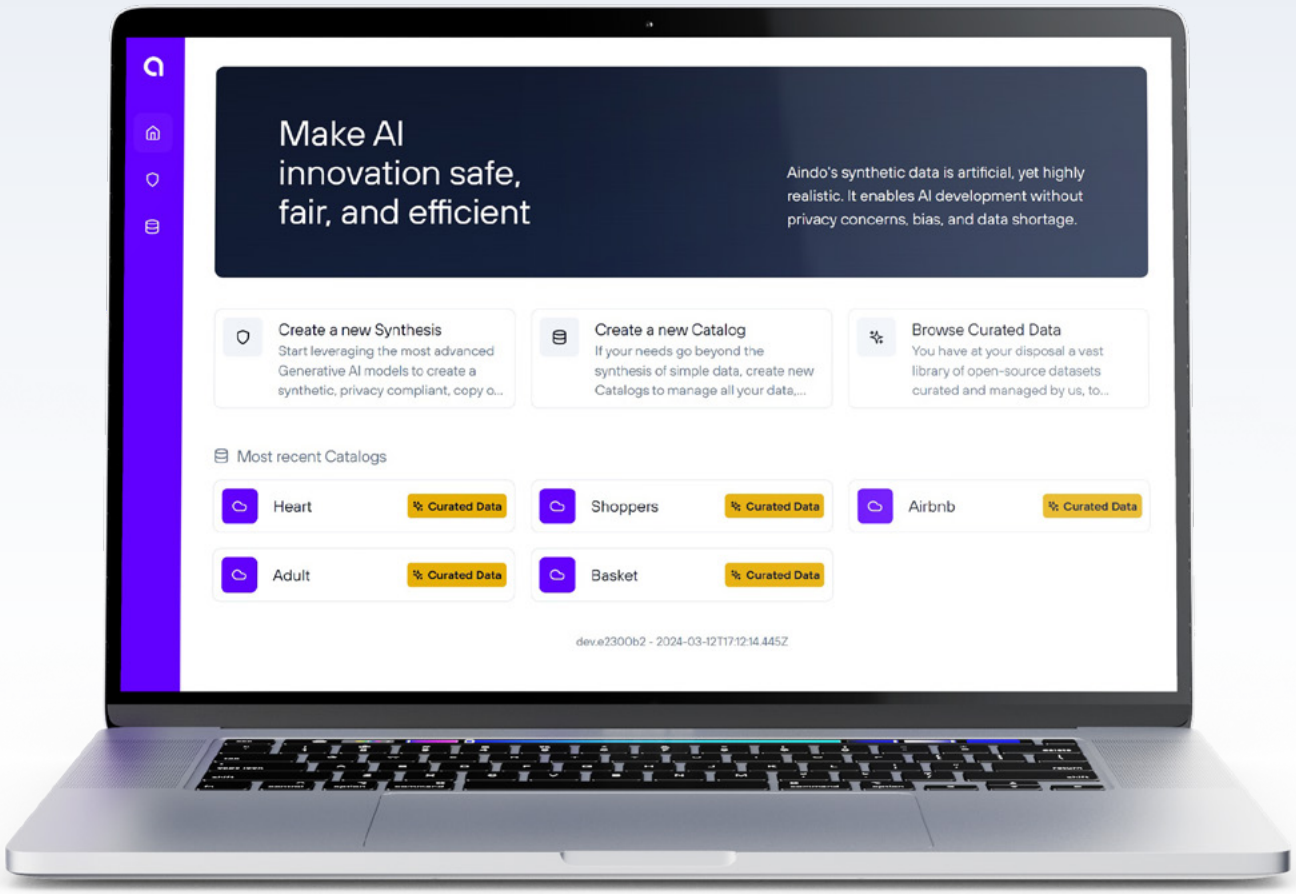
- **Conversion Model** – Predicts whether prospective customers would convert to paying customers.
- **Frequency Model** – Estimates the number of claims over a specific period.

These models were trained on two internal datasets: a customer quotations database and a policies database.

