



# DATA-POWERED INNOVATION REVIEW

WAVE VI  
2023

Capgemini 

# FOREWORD



**NIRAJ PARIHAR**

CEO, Insights & Data,  
Capgemini

I'm happy to invite you to a deep dive into the fascinating world of data-powered innovation, where the possibilities are as vast and diverse as the coral reefs of the ocean.

Yet again, the cover page of this edition of our innovation magazine (which is not generated by AI, mind you) inspired me to explore the relationship between nature and the technology business environment. The cover visual features a striking, zoomed-in image of a coral reef – a reminder of the intricate ecosystems that underpin our planet's biodiversity. Yet, this image is much more than just a beautiful photograph. It serves as a metaphor for the complex business ecosystem in which we operate, where companies, customers, employees, and partners are interconnected in myriad ways.

To navigate this delicate and valuable ecosystem successfully, we need data and AI. Data makes us collaborate to better understand and act on market trends, consumer behaviors, business performance, and so much more. AI augments our abilities by enabling us to analyze vast amounts of data, make better predictions, and automate routine tasks; next of course to unleashing the knowledge and creativity of both the organization and all of us individually.

The three main themes of this edition underpin this.

Where we see the emerging impact of data and AI on strategy, it often revolves around the dynamics of collaborative, increasingly diverse data ecosystems. But just as much, we see the ownership of data coming closer and closer to the business side, managing data as a first-class product.

"Augmentation" sure is the word of the day. You will therefore not be surprised – with such a strong AI undercurrent going on – to see various articles homing in on the perks and perils of generative AI.

Then there is always the quest to activate data for sustainability and a better society, or as we simply put it: "positive futures." We feature for example an introduction to the 6th edition of our Global Data Science Challenge, focusing on biodiversity.

So, welcome to our "reef-reshing" world of data and AI innovation. I hope you'll find it making a splash!

# EDITOR'S NOTE



RON TOLIDO

CTO, Insights & Data,  
Capgemini

The data-powered Innovation Review is back with a bang, now in its 6th edition. Packed with cutting-edge ideas and best practices, this treasure trove is all about unleashing the full innovative potential of data and AI in business.

We said it before: for Capgemini's data-powered innovation movers and shakers, this sleek magazine is not the guide to life, the universe, and everything (we all know it's 42 anyway). Instead, it's a powerful catalyst for collaboration, bringing together sharp minds to brainstorm innovative topics, trends, and stories, tap into top experts and technology partners, and jointly create unique content. And the action doesn't stop there. There is a long, swishing tail of follow-up posts and events to keep the momentum going and drive the conversations forward.

Nowadays, it seems every second (and third, and fourth, and fifth...) innovation discussion revolves around generative AI. We take a bit of pride in having flagged the topic right from our first edition – focusing on the use of generative AI for creativity, research and development, legacy code migration, and language support. And sure enough, the **Data for Augmentation** section of this edition is at the core, diving deeper into both the perks and the perils of the new wave of generative AI that we are all part of now. As you will appreciate, a part of the content may be outdated before it is even published. This is the pace of change we are currently seeing. Ironically, this time "no AI was harmed" in the making of either the articles or the visuals of this edition – which rather sooner than later will become something noteworthy.

Well, is there still life outside of generative AI? We think so.

In the first section of the magazine – **Data for Strategy** – we explore how every business now is a data business, bringing the power of data not only as a high-priority boardroom topic, but also at the middle of a cultural and behavioral shift. The principles of the data mesh – as technically as they have been interpreted by some – manifest themselves increasingly

on the business side, emphasizing federated governance, domain-owned data and (arguably above all) managing data as a serious, top-class product, just like the other products and services of the organization. And look at how collaborative data ecosystems evolve, all the way up to the notion of "data cleanrooms" in which even the fiercest of competitors can share their machine learning data for mutual benefits.

In what is a tradition by now, we finalize the magazine with a section about how data and AI can be used to create a better planet and society. In **Data for Positive Futures**, we display some inspiring stories, including one coming from the winners of our internal T4PF innovation hackathon, which shows how data – in surprising ways – can be activated to support the biodiversity of Miyawaki "urban forests." And talking about hackathons: the topic of our 2023 Global Data Science Challenge – this time in collaboration with AWS and the Dutch Naturalis Biodiversity Center – is introduced as well. It's about bugs. But not in an IT way. Or is it?

