GSMA
Foundry

SEPTEMBER 2025
CASE STUDY

Overcoming **Data Bottlenecks and Scarcity**

Rockfish Data is helping enterprises to desensitise real-world data and to train AI models

Executive Summary

Harnessing real-world data to train AI models can face major technical, commercial and regulatory obstacles. These obstacles can result in both data bottlenecks and data sparsity. Rockfish Data is employing generative AI to help enterprises overcome these challenges. The three-year-old company has developed AI models that can analyse real-world data and then either replace it or augment it with synthetic data.

Rockfish Data's models generate a representation or statistical abstraction of the real-world operational data, which doesn't leave the customer's environment. For example, it could analyse the distribution, magnitude, timing and nature of financial transactions, as well as correlations between them, made by people in specific locations. It would then create a synthetic data set that is faithful to these patterns and correlations, but without any information that could be used to infer anything about the real-world transactions or the people that made them.

Alternatively, Rockfish AI's models can be used to alter the real-world data so as to test a particular hypothesis. They can be configured to change specific fields in the real-world data, such as the take-up rate for an autopilot feature in a vehicle, and then see how that would impact other fields.

Rockfish Data already has some notable customers, including Ford Motor Company, Conviva, US public sector agencies and Deutsche Telekom. It is initially targeting three sectors – cybersecurity/telecoms, financial services and supply chains – and supporting a wide range of use cases, from countering fraud and predictive maintenance to testing product concepts and anticipating scenarios.

So far, the synthetic data produced by Rockfish Data has had a major impact on the effectiveness of AI model training. The company says the accuracy of some models has increased by 20 to 30 percentage points. The solution has also enabled a significant reduction in the time it takes customers to develop new products and services.

Without the right data, artificial intelligence (AI) is far from intelligent. Even the most advanced AI models depend on high-quality and relevant data to work well.

As organisations across the economy strive to harness AI, many are hamstrung by either data bottlenecks or data sparsity or both. That's the view of Muckai Girish, who has 30 years' experience in the tech and telecoms sector, including with AT&T and Juniper Networks.

"You need some data, but it is not with you, it is with somebody else, either with a vendor or a partner or a customer, or another division," Muckai Girish notes. "They are not able to give you the data as is, because of either confidentiality or some sort of regulatory or compliance issue." In many jurisdictions, the sharing of personally-identifiable data is highly regulated. At the same time, many businesses don't want to share commercially-sensitive information, even with partners and suppliers.

> Even though there is a lot of data, in many instances it's like being in an ocean sitting on a boat: You want to drink sparkling water, but the only water that they have is salt water.

Dr. Muckai Girish - CEO and co-founder Rockfish Data

Now, the CEO and co-founder of Rockfish Data, a three-year-old company based in San Francisco Bay Area, Muckai Girish is focused on enabling enterprises to overcome both data bottlenecks and data sparsity. For telecoms operators, which capture sensitive behavioural data, data bottlenecks are a particularly big issue. Telcos may wish to share their data with vendors developing AI models, for example, but are unable to do so because it will include IP addresses and other information that can point to specific individuals.

Data sparsity is a different, but equally challenging, issue. In this case, there may be insufficient real-world data to enable an AI model to test a particular hypothesis or run a specific scenario. This can be a problem if a business wants to gauge the impact of relatively rare events. For example, a lack of real-world data can make it difficult for AI models to stress-test cyber-security or the resilience of a supply chain to a particular geopolitical scenario. Other examples include the need to assess a bank's vulnerability to a new form of fraud or the potential impact of an unusual network fault (in the case of a telecoms operator).