In order to assess the feasibility and effectiveness of synthetic data technology, BeRebel wanted to replicate these AI models by training them on synthetic datasets generated using Aindo's platform. The performance of the synthetic-data-trained models was then to be compared with that of the real-data-trained models.

Aindo's technology

To meet BeRebel's needs, Aindo deployed its state-of-the-art synthetic data platform. The platform generates synthetic datasets that mimic the statistical properties and utility of real data while safeguarding privacy and ensuring regulatory compliance. For this project, the platform was securely deployed on BeRebel's infrastructure, ensuring that real data remained onsite and protected at all times.



Process and setup

Aindo's team collaborated with BeRebel to define the operational requirements for synthetic data generation. This included defining the structure, data types, and volume of the datasets, as well as identifying the computational resources needed for the process.

Following the setup and training phase, BeRebel didn't require further assistance from Aindo.

Then Aindo trained BeRebel staff to use Aindo's platform. Following the setup and training phase, BeRebel independently managed the synthetic data generation process without requiring further assistance from Aindo.

Datasets used for synthetic data generation

BeRebel's datasets, central to their AI models, included two key databases:

- **Quotation Database**
  - Contains approximately 224,000 observations
  - Includes 12 variables (5 categorical, 7 numerical)
  - Used to analyze conversion rates, predicting whether potential clients who request quotes will become paying customers.
- **Policy Database**
  - Contains approximately 167,000 observations
  - Includes 10 variables (5 categorical, 5 numerical)
  - Used to develop a claim frequency model, estimating the number of claims over a given period.

Both datasets provide key insights into customer behavior and risk assessment, making them essential for building accurate and efficient AI models in the insurance sector.

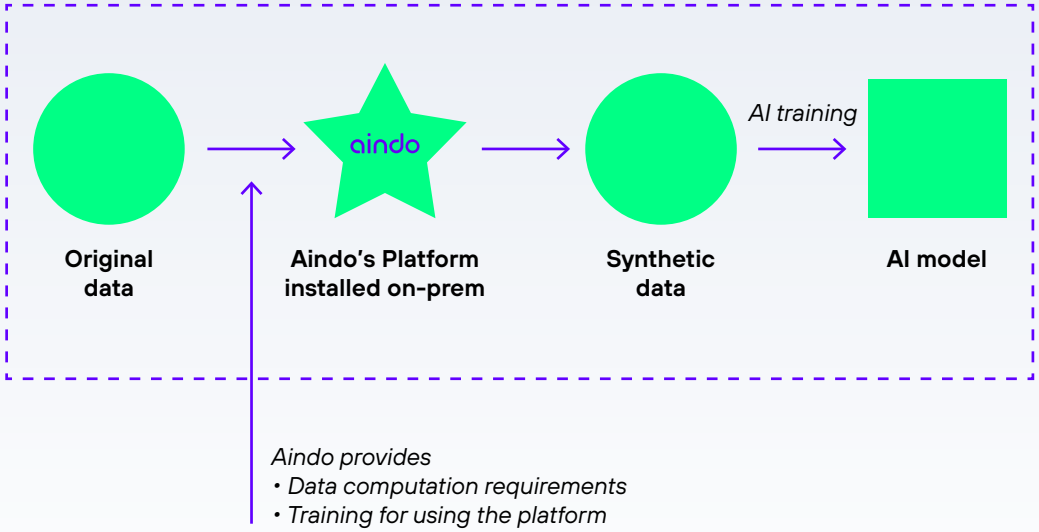
Synthetic data generation and deployment

BeRebel used Aindo's platform to independently process their datasets and generate synthetic data entirely within their secure environment, ensuring that no sensitive information left their premises. This autonomous approach safeguarded BeRebel's real data while enabling them to fully leverage the benefits of synthetic data.