In any case, this magazine is a product of some serious brainpower, with leading experts from Capgemini, a very modest amount of generative AI (there: admitted), and key partners like Google, Starburst, Microsoft, Snowflake and Databricks all chipping in. And the best part? The contributors are more than happy to help you with your own data-powered innovation journey, so don't hesitate to reach out to them. As stated, there's even more excitement in store, with follow-up articles and live events featuring the makers and some special guests.

Stay tuned!

## DATA FOR STRATEGY

### REWARDS VERSUS RISKS IN THE METAVERSE: 6

Balancing personalization with data privacy

*Liz Henderson, Capgemini*

### POWER TO THE ANTS: 9

Action Mesh activating

*Malgorzata Ilczyk &*

*Arne Rossmann, Capgemini*

### SHARING WITHOUT SHOWING: 13

Data clean rooms allow for unprecedented collaboration

*Jennifer Belissent, Snowflake*

### DATA AS A PRODUCT MEETS DATA GOVERNANCE: 18

A match made in heaven

*Roosa Sääntti & Minna Lind, Capgemini*

### DATA IS THE BUSINESS: 22

Driving a collaborative data

ecosystem

*Dinand Tinholt, Reshma Bhatt*

*Capgemini & Rob Saker, Databricks*

### BECAUSE I WAS INVERTED: 26

Getting data on time

*Hemavathi Thiruppathi &*

*Yashowardhan Sowale, Capgemini*

### UNLEASHING THE DATA MESH REVOLUTION: 31

Empowering business with cutting-edge data products

*Dan O'Riordan, Capgemini & Andy*

*Mott, Starburst*

### THE RACE TO PRODUCTIZE DATA: 35

The slow and steady tortoise won't win

*Aujun Sud, Capgemini*

## DATA FOR AUGMENTATION

### GENERATIVE AI FOR THE ENTERPRISE: 41

Beyond the hype

*Mark Oost, Marijn Markus,*

*Dr. Sergey Patsko, Capgemini*

### THE GENERATIVE AI ARMS RACE: 45

Why responsible AI matters now more than ever

*Lee Hickin, Microsoft &*

*Aruna Pattam, Capgemini*

### BIGGER, SMALLER, SMARTER, SAFER: 48

The competing pressures disrupting AI

*Mark Roberts & WeiWei Feng,*

*Capgemini*

### DISRUPTION MANAGED: 52

Bringing "gen AI-in-the-room"

*Christopher Scheefer, Capgemini*

### I SEE YOU: 56

Opening the AI black box

*James Wilson, Pantelis*

*Hadjipantelis, Capgemini*

### GENERATIVE AI: 60

A powerful tool, with security risks

*Matthew O'Connor, Google*

*Cloud*

### SERENDIPITY SYSTEMS: 63

Design for AI

*Charles Aubert &*

*Chloe Cheau, Capgemini*

## DATA FOR POSITIVE FUTURES

### FROM BIG DATA TO RIGHT DATA: 68

Data platforms in the age of eco-responsibility

*Nicolas Ydder & Nicolas*

*Claudon, Capgemini*

### GLOBAL DATA SCIENCE CHALLENGE FOR A SUSTAINABLE FUTURE: 71

The Biodiversity Buzz

*Dr. Dan Stowell & Bart Braun,*

*Naturalis Biodiversity Center*

### DATA MASTERS DRIVE THE FUTURE OF INNOVATION: 74

Forward-thinking

companies scale and

transform with data and AI

*Ivar Aune, Thomas Svahn &*

*Magnus Carlsson Capgemini*

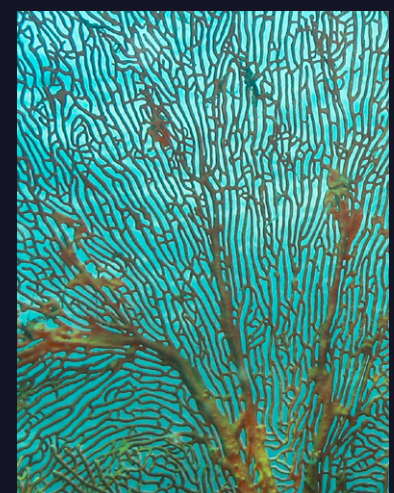
### CREATING URBAN FORESTS: 77

With a little help from data analytics and

visualization

*Anil Kandpal, Douglas Hoover*

*& Daniel Freytag, Capgemini*





# DATA FOR STRATEGY

## REWARDS VERSUS RISKS IN THE METAVERSE: 6

Balancing personalization with data privacy  
*Liz Henderson, Capgemini*

## POWER TO THE ANTS: 9

Action Mesh activating  
*Malgorzata Ilczyk & Arne Rossmann, Capgemini*

## SHARING WITHOUT SHOWING: 13

Data clean rooms allow for unprecedented collaboration  
*Jennifer Belissent, Snowflake*