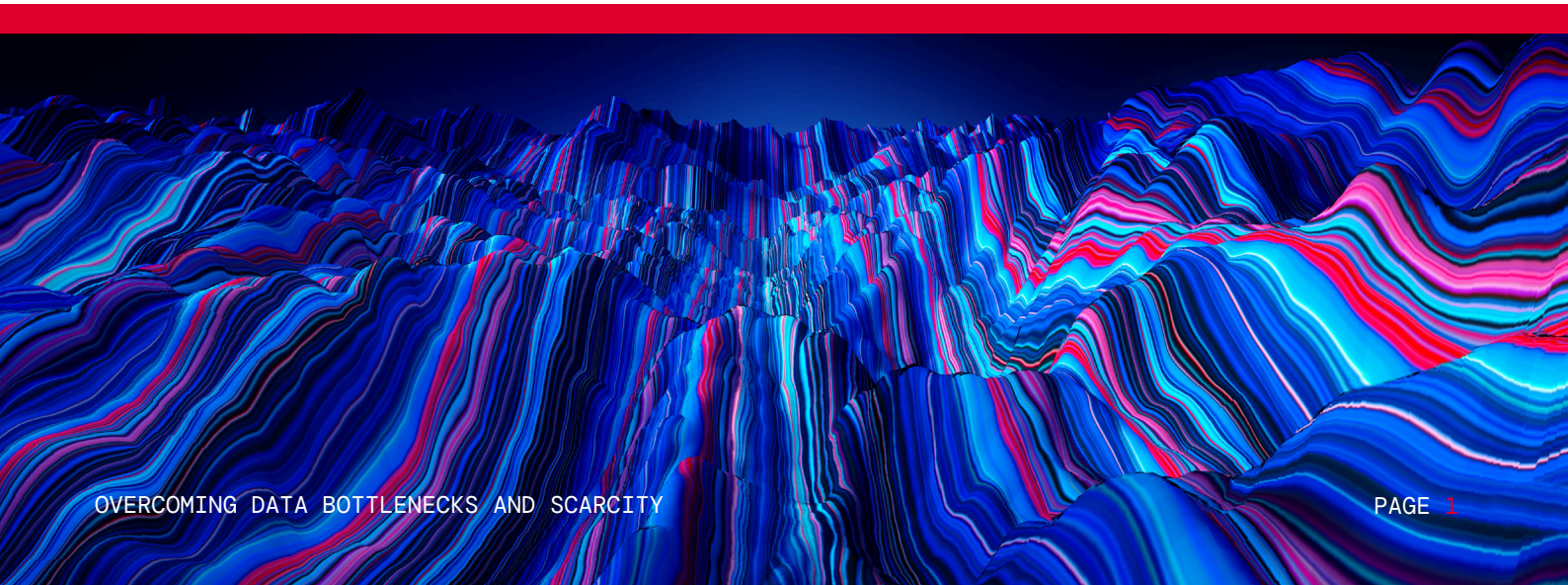
"Even though there is a lot of data, in many instances it's like being in an ocean sitting on a boat: You want to drink sparkling water, but the only water that they have is salt water", Muckai Girish says.