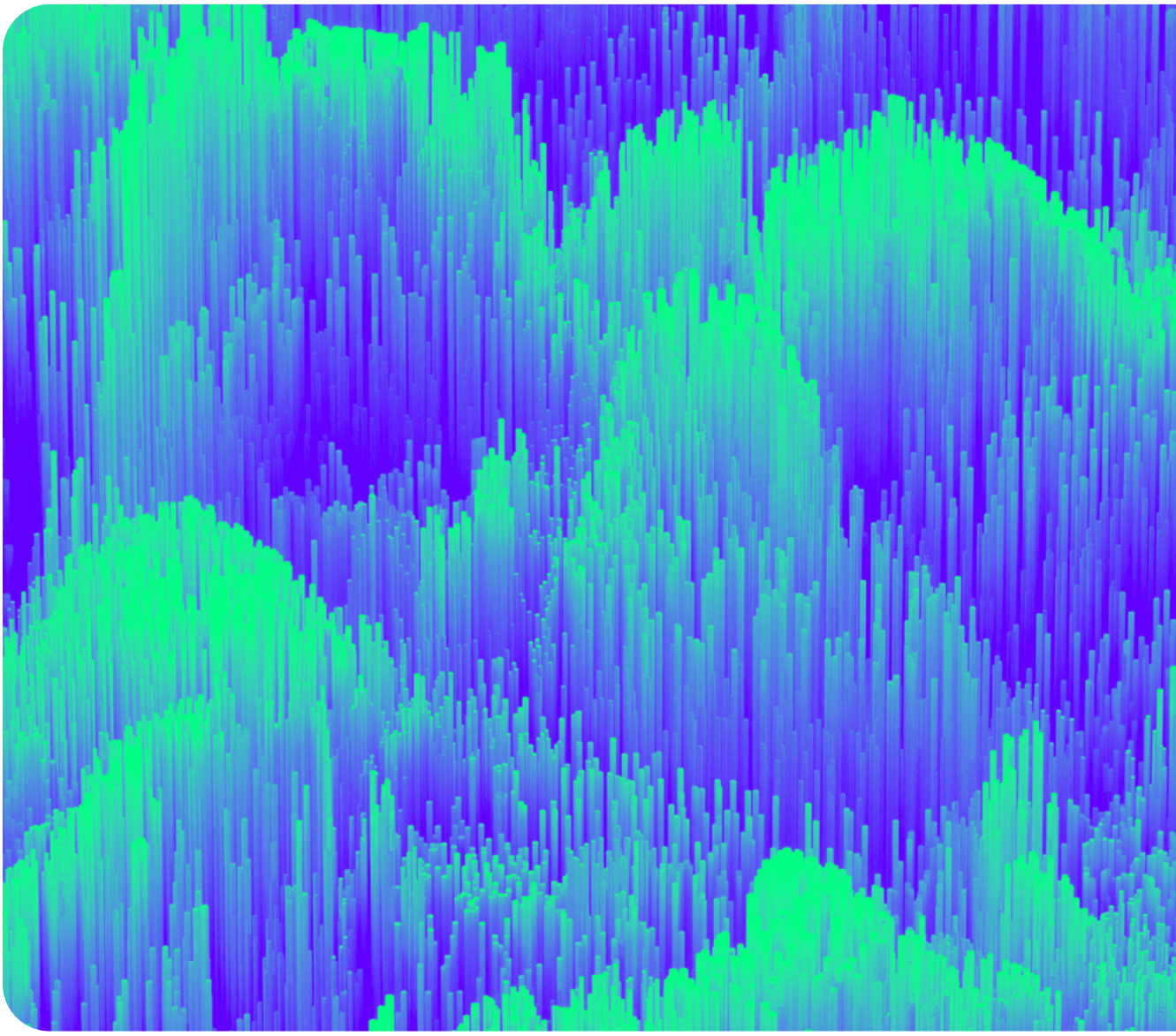
The deployment process was seamless. After a brief training session provided by Aindo and an initial assessment of computational requirements, BeRebel staff were able to independently generate synthetic datasets. Thanks to the platform's intuitive interface, they could easily replicate the statistical properties and analytical value of their real data, creating synthetic datasets that were fit-for-purpose and ready for AI model development.

Seamless deployment process and an intuitive platform interface.