## DATA AS A PRODUCT MEETS DATA GOVERNANCE: 18

A match made in heaven  
*Roosa Sääntti & Minna Lind, Capgemini*

## DATA IS THE BUSINESS: 22

Driving a collaborative data ecosystem  
*Dinand Tinholt, Reshma Bhatt Capgemini & Rob Saker, Databricks*

## BECAUSE I WAS INVERTED: 26

Getting data on time  
*Hemavathi Thiruppathi & Yashowardhan Sowale, Capgemini*

## UNLEASHING THE DATA MESH REVOLUTION: 31

Empowering business with cutting-edge data products  
*Dan O'Riordan, Capgemini & Andy Mott, Starburst*

## THE RACE TO PRODUCTIZE DATA: 35

The slow and steady tortoise won't win  
*Aujun Sud, Capgemini*

# REWARDS VERSUS RISKS IN THE METAVERSE

Balancing  
personalization  
with data privacy



LIZ HENDERSON

Executive Advisor, Capgemini



As deeper immersion into the metaverse promises to seamlessly connect physical and virtual worlds, the opportunity for companies to provide personalized offerings to customers seems infinite. But brand reputation is at stake regarding the limits of data compliance, especially as privacy legislation evolves alongside technology.

The metaverse is rapidly gaining in popularity as a new digital space for commerce, entertainment, and social interaction. McKinsey & Company estimates that the worldwide value of the metaverse could reach [\\$5 trillion by 2030](#), and more than \$120 billion has already been invested in this sphere. Many start-ups receive funding at all rounds for producing live-stream 3D hologram images. Also, of course, there is the recent announcement of Apple's Vision Pro, its first 'spatial computer' that leaves no one untouched.

Its potential is [summed up well by Satya Nadella](#) chairman and CEO of Microsoft: “The metaverse is here, and it’s not only transforming how we see the world but how we participate in it – from the factory floor to the meeting room.” But as technology matures to shape the metaverse of the future, associated risks also increase significantly for companies that collect user data to expand their brand’s reach.

## Personalizing the metaverse

Imagine working in a laboratory where you’re virtually guided through health and safety protocols without being exposed to a hazardous environment. Or if you’re learning to manage internet shopping deliveries for a fulfillment center but don’t need to travel to a classroom for scheduled training sessions; instead, you can be “hands-on” in a virtual setting that teaches you how to identify, sort, and ergonomically lift various items.

And on your downtime, the metaverse opens up another world of experiences. Like playing virtual tennis against Roger Federer at the US Open and feeling the power of his 140 mile-per-hour serve. Or enjoying an immersive shopping excursion at home, “trying on” luxury goods in a boutique powered by augmented reality (AR) and virtual reality (VR).

The metaverse continues to evolve, and personalization has emerged as a key differentiator for users, who expect experiences tailored to their individual needs and preferences. Consumers expect companies to be well-informed about their personal information during a service interaction. As enterprises expand their offerings to capitalize on emerging revenue streams in this

*“The metaverse is here, and it’s not only transforming how we see the world but how we participate in it – from the factory floor to the meeting room.” – Satya Nadella, chairman and CEO at Microsoft*

new virtual economy, reputational risk must also be top of mind.

In recent years, even well-established tech giants have been hacked and, in March 2023, OpenAI announced a [data breach of ChatGPT](#), the company’s AI chatbot. These examples represent a violation of trust with customers, and the fallout can have far-reaching effects. For public companies, [stock prices often spiral post-breach](#), while private companies can face hurdles such as raising capital from risk-averse investors.

## Engineering a cohesive virtual user experience

The training and entertainment use cases outlined above are already achievable. Yet, the next level of potential for companies and customers alike lies in the seamless integration of these disparate experiences.

So, what does that next-level experience look like?

Imagine a virtual shopping delivery where you drive to your customer’s address, observing the

neighborhood children rallying with Roger Federer. Then you cruise past a testing facility, getting glimpses of new employees learning how to handle hazardous materials, thereby ensuring the safety of the community, and avoiding dreaded emergency evacuations.

