
experian
NAVIGATING NEW FRONTIERS
A strategic exploration of key
GenerativeAI
use cases within credit risk

Contents

> GenAI is likely to be one of the most powerful catalysts for innovation that we have ever seen. As we navigate this new era of possibilities, we have the opportunity to redefine the way we manage and analyse data to develop meaningful insights.

Jorge Bordona
CIO, Experian EMEA & APAC

The recent advances in Generative AI (GenAI), and in particular Large Language Models (LLMs), have swept across our global community like a tsunami. Never before have we held the keys to a tool that can unlock our shared processing code – language. But with all the hype around GenAI, it can be challenging to identify practical use cases to invest in and pursue.

To understand GenAI capabilities from a high-level perspective, imagine language as a jigsaw puzzle with a vast number of pieces. In this metaphor, the 'pieces' represent words and phrases, and the 'patterns' represent the meaningful arrangements of these pieces. GenAI can analyse the relationship between different pieces and identify correlations to reveal a complex picture. Unlike traditional AI, it not only identifies patterns but can produce new content on demand.

As we focus on the world of credit risk, GenAI offers opportunities to improve productivity by automating the process of credit risk assessment, thus reducing the time and resources required. In addition, it can simplify analytics, help meet compliance and regulation requirements, and enhance customer experience. With all these options, selecting the right use cases to invest in is critical for positive ROI.

This guide is intended to help shine a light on some of the most promising use cases offered by GenAI in credit risk analytics and fraud prevention. As a global technology leader, Experian has always been at the cutting edge of innovation, and our large team of data scientists is constantly finding new ways to optimise our clients' business processes whilst ensuring regulatory compliance.

We can help you chart the best course through this new frontier and maximise the potential of this exciting technology.

GenAI vs traditional AI

When considering which GenAI use cases to invest in, it helps to clearly differentiate between the capabilities and limitations of GenAI compared with traditional AI. Although there are similarities between these distinct classes of AI, there are some fundamental differences.

According to Experian's research, AI and its subset Machine Learning (ML) are currently being used by roughly a third of businesses across a wide variety of applications. These range from credit risk assessment and portfolio monitoring to fraud detection and identity authentication. More than half (54%) of the firms in our survey that are using AI/ML find the productivity gains have already offset the initial cost.

An essential component of these traditional AI/ML models is that they are trained on known datasets to identify correlations between features that may not be apparent to a human. This advanced pattern recognition allows them to make highly accurate predictions based on previous data. For example, to assess the credit risk of a new customer or create a fraud probability score.

AI/ML use within Financial Services and Telcos

	Fraud models	Credit risk models	Portfolio monitoring	Product enhancement	Identity authentication	Customer service
We are already using AI/ML	35%	34%	34%	33%	30%	36%
We aren't using AI/ML yet, but plan to use AI/ML in the next 12 months	39%	40%	39%	40%	39%	37%
We have no plans to use AI/ML	24%	25%	25%	26%	29%	27%

Base: 889 EMEA & APAC decision makers at Financial Services and Telco providers
Source: Experian research conducted by Forrester Consulting, July 2023

An important aspect of these AI models is their 'explainability', which refers to the ability to understand and interpret the decisions made by AI. With traditional AI models, it's possible to identify which features have contributed to the output, even when using complex neural networks. This transparency in decision-making is crucial as it allows users to trust and verify decisions based on AI.

The importance of explainability will be further highlighted by the upcoming EU AI Act. This legislation will make transparency in AI decision-making a legal requirement for credit assessments, setting a likely blueprint for the rest of the world to follow. The AI Act will also impose other transparency requirements for GenAI – irrespective of the use case.

This tenet of explainable AI is at the forefront of Experian's responsible and ethical use of this technology. All of the models that we have developed over the past 25 years have followed this principle. And our explainability solutions can be retroactively applied to existing models to ensure our clients' in-house models will comply with future legal frameworks.

How is GenAI different?

In contrast to traditional AI, the foundation LLMs that make up the backbone of GenAI are trained on unknown datasets consisting of vast quantities of text from the internet. Although these models can then be fine-tuned on known datasets for specific use cases, they remain unexplainable – meaning the features and weighting of the features that contribute to the output are unknown.

Until this explainability limitation is resolved, it means that GenAI models are not suitable for credit risk assessment. However, the core capabilities of GenAI – natural language processing – outperform traditional AI, and this is where they offer the greatest opportunity.