The data workflow, illustrated in **[Figure 1]**, shows how BeRebel managed the synthetic data generation process by installing Aindo's platform on-prem and without sharing any of the original data externally.



[Figure 1]
Overview of the data workflow during the project.



AI model development

Using the newly generated synthetic datasets, BeRebel successfully developed two AI models tailored to their objectives:

- **Customer Conversion Model** – Predicting the likelihood of prospective customers becoming paying clients.
- **Claims Frequency Model** – Estimating the expected number of claims over a specific period.

These synthetic-data-trained models were evaluated alongside their real-data-trained counterparts to ensure comparable performance, accuracy, and utility. The results confirmed the reliability and effectiveness of synthetic data for advanced AI model development in the insurance sector.

Outcome

Model performance

- Conversion Model: Accurately predicted customer behavior, achieving performance metrics nearly identical to those of models trained on real data.
- Claims Frequency Model: Estimated claim counts with an average forecast error of just 0.8%.

The evaluation demonstrated that **synthetic data-trained models performed with less than a 1% difference compared to real data-trained models**. This confirms synthetic data's ability to maintain high utility while fully protecting sensitive information.

Conversion model [A]

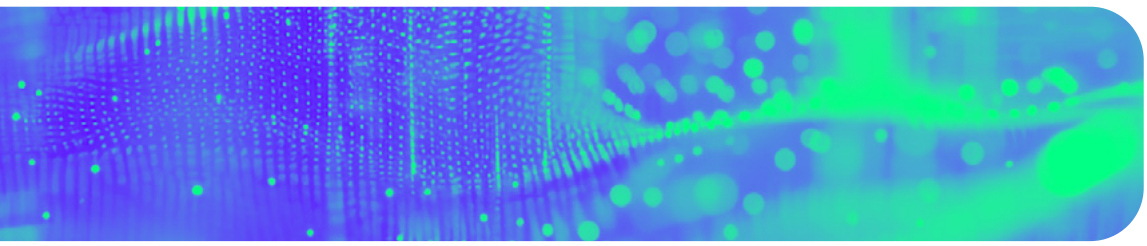
	Data model royals	Data model synthetics
Precision	0.94	0.93
Recall	0.94	0.93
F1 score	0.94	0.93
AUC-ROC	0.97	0.97

Frequency model [B]

	Observed	Expected
Frequency	12.66%	12.76%
# of accidents	1803	1817

[A]
Performance evaluated using precision, recall, F1-score, and AUC-ROC.

[B]
Predictive accuracy assessment: observed and predicted claims using the mean prediction error.