Realizing such an integrated vision requires significant collaboration across multiple companies, creators, and policymakers. This type of partnership is indeed possible, as evidenced by the financial industry’s achievement with open banking. Despite the inherent complexity and risk in this sector, financial institutions have teamed up to ensure the safe transfer of funds and user data between entities. They’ve also adopted strict security and compliance measures for sharing user data. Privacy remains paramount, while customers have greater control over their finances with no need to visit a brick-and-mortar bank.

To engineer a seamless, user-centric experience, metaverse developers need data that is accurate and current. This also amplifies the importance of data management and governance when providing personalization, while mitigating privacy risks at the outset.

## Creating data intelligence from information

Personalization is achieved in the metaverse by employing data to customize experiences for individual users. If you’ve already purchased a luxury item from the comfort of your couch, immediately getting ads for the same product on your device isn’t likely to engender trust in how a brand is using your data.

# METAVERSE  
# DATAMANAGEMENT  
# PERSONALIZATION  
# DATAGOVERNANCE

# Innovation takeaways

## PERSONALIZATION IS A KEY DIFFERENTIATOR IN THE METAVERSE

It can be achieved through the collection of accurate and up-to-date data, such as user preferences and purchasing and activity history.

## PRIVACY IS ONE OF THE MOST SIGNIFICANT CHALLENGES IN DATA MANAGEMENT

As the metaverse expands, privacy is paramount and users must maintain their rights and control over their data, aligned with laws and regulations. Developers must work closely with policymakers and creators to ensure data is managed and governed effectively.

## THE USER EXPERIENCE IS MISSION-CRITICAL

Developers must ensure that they don't undermine the user experience.

Companies need to be smart and selective about the data they use. For instance, harnessing the power of first-party data from their own channels and audiences – such as user preferences, purchasing history, activity history, and more – can help developers and marketers deliver accurate personalization. Additionally, making managed and governed data available to teams leads to insights-driven analytics, which allows for a better understanding of user behaviour in the short and long term.

## Addressing data-management challenges in the metaverse

The metaverse is decentralized, making it difficult for developers to manage the sheer volume of data and ensure its currency, reliability, and accuracy. And again, collecting and storing user data comes with responsibility. The complexity and related risk are increased for companies that operate globally, where laws and regulations vary. Businesses offering products or services in parts of the US and Europe, for instance, need to be aware of their obligations under the California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR).

Sometimes, perception becomes reality for brands. If an enterprise's practices on user data don't appear to be airtight, it might find itself in trouble over privacy concerns, like social media giant TikTok has. Some governments have recommended limits around use of the app, while others have imposed full-scale bans. In the digital world, cases are already being litigated regarding intellectual property and harassment. Companies must protect their brand reputation and investment in this emerging world.

Data management and governance solutions are crucial. Developers must ensure they adhere to privacy regulations, manage data effectively, and create a user experience that is enjoyable and engaging. Collaboration and communication across multiple companies, creators, and policymakers opens a world of opportunity while ensuring both success and safety in the metaverse.



# POWER TO THE ANTS:

## Action Mesh activating



**MALGORZATA ILCZYK**

Enterprise Architect & Azure Solutions Architect, Capgemini



**ARNE ROSSMANN**

Chief Architect Data & AI for Intelligent Industry, Capgemini



The Action Mesh is a game-changer for various industries, providing a more efficient, responsive, and resilient system. The benefits of this new model are clear, including greater efficiency, higher productivity, and faster decision-making. As technology continues to advance, we can expect to see more industries moving towards a decentralized model of data processing and decision-making, resulting in a more connected and responsive world.

In recent years, we have experienced a rapid increase in the amount of data generated by businesses, organizations, and individuals. As a result, we have also seen a rise in the need for data processing and analysis to extract insights and drive decision-making, ideally on-demand and in real-time. Traditionally, this has been done through centralized data platforms, which serve as a hub for collecting, storing, and processing all types of data from inside and outside the organization.

However, with advancements in local computing power and connectivity, we are now witnessing a paradigm shift towards decentralization in data collection, processing, and action-taking. This means that decisions can now be made where

the action occurs, directly on the edge, without relying on a central cloud platform.