Using GenAI in conjunction with traditional AI models can significantly enhance the predictive accuracy of credit risk and fraud models. This is achieved by using GenAI to fast-track the analysis and preparation of the data used to train traditional AI models. Moreover, the incredible mastery that GenAI has over language means that it can provide model diagnostics and suggested improvements in a simple and easily understandable format.

In its current form, GenAI is not a replacement for traditional AI/ML credit risk and fraud models but rather an exciting tool that can elevate the process of developing and monitoring these models.

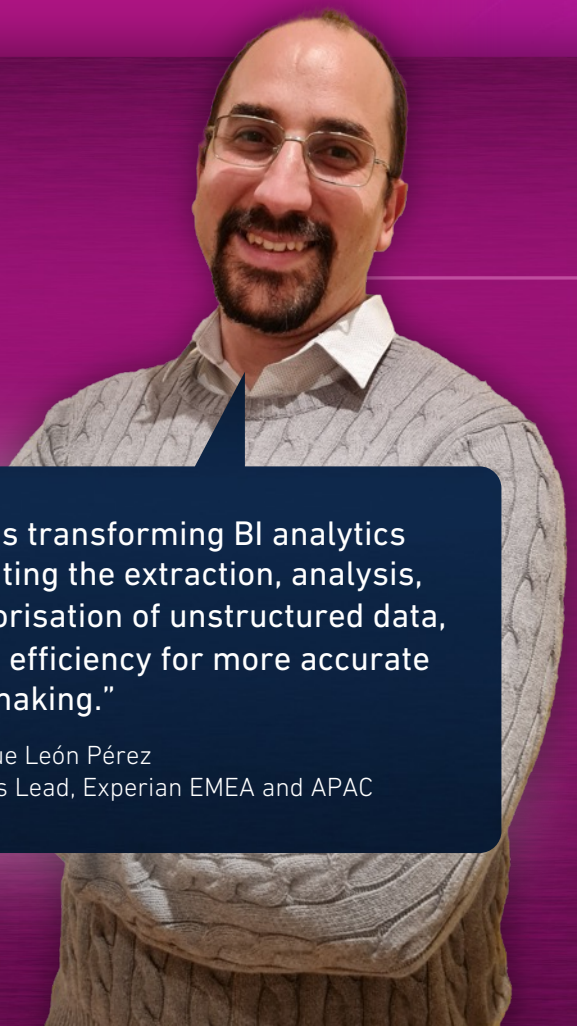
Where can GenAI provide the most value?

In two words – automation and productivity. Although the adoption of traditional AI and ML within credit risk is rapidly increasing, there are still many manual processes. By automating these tasks with GenAI, businesses can reduce the time and resources needed to complete them, bringing efficiency and productivity to a peak.

GenAI represents an AI inflection point, where the computer processing power and algorithms available can do more than analyse vast datasets and provide insight. We have now moved beyond that capacity, as GenAI can use this insight to suggest actions. This significant developmental step – from insight to action – represents a step change in AI capability.

Generic LLMs can be fine-tuned for a huge variety of specialised tasks. Experian is actively exploring a range of new use cases and is already working with clients to provide specific solutions. The use cases highlighted in this report focus on credit risk and fraud decisioning and illustrate three practical applications of this technology that businesses can implement today.

Experian's top GenAI applications



"Our LLM is transforming BI analytics by automating the extraction, analysis, and categorisation of unstructured data, enhancing efficiency for more accurate decision-making."

Tomás Enrique León Pérez
Data Analytics Lead, Experian EMEA and APAC

To better understand where where GenAI can provide the most value, we asked two of Experian's top data scientists for insight into which use cases they are most excited about and what the implications of these innovative solutions are for businesses.

1 Business Information (BI) data extraction

Application

The ability to automate the evaluation of unstructured data from business reports and documents for integration into risk models.

Business value

- Reduced costs
- Faster execution
- More comprehensive analysis
- Improved model predictive accuracy

One of the most relevant GenAI use cases that I see is in the preparation of Business Information (BI) data. These datasets are usually made up of large volumes of unstructured data, such as PDF reports, profit and loss (P&L) statements and legal documents. Up until now, the extraction of this data has required the manual review and labelling of data before it could be included in an ML risk assessment model. With the right LLM, this process can be automated, and the accuracy of the final model can be improved.

We are in the final stages of completing a proof-of-concept (POC) for this specific GenAI use case, and we have seen some remarkable results. A correctly trained LLM model is capable of reading and understanding these large PDFs, which are often over 100 pages in length and require a significant investment of time and effort to read and analyse.