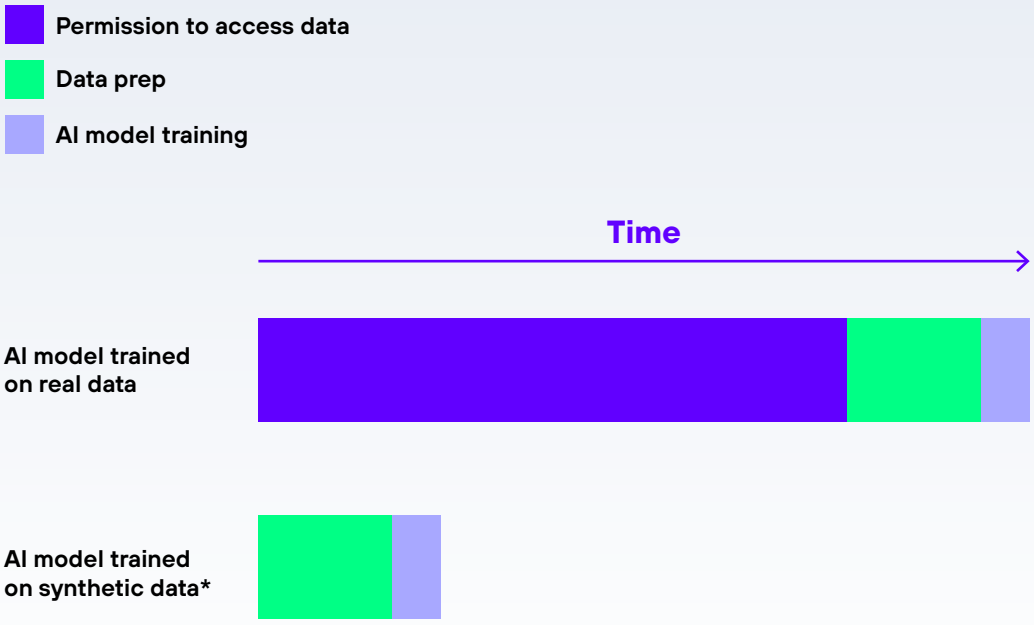
Improved data availability

Synthetic data is a powerful privacy-enhancing technology that unlocks the potential of personal data while ensuring compliance with stringent privacy regulations.

At BeRebel, **obtaining data for AI model development has traditionally been a time-consuming and resource-intensive process**. The typical workflow requires navigating complex privacy compliance protocols, securing extensive permissions, and involving multiple departments, including purchasing, legal, data protection, engineering, and ICT. On average, preparing data for AI models takes 4 to 6 months per project, with permissions expiring after limited periods, necessitating a repeat of the process for each new project.