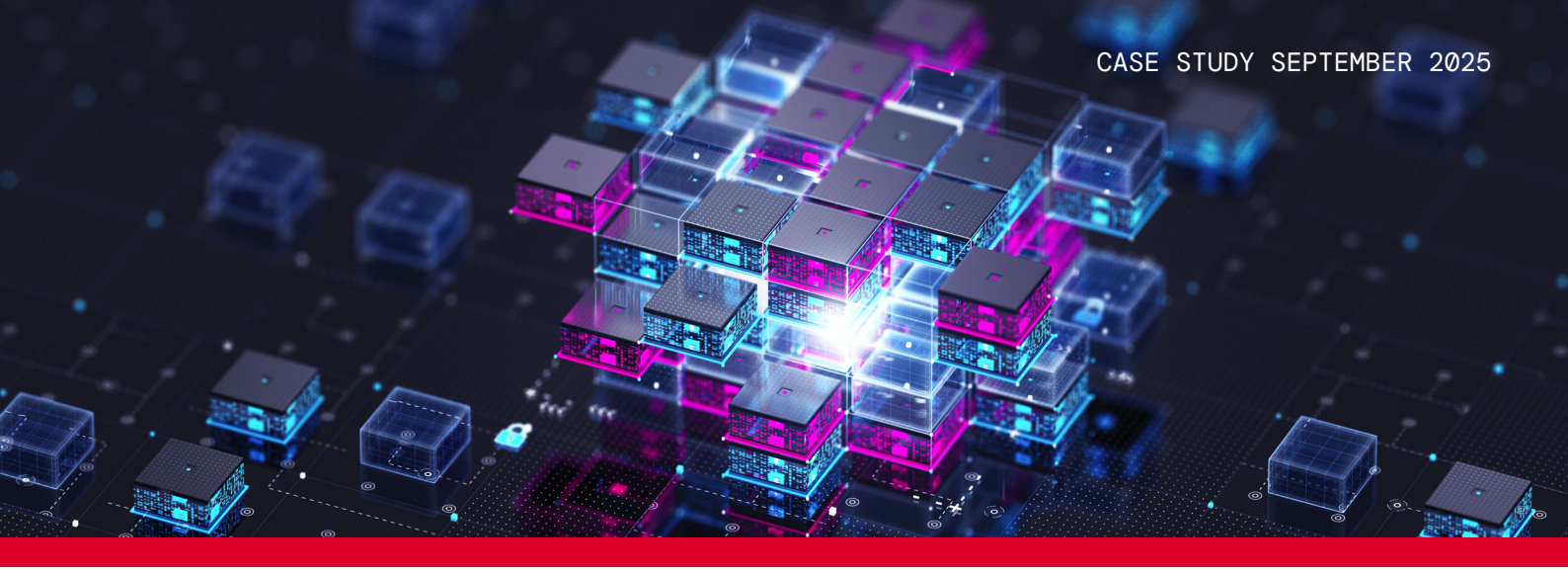
Distilling value from data

This analogy is likely to resonate with telecoms operators - their networks generate oceans of data, but they often struggle to distil value from that. "Unfortunately, they are not able to leverage it to the extent that they could," notes Muckai Girish. "They just haven't been able to use it for everything from customer experience to operating networks well to overall business outcomes. All of that could be done so much better, and AI actually provides a fabric to make all that happen. But you need to really be able to use the data to do that."

Telcos may face regulatory and technical challenges transferring and aggregating data from multiple systems and countries. The cost of storage can also mean that operators need to delete large amounts of data every week, while only retaining limited snapshots of the activity on their networks. As a result, much of the granularity that AI models need to become highly sophisticated may be lost.

For many businesses, data bottlenecks and sparsity will be major obstacles to realising the full potential of AI. Under pressure from investors, companies need to be able to convincingly demonstrate that they can harness AI to improve business outcomes, whether that be in terms of revenue, profitability or





other key performance metrics. AI can, for example, help a business gauge the likely return it will make on specific investments, whether they be in advertising campaigns or new product features, and then make better decisions accordingly. As well as impacting investors' confidence, a failure to embrace AI could see a business become increasingly uncompetitive over time.

In particular, effective use of AI can be the key to maintaining a good customer experience. Muckai Girish gives the example of systems that are designed to help financial services companies detect and prevent fraud. AI is crucial to ensuring that these systems don't block too many legitimate transactions or interactions, while also ensuring that suspicious events are flagged, investigated, and where necessary, blocked. If the system gets this balance wrong, then customers either encounter too much friction or could be defrauded.

Similarly, in the automotive sector, AI is playing an increasingly important role in determining what features to integrate into a vehicle, both to optimise its appeal to certain customer segments and the cost of production. Again, the balance needs to be right or a potential customer may switch to a different brand.

A deep fake for a good cause

To overcome the issues of data bottlenecks and data sparsity, Rockfish Data has developed AI models that can analyse real-world data and then either replace it or augment it with synthetic data. One way to think about this synthetic data is "a deep fake for a good cause," says Dr. Muckai Girish.

Two of his fellow co-founders at Rockfish Data, Dr. Giulia Fanti and Dr. Vyas Sekar, are professors at Carnegie Mellon University, which is at the cutting edge of the development of AI. Rockfish Data's approach is to use generative adversarial networks (GAN) models, supplemented by other AI tools, to generate a representation or statistical abstraction of the real-world operational data, which doesn't leave the customer's environment. For example, it could analyse the distribution, magnitude, timing and nature of financial transactions, as well as correlations between them, made by people in specific locations. It would then create a synthetic data set that is faithful to these patterns and correlations, but without any information that could be used to infer anything about the real-world transactions or the people that made them.

Alternatively, Rockfish Data's models can be used to alter the real-world data so as to test a particular hypothesis. They can be configured to change specific fields in the real-world data, such as the take-rate for an autopilot feature in a vehicle, and then see what the impact would be on other fields.

However, before generating any synthetic data, Rockfish Data performs a significant amount of pre-processing to ensure that the synthetic data its solution produces is both accurate and realistic. This pre-processing might involve straightforward considerations, such as ensuring the model places a city in the right country, or more complex considerations, such as ensuring a proposed vehicle configuration can actually be built within the operational and commercial constraints facing an automaker. "Being able to account for these domain-specific requirements and constraints is a very critical aspect," notes Muckai Girish.