### 5G: The catalyst for real-time action

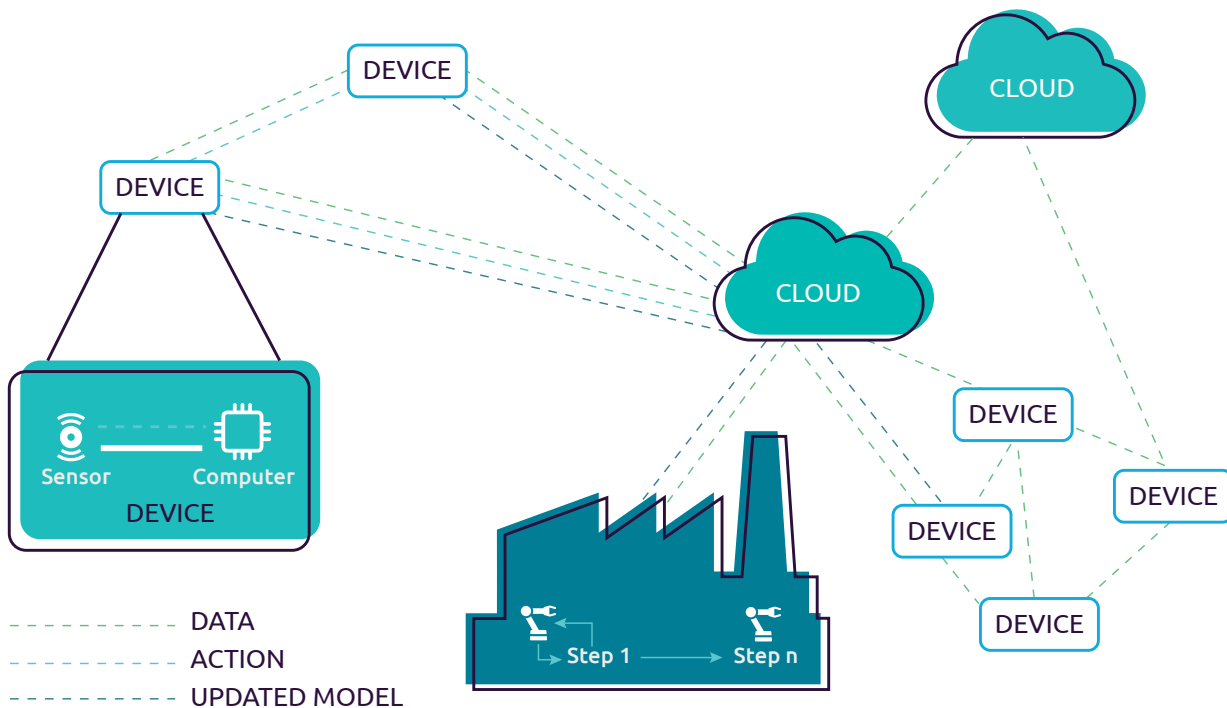
With 5G and the massive increase in computational performance, smaller devices are becoming increasingly capable of [performing complex tasks without relying on a central cloud platform](#). 5G is a key technology enabler for what we call an "Action Mesh," as it brings:

- high bandwidths, with speeds up to 10 Gbps, up to 100 times faster than 4G
- ultra-low latency network in broad areas

- higher device coverage, with up to 100 times more devices per unit area than 4G.

Add to that a [90 percent reduction in energy usage](#) compared to 4G, thus increasing the efficiency of data collection and processing on the edge.

With these advancements, there is no longer a need to slow down decision-making, and actions drivers can be moved from the centralized data platform directly to the edge, enabling automated actions in response to data in real-time.



## THE ACTION MESH IS AN ACTIVE NETWORK OF ACTION-TAKING AND CONNECTED PARTIES

It is essential to note that these enabled devices do not act in isolation: it's the complete opposite. The parties within the Action Mesh are extensively connected, sharing data, and use regularly updated AI models, ensuring that all actions taken are known to the enterprise and can be distributed across the mesh for AI model improvements.

With the latest technologies like [NVIDIA's GPU-accelerated solutions](#) or [Microsoft's multi-access edge compute](#), edge AI models can be trained on large amounts of data in the cloud and then deployed to local devices for real-time decision-making. This means that devices can respond to data immediately. For example, self-driving cars can make decisions about steering, braking, and accelerating based on local data from cameras and sensors, without sending data first to a central server for processing.

In combination with 5G, these edge AI technologies accelerate the digital transformation of enterprises across all industries. 5G provides the underlying connectivity for billions of devices, extending the reach of AI algorithms and applications to all connected objects at the edge, enabling new use cases and new markets.