In contrast, **synthetic data drastically reduces these hurdles**. Once Aindo's platform is installed - a process that requires an initial data processing agreement (DPA) and internal approval from BeRebel's DPO - the time-to-data for an AI model was reduced by more than 75%. Synthetic data eliminates the need for repeated compliance reviews, significantly reducing administrative overhead and operational costs.

After installing Aindo's synthetic platform, BeRebel can now independently and effortlessly obtain fit-for-purpose training datasets within days instead of months, streamlining their AI development pipeline and freeing up resources across teams.



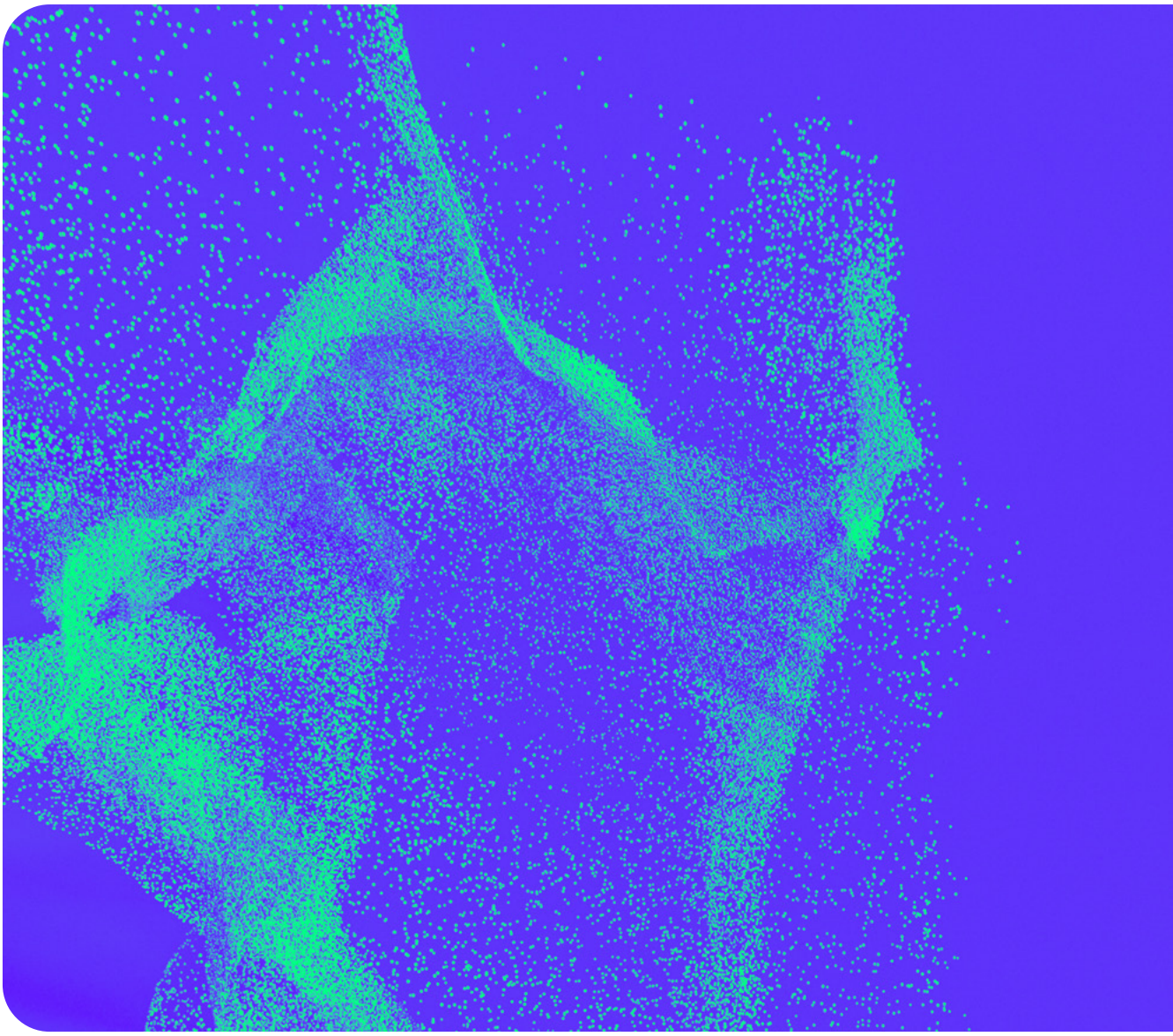
[Figure 2]
Comparison of time-to-data for models trained on real data compared to models trained on Aindo's synthetic data. (*estimated time after Aindo's platform is installed)