The LLM can, almost instantly, review these documents and then summarise the relevant data points required for the task. That means it is finding contextual information rather than simply copying all the information from one place to another. What's more, they can then classify each data point and add them to the database that we use to develop the risk score model. This capability brings a lot more information to our databases, enabling us to provide more accurate responses to our clients to improve their decision-making.

If you think about how data is currently extracted from these P&L or PDF documents, it often involves a team of highly experienced financial experts. This process requires a huge amount of time and resources to evaluate and extract the key data points. The final step is to add their own insight.

In contrast, with an LLM model, all of the tedious and time-consuming copy-and-paste work can now be automated. Not only can the LLM extract the relevant data, but it can also categorise it and automatically include it in the database. The end result is that it frees up time for the experts to provide more value in the process with better insight that can ultimately improve predictive accuracy.

When we consider many of the GenAI use cases, a lot of the focus has been on the creation of new text and images. However, for me, this capacity of LLMs to analyse and summarise documents is more exciting in terms of the benefits that it can deliver to BI data analytics. It allows lenders to be far more efficient across the whole process by reducing the time involved and increasing the number of documents analysed per annum.

I would also like to add that although this capability may seem to be replacing human workers, we don't see it that way. What it means is that you will have to rearrange your business ecosystem, but humans are still a vital part of the process. Especially when it comes to validating the output from the GenAI models. This is essentially a new position that is created when applying this technology. Moreover, it will allow humans to focus on higher-value tasks rather than more time-consuming, data-intensive and repetitive tasks.

In our current global economic context of uncertainty and elevated risk, the ability to improve the speed and accuracy of lending decisions has never been more critical. With this technology, we can make material gains in the accuracy of those decisions. How? By improving the quality of the source data that we use in the risk models. The end result is enhanced predictive accuracy, with a reduction in costs through efficiency and automation.

2 Model monitoring

Application

Automate model diagnostics with a GenAI assistant.

Business value

- Reduce the time required to monitor model performance and diagnose model drift.

The second application of GenAI that I'm excited about is model monitoring. We are fine-tuning an LLM that can be used to monitor credit risk model performance over time and identify where performance has degraded. In addition to recognising any issues, the LLM can then diagnose what the problem is. The final step is suggesting solutions to rectify the problem and improve the model performance.

With this capability, the time required for model diagnostics can be greatly reduced. Rather than depending on an analyst's predictions for model improvement, the LLM can constantly monitor the model and instantly suggest remedial action in a simple and understandable way if the performance starts to decline.

This solution can be added to existing analytics platforms to streamline the process of model monitoring. It forms a part of our Model Monitoring Toolbox (MMT), which drastically simplifies the process involved with consistently maintaining the predictive accuracy of risk models.

Another advantage of this new solution is that it will provide insightful metrics in a readable format. This empowers a broader audience – beyond the technical data analysts – with the ability to understand the current situation for any scorecard.

Our model-monitoring LLM is being trained on industry-specific data, and it can fill a previous gap in model health analysis by responding to a variety of questions and providing relevant advice. This use case is in the MVP stage and is already delivering great results.

GenAI and model monitoring

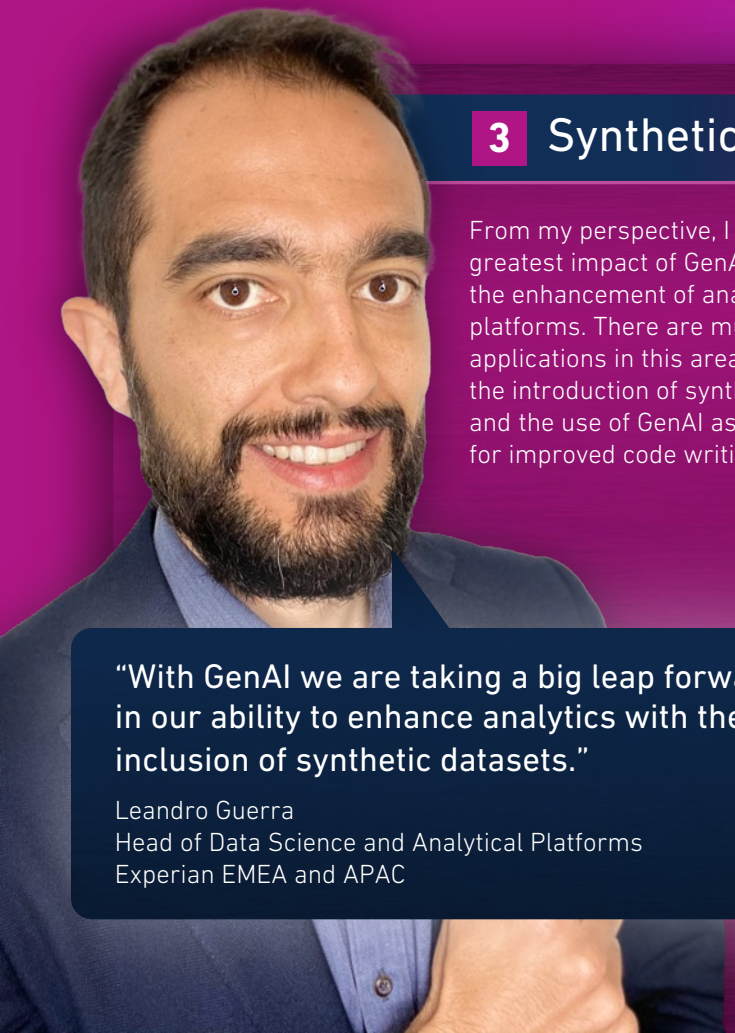
What is it?