### Action Mesh activating – A solution for diverse business areas

The shift towards decentralization has significant implications for businesses and organizations in various industries. One example of this is seen in the mining and quarrying industry, where [Volvo Autonomous Solutions](#) has been leading the way in downsizing its fleets of mammoth machines and replacing them with smaller,

*“The Action Mesh is the future of decentralized intelligence, bridging the gap between data and real-time decision-making, and transforming industries with unprecedented levels of efficiency, productivity, and connectivity. Powered by 5G and the latest cutting-edge technology, the Action Mesh is set to revolutionize the way we live, work, and interact with the world around us.”*

electric-powered vehicles. This change is possible due to the decentralization of data collection and decision-making, which has enabled more efficient and cost-effective management of these fleets.

Another critical area of use cases can be found in the context of [Smart Cities](#). By having the capability for vehicles and objects to communicate and share their taken actions, scenarios on passenger enablement (such as warning on weather conditions, optimization of passenger and vehicle transit flow, etc.) arise. This leads to an even further enhanced “user experience” within the city and making life in urban areas more enjoyable. For instance, cameras and weather sensors can detect if passengers want to cross streets during harsh weather, traffic lights for cars can be adjusted, and the car traffic in the area can be slowed down to give passengers more time without causing traffic jams. These opportunities are not only found within cities, but also on a broader scale. [The Hamburg Port Authority](#) has multiple projects running to

leverage the possibilities of these new technologies.

The benefits of the Action Mesh are clear, including greater efficiency, higher productivity, and faster decision-making. By distributing data processing and decision-making capabilities across a network of connected devices, we can create a more efficient, responsive, and resilient system. As technology continues to advance, we can expect to see more industries moving towards a decentralized model of data processing and decision-making, resulting in a more connected and responsive world. While there are still challenges to be addressed, the positive impact of this new model is undeniable.

#ACTIONMESHACTIVATING #5G  
#EDGEAI #INNOVATION  
#DATAPOWERED  
#REALTIMEDECISIONMAKING

# Innovation takeaways

## 5G AS ENABLER

With 100 times faster bandwidth, ultra-low latency, and 100 times more devices in an area, 5G is the enabler for Action Mesh activating.

## AI ON THE EDGE IMPROVES ACTION TAKING

Massive improvements in edge computing, including GPU-powered devices, enables computation at the edge.

## APPLICABLE IN A MULTITUDE OF USE CASES

With use cases spanning all sectors, from autonomous driving to manufacturing and logistics to smart cities, Action Mesh activating is the solution for many current and future challenges.

## Or, as ChatGPT summarizes it:

“The Action Mesh is a game-changer for various industries, providing a more efficient, responsive, and resilient system. The benefits of this new model are clear, including greater efficiency, higher productivity, and faster decision-making. As technology continues to advance, we can expect to see more industries moving towards a decentralized model of data processing and decision-making, resulting in a more connected and responsive world. While there are still challenges to be addressed, the positive impact of this new model is undeniable. Embracing the Action Mesh with cutting-edge technology will undoubtedly pave the way for a brighter, more innovative future.”

*Disclaimer: This article was written with the help of GPT-4*

# SHARING WITHOUT SHOWING

Data clean rooms  
allow for  
unprecedented  
collaboration



JENNIFER  
BELISSENT

Ph.D., Principal Data Strategist,  
Snowflake



Imagine the potential for secure data collaboration. With the boundaries between different companies, organizations, and entire industries blurring, the use cases are endless. Organizations can perform joint data analysis and train machine-learning (ML) models while ensuring that confidential information will stay protected from their sharing partners. It's all happening in the world of data clean rooms.

Pharmaceutical companies [can identify the best hospitals for clinical trials](#) with a look-alike analysis against patient records. Insurance companies can collaborate to identify fraudulent claims. Media outlets can offer [premium placement to advertisers](#) to ensure targeted messaging. Loyalty programs can [deliver truly personalized services](#) across hotels, airlines, and other services. Telecom operators can collaborate with location data to enrich those personalized services. Emergency and [social services](#) can [collaborate to help those in need](#).

Yet in many cases, the relevant data is personal information, and protected by privacy laws and bonds of trust. How can that data be shared?

## The use cases for secure collaboration with data clean rooms are endless

Imagine the following scenario.