Enhanced privacy protection

Aindo's synthetic data fully meets GDPR and sectoral compliance requirements, as confirmed by IVASS.

With synthetic data, BeRebel can confidently develop AI models without requiring additional reviews or approvals from their DPO. Since Aindo's synthetic data is non-personal, it can be reused for any purpose, at any time, without compromising privacy or compliance. Additionally, Aindo's platform integrates built-in data protection measurement tools to ensure highest privacy standards. For more technical details about these features, visit Aindo's documentation page https://docs.aindo.com/.

Aindo's synthetic data fully meets GDPR and sectoral compliance requirements.



Summary

Aindo's synthetic data technology empowered BeRebel to develop robust and accurate AI models while addressing the complexities posed by stringent privacy regulations. The project demonstrated synthetic data is a transformative alternative to real data, offering a powerful combination of utility, privacy, and efficiency.

Key outcomes:

1

Identical performance

Models trained on synthetic data achieved metrics nearly identical to those trained on real data, with less than a 1% performance loss.

2

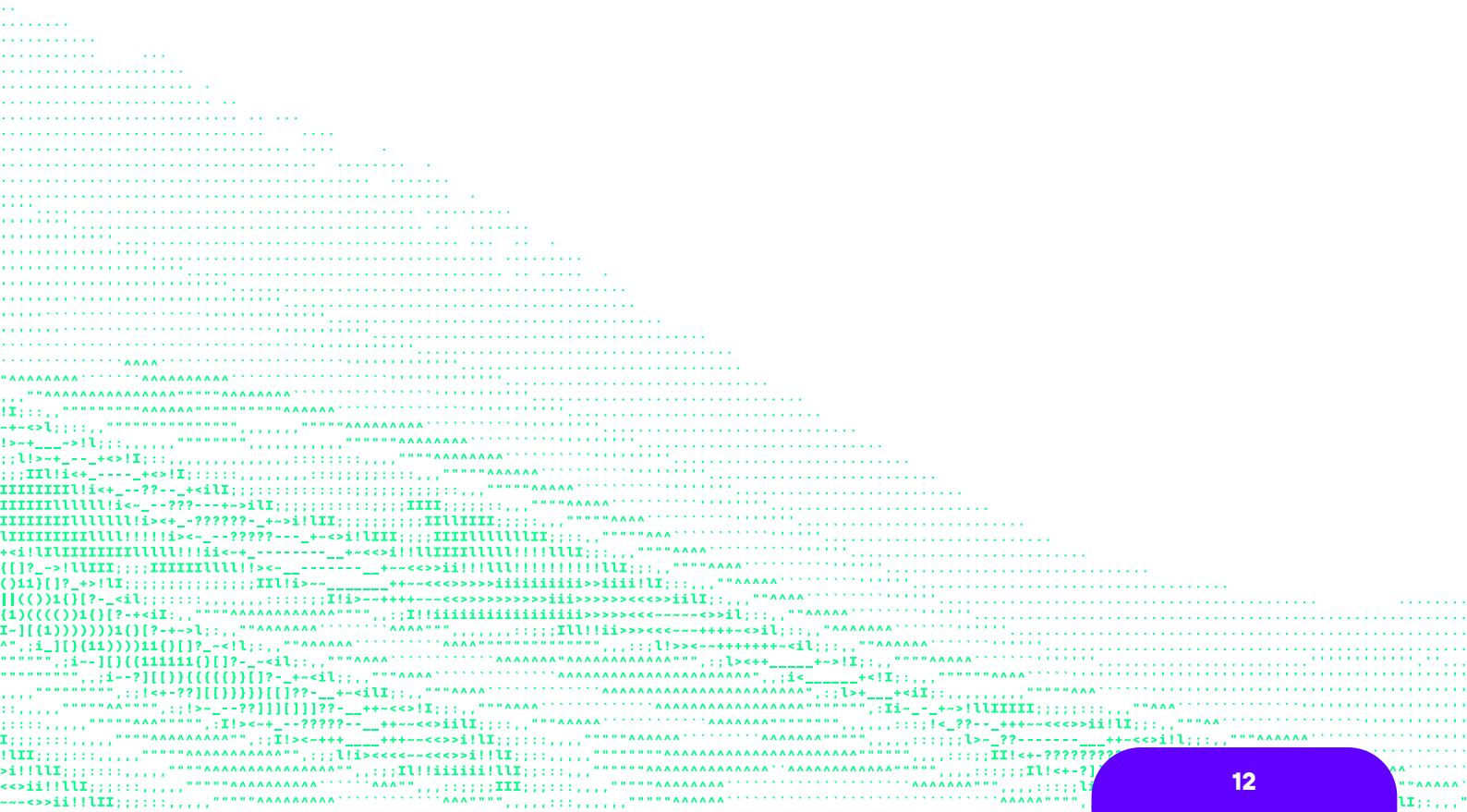
Enhanced privacy protection

Synthetic data fully met GDPR and sectoral compliance requirements, as verified by IVASS.

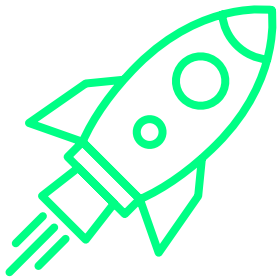
3

Time-to-data reduction

Synthetic data reduces the typical time-to-data by more than 75%, involving fewer stakeholders and significantly accelerating the development process.



Next steps



By reducing time-to-data, ensuring compliance, and maintaining model accuracy, synthetic data is unlocking new possibilities for secure data sharing and AI-driven innovation in the insurance sector.

The next phase of the project will begin in January 2025 and it will focus on assessing the effectiveness of synthetic data for data pooling across multiple insurance providers. This phase will aim to create a larger, integrated dataset to enhance the development of data-driven solutions, optimize analytics, and further validate synthetic data's utility for collaborative innovation.

Partner with Aindo

Unlock the full value of your data with Aindo's cutting-edge synthetic data technology.

Partner with us today and lead the way in secure, innovative data solutions.

aindo

Be Rebel

Contacts:

hello@aindo.com
+39 040 375 5640
Località Padriciano 99
34149, Trieste (TS), Italy