Diverse use cases for synthetic data

Although its engineering team only started work in early 2023, Rockfish Data already has some notable customers, including Ford Motor Company, Conviva, US public sector agencies and Deutsche Telekom. Focused initially on the North American and European markets, Rockfish Data is planning to raise additional venture funding in mid 2026 to help the business grow further.

It is targeting three major groups of customers – cybersecurity/telecoms, financial services and supply chains – and supporting a wide range of use cases, from countering fraud and predictive maintenance to testing product concepts and anticipating scenarios. Although its early customers are a diverse group, they face common challenges. One of the most significant is the effective testing of new products and solutions.

Synthetic data can be employed in place of the "stale data" that is often used to test new product concepts, notes Muckai Girish. In one case he encountered, test data was over ten years old. "They test it and say everything is hunky dory. But when it rolls it out, everything falls apart," he notes. "This is a serious challenge in every industry, and every public sector we work with."

Another popular use case is the development of synthetic data to support product and solution demos – Rockfish Data draws on the real-world data to effectively create a benign deepfake – a demo of a realistic, but synthetic, customer using a specific product or solution.

> They test it and say everything is hunky dory. But when it rolls it out, everything falls apart. This is a serious challenge in every industry, and every public sector we work with.

Dr. Muckai Girish - CEO and co-founder Rockfish Data

To date, the synthetic data produced by Rockfish Data has had a major impact on the effectiveness of AI model training. Muckai Girish says the accuracy of some models has increased by even 20 to 30 percentage points. The solution's support for scenario planning and testing new hypotheses has also led to a significant reduction in the time it takes customers to develop new products and services, he adds. Similarly, there has been a marked reduction in the time it takes customers to create new product demos.

Rockfish Data provides customers with a dashboard that shows how their synthetic data measures up in terms of both protecting privacy and in terms of its fidelity to the original data. These scores are based on a range of factors, such as the number of times lines are repeated in the synthetic data and how close the correlations between fields are to those between the fields in the original data.

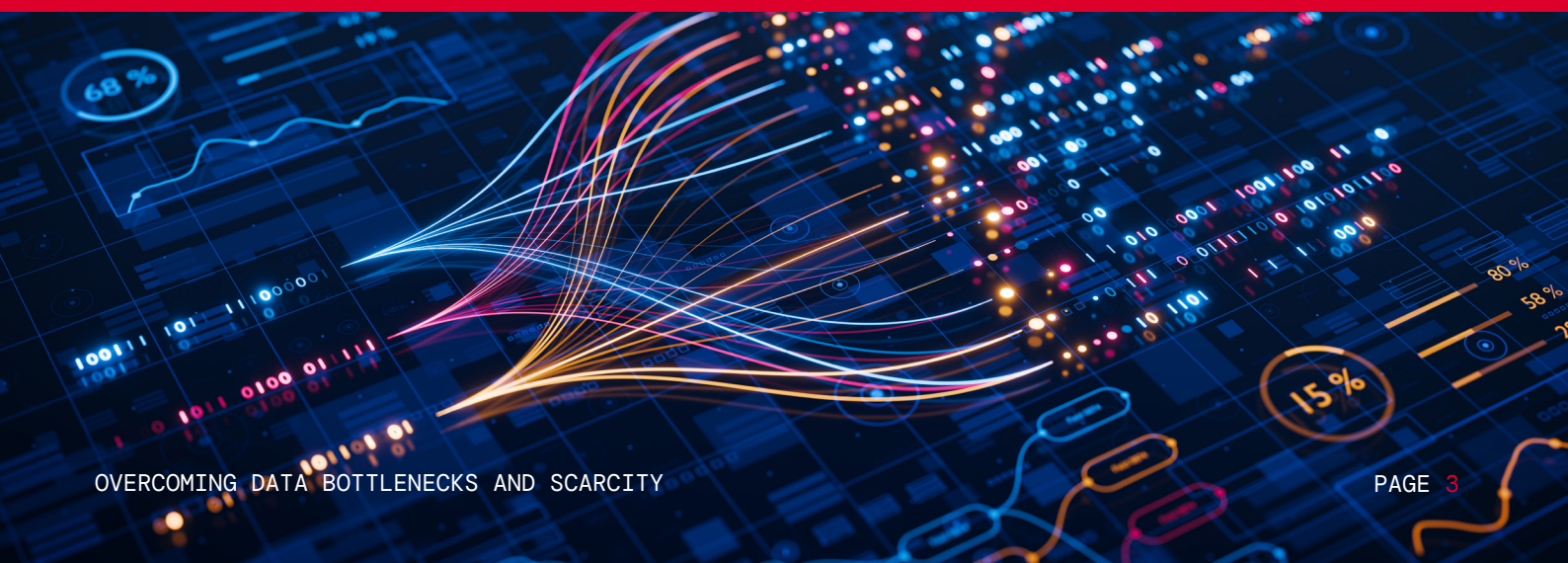
More channels, better products, greater integration