A crowd of spectators is watching a big game and the teams are tied. The tension mounts. The fans grow restless. He shoots. He scores! The roar of the crowd can be heard all the way down the neighborhood street. And all the consumer brands want to know who is watching and how to reach these audiences. Yet, these sports fans are watching the game in the privacy of their homes, and the network they're watching on must legally protect their data.

How can these media outlets share their viewer data – or the insights

*“Sharing without showing? You bet!”*

from it – without violating data protection laws and the trust of their subscribers?

It turns out that a similar question was posed by an academic in the early 1980s. Professor Andrew Yao introduced the problem: Alice and Bob, both millionaires, want to know which of them is richer but neither wants to reveal his or her exact wealth. Through complex mathematical proofs, [Yao's Millionaires' problem](#) was solved, proving it is possible to share insights without showing the underlying data. Fortunately, modern methods do not require arduous manual calculations.

## Increased demand for data sharing

For potential advertisers or anyone who wants to collaborate with data, that's great news. Data sharing and collaboration deliver business value. A recent Capgemini study, [Data Sharing Masters](#), found that companies with collaborative data ecosystems reported better business outcomes including new revenues, reduced costs, increased productivity, and greater customer satisfaction. And that promise has spurred new data ecosystem initiatives.

Companies have long used their own data to better understand their customers or to improve operations.

Increasingly, data teams turn to external data sources to enrich their internal data and enhance analytics. Budgets for external data are significant and growing. In a [recent survey](#) conducted by external data platform Explorium, 22 percent of respondents said they were spending more than \$500,000 on external data, with 13 percent saying they spent more than \$1 million (up from 7 percent from a similar survey in 2021).

Customer data was the number one type of data acquisition: 52 percent purchased data on companies, followed by 44 percent purchasing demographic data. And the number of sources has grown as well: 44 percent of firms acquire external data from five or more providers. That's up from only 9 percent the previous year. However, procuring external data is not without challenges, with regulatory constraints often topping the list. Concerns about GDPR or other privacy regulations loom large, and for good reason.

## Introducing modern data clean rooms

Not long ago, data sharing meant copying and sending files to a partner. That practice certainly complicated data governance. Short of a manual audit, knowing who accessed the data and for what purpose was impossible. Now, using the principles demonstrated by Yao's millionaires, two or more parties can derive insights from data without revealing the underlying information.

With a [Snowflake Global Data Clean Room](#), each party controls its own data, allowing governed, controlled analytics by other parties. That is to say, each party specifies who can access the data and for what purpose. Let's take a look at how it would work with Yao's two millionaires, Alice and Bob.

First, each party creates a table with the data to be shared. Then one party, let's say Bob, creates a table

to store allowed statements. This is where the queries that Bob will allow another party to run against his data will be maintained. He then creates an access policy granting use of these statements, and applies this access policy to his data table.

Next, Bob defines the exact statement or query he will allow, and inserts it into his "allowed statements" table. The statement includes the comparison of their

wealth and the answers that will be returned in each case: "Bob is richer," "Alice is richer," or "Neither is richer." Finally, he grants Alice permission to access and use his data for only this specific purpose. Alice then asks the question in the form of the specified query and receives the response: Bob is richer. Sorry, Alice.

**ALICE**

**1**

```
use role alicel;
create or replace table
alicel.data.my_wealth as select 1110000
as wealth;
```

Alice creates a table with her wealth value.

Alice then asks that question in the form of a query and receives an answer.

**3**

```
select case
when bob.wealth > alicel.wealth then 'bob is richer'
when bob.wealth = alicel.wealth then 'neither is richer'
else 'alice is richer' end
from bob1.data.my_wealth bob,
alicel.data.my_wealth alicel
where exists (select table_name from
alicel.information_schema.tables where
table_schema = 'DATA' and table_name =
'MY_WEALTH' and table_type = 'BASE TABLE');
```

Row	CASE WHEN BOB.WEALTH > ALICE.WEALTH
1	bob is richer

**BOB**

Bob does the same but creates a row access policy that establishes what question Alice can ask and what answer she can receive using allowed statements.

**2**

```
use role bob1;
create or replace table bob1.data.my_wealth as select 1250000 as
wealth;
create or replace secure view bob1.data.my_wealth_v as select
* from bob1.data.my_wealth;
create or replace table bob1.data.allowed_statements (
statement varchar(20000) );

create or replace row access policy bob1.data.data_firewall as
(wealth integer) returns boolean ->
current role() IN ('ACCOUNTADMIN','