LLMs can be trained to provide a human-like response to a variety of questions and provide advice based on a given context.

How does it work?

The LLM is trained on industry-specific data, so it can analyse the health status of a model and suggest the next steps for optimal portfolio management.

Enhancing credit risk analytics



"With GenAI we are taking a big leap forward in our ability to enhance analytics with the inclusion of synthetic datasets."

Leandro Guerra
Head of Data Science and Analytical Platforms
Experian EMEA and APAC

3 Synthetic data

From my perspective, I see the greatest impact of GenAI being the enhancement of analytics platforms. There are multiple applications in this area, such as the introduction of synthetic data and the use of GenAI assistants for improved code writing.

Application

- Compliance and regulation requirements
- Simplify data sharing
- Overcome data scarcity
- Code more efficiently

Business value

- Greater predictive accuracy from credit risk and fraud models
- Reduced complexity involved with data sharing across borders and regions

The first GenAI application I would like to highlight is the generation of synthetic data. Data drives profitability. And while traditional AI is taking decision-making and fraud prevention to new heights, this progress is heavily dependent on vast quantities of data. Not only do these models require adequate data, but the datasets must be sufficiently balanced to deliver the most stable and accurate predictions.

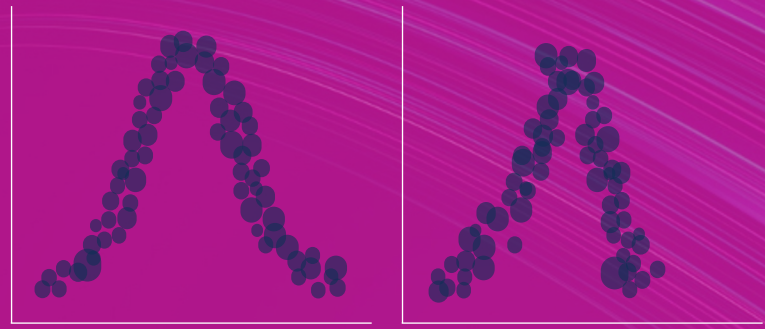
Imbalanced or insufficient datasets can be augmented by using GenAI to produce synthetic data. The most effective way to do this is with a Conditional Tabular Generative Adversarial Network (CTGAN). This technique delivers the most complex distribution of data and can include multiple data types in each model.

CTGAN synthetic data has three primary use cases

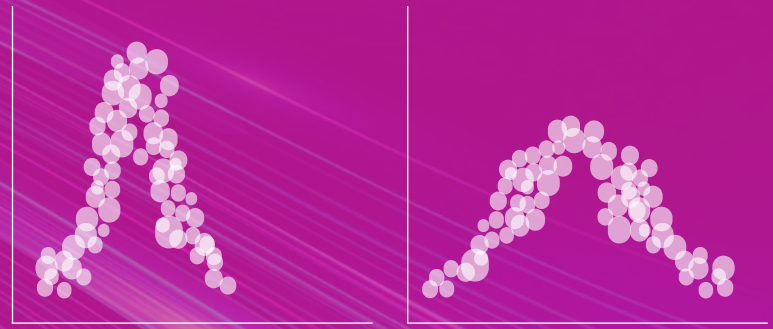
Overcoming compliance issues around data sharing	Improving the predictive accuracy of models	Simulating hypothetical scenarios

What is GenAI synthetic data? Synthetic data is produced from real data, so it retains all the relevant statistical characteristics.

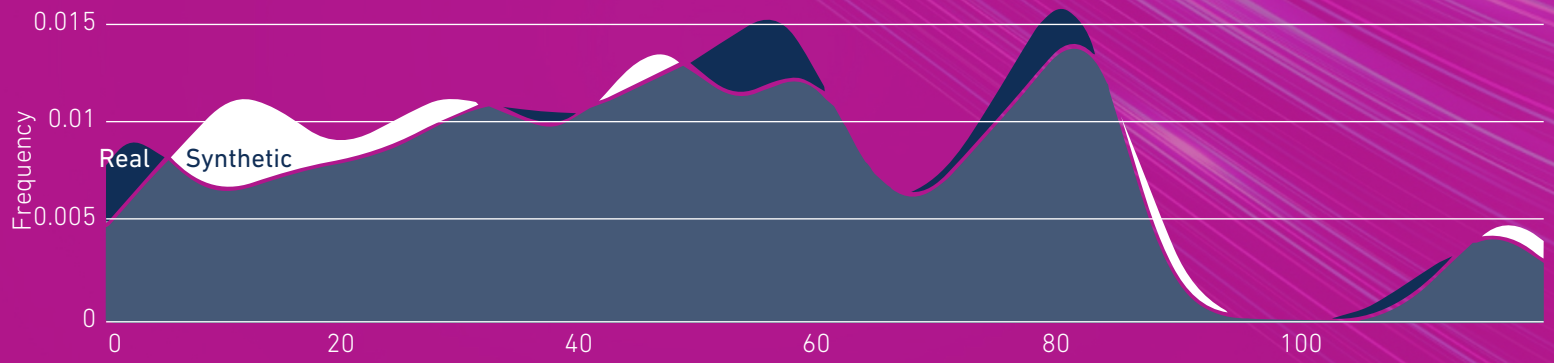
Original data



Synthetic data



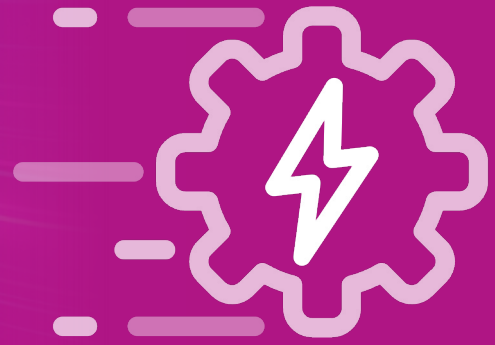
Real vs. Synthetic data



Overcoming compliance issues around data sharing

As restrictions around privacy and data sharing become more demanding, synthetic data becomes more appealing as an alternative to real data. The advantage of synthetic data is that it represents the same statistical characteristics as the original dataset, yet it is fully anonymised, and therefore there are no restrictions associated with sharing it.

This is one of the biggest benefits that synthetic data provides in that it removes any Personally Identifiable Information (PII) in a dataset, and the original sensitive information cannot be inferred from the synthetic version. As a result, it can be freely shared between business units and across borders as it complies with current privacy regulations around the world.



Improving the predictive accuracy of models

According to Experian's recent AI research, the biggest data-related challenge that businesses face is a lack of data to assess the creditworthiness of consumer and business customers. This issue can be directly addressed through the use of synthetic data, which accurately mimics the properties of real-world data. However, to train CTGANs to produce highly relevant synthetic data requires a sufficiently representative training dataset.

Experian is uniquely positioned to help our clients produce synthetic data via our comprehensive bureau datasets. Once the data imbalance within a model is identified, we can produce the data needed to correct this imbalance. The results of adding synthetic data to a model can be considerable, in some cases as much as a 20-point improvement in the Gini coefficient.

Fraud detection ML models are another area where synthetic data can deliver large improvements. In real-world fraud training data, the instances of fraud are often around 0.5% to 1%. Whereas the optimal model training data ratio would be around a 50/50 split between genuine customers and fraudsters. Synthetic data can resolve this imbalance to improve the accuracy of the fraud model.

Simulated hypothetical scenarios

Synthetic data can also be used to model hypothetical economic situations, such as periods of elevated inflation or a recession. This data is essential to simulate the impact of these events by stress-testing their portfolios. Understanding how a portfolio will react under these extreme conditions can help organisations identify any potential risks and take steps to mitigate them in advance.

Another area where synthetic data can be highly valuable is when entering new markets. In this situation, businesses may lack sufficient historically relevant data to develop accurate credit risk models. Synthetic data can be used to emulate potential credit behaviour in a new market to improve the predictive accuracy of credit decisions.

4 Improved coding efficiency

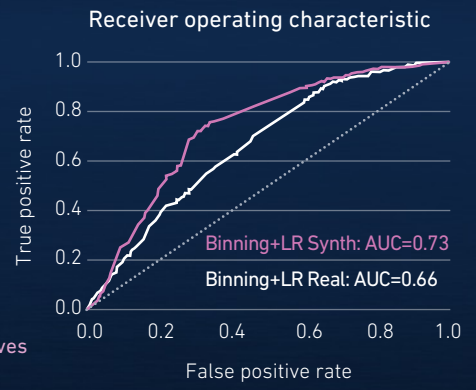
The last GenAI application I would like to mention involves improving coding efficiency. We have integrated an AI assistant into our model development platform that helps you to code faster. Although this use case still has some limitations from a technical point of view, it can significantly enhance productivity when coding.

An important point around this application is that it allows your data science analytics platform to be an enabler for GenAI within your business. With the help of this coding prompt assistant, it becomes much easier to create your own LLMs to perform a variety of specialised tasks.

How does synthetic data enhance model development and accuracy?

- Overcome data scarcity challenges to improve the predictive accuracy of models.
- Remove PII from datasets to comply with data sharing restrictions between business units and across borders.
- Create the data required to simulate hypothetical scenarios.

The graph shows the reduction in false positives achieved by using synthetic data.



CASE STUDY
Tier 1 Italian Bank

A leading Italian bank used a CTGAN synthesiser to create an additional 87% of data to augment their own database for credit risk decisioning.

By incorporating this data into their models, they we able to achieve the following:

11% increase in model performance

$1m estimated increase in annual revenue

Future applications of GenAI

The velocity of GenAI innovation is considerable. In a little over a year, this technology has become a household name, and potential future applications are being discussed in boardrooms and at dining room tables across the globe.

Such rapid uptake means the volume of applications is growing exponentially. If you consider how fast the GenAI algorithms are evolving, it seems likely that we will soon see significant improvements in the accuracy of their outputs and the number of tokens they can ingest. This will further expand how we use this technology. Perhaps the biggest restricting factor for GenAI applications in credit risk assessment – the unexplainable nature of the outputs – will also be resolved.

Here are a few examples of potential GenAI use cases that could have a significant impact in the near future:

Dynamic risk assessment

GenAI could enable more dynamic risk assessment, where credit risk models are constantly updated in real time based on the latest data. This process could be automated so that macroeconomic factors are analysed, categorised, and the relevant data ingested into a model as the events happen.

Personalised credit products

Financial institutions could offer credit products tailored to the specific needs and risk profiles of individuals. By analysing each customer's income, financial goals, debt level and expenses to offer services specifically tailored to their needs. The LLM could then interact with the customer via their preferred communication channel and language in a highly responsive way.

More ethical lending practices

To mitigate the risk of bias in credit models, GenAI could help promote ethical lending practices by analysing training datasets and identifying any potential bias. This would enhance ethical lending in the credit risk assessment process.

Maximise GenAI with Experian

As a global leader in data, analytics and technology, Experian is actively exploring over 40 different use cases for GenAI. Our large multinational teams of data scientists have decades of experience working with AI/ML and have successfully developed and implemented thousands of models for businesses across the globe.

This solid foundation of expertise is a critical factor when exploring the potential that GenAI offers. It gives us an in-depth understanding of the benefits, as well as the challenges, involved with implementing this new technology.

At the core of our purpose is the use of technology to drive automation, efficiency and profitability in a safe and responsible way. Our approach to AI ensures compliance with regulatory requirements for accounting, auditing and model explainability.

We encourage you to reach out to us, to discuss how your business can take advantage of this exciting technology. The GenAI use cases we have highlighted in this guide are only the beginning, and in the coming months, we will continue to update you on the ongoing evolution of this critical technology.

Our team of specialised consultants is ready to help you through each stage of identifying and developing the right GenAI applications for your business. Contact us today to speak to a local consultant and fast-track your automation and productivity with GenAI.

Experian can help you leverage the power of

GenerativeAI

- Best-in-class synthetic data from our comprehensive global datasets
- Simplify model monitoring and diagnostics for peak model performance
- Automate BI data extraction, analysis and categorisation to enhance productivity

Contact us

experian

Data | Analytics | Technology | Expertise