In 2025 and 2026, a key priority for Rockfish Data is developing its channels to market. As well as dealing directly with customers, its solution is becoming available through the hyper-scalers, AWS, Azure and Google Cloud Platform, and through data management specialists, such as Databricks and Snowflake. "As well as wanting to grow the market, we want to ensure there is a clear product market fit for us," adds Muckai Girish. "We have a bunch of initial customers, but we want to make sure that we have this repeatable pattern in multiple ways."

From a product development perspective, priorities include further improving the scale, performance and efficiency of Rockfish Data's AI models, as well as their usability. The start-up is working on a conversational interface, that will complement the existing SDK and web user interface. For example, a telecoms engineer will be able to verbally instruct the system to "generate me one million data points for a radio network that is in a certain part of the town with outages being 10 times what they are normally," explains Muckai Girish.

Rockfish Data is also working on integrating its models with various other systems, as it seeks to ensure it is employing the optimum combination of models for a particular use case. In the telecoms industry, it is engaging with some of the major network vendors to integrate its models into their roadmaps, as they develop increasingly automated radio and core network systems.

Muckai Girish, whose parents worked on telecoms switchboards when he was a child in India, believes the industry's ability to capture valuable data is a largely untapped gold mine. "There is so much potential for the telecom ecosystem because it



connects people. And without connectivity, none of us can even breathe anymore," he says. "It is a world where telcos can literally make such a big difference to your life. But using data correctly has been a significant challenge. Now, we are at a stage where AI and data approaches, like ours, can transform that very quickly and make telcos far more sophisticated."

> It is a world where telcos can literally make such a big difference to your life. But using data correctly has been a significant challenge. Now, we are at a stage where AI and data approaches, like ours, can transform that very quickly and make telcos far more sophisticated

Dr. Muckai Girish - CEO and co-founder Rockfish Data

About the GSMA

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Representing mobile operators and organisations across the mobile ecosystem and adjacent industries, the GSMA delivers for its members across three broad pillars: Connectivity for Good, Industry Services and Solutions, and Outreach. This activity includes advancing policy, tackling today's biggest societal challenges, underpinning the technology and interoperability that make mobile work, and providing the world's largest platform to convene the mobile ecosystem at the MWC and M360 series of events.

For more information, please visit the GSMA corporate website at gsma.com

Follow the GSMA on LinkedIn: @GSMA.

About the GSMA Foundry

The GSMA Foundry is the go-to place for cross-industry collaboration and making positive change happen, supported by leading technology organisations and companies. By bringing together members and key industry players, engaging, and unifying the end-to-end connectivity ecosystem, the GSMA is solving real-world industry challenges.

Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. This enables the mobile industry's mission: to connect everyone and everything to a better future.

Find out more, or submit a new project idea, at gsma.com/Foundry

Rockfish Data

Enterprises are stymied by data bottlenecks including sharing constraints and sparsity, while productising AI, testing AI agents and making operational decisions. Rockfish's outcome-centric synthetic data platform is purpose built for cross-silo data sharing and smart data augmentation for overcoming these bottlenecks. Rockfish Data's foundational innovation originated based on years of research at Carnegie Mellon University, the birthplace of AI. Rockfish Data is trusted by several enterprises and public sector agencies.

For more information, visit: www.rockfish.ai

About this case study

This case study is for information only and is provided as is. The GSM Association makes no representations and gives no warranties or undertakings (express or implied) with respect to the study and does not accept any responsibility for , and hereby disclaims any liability for the accuracy or completeness or timeliness of the information contained in this document. Any use of the study is at the users own risk and the user assumes liability for any third party claims associated with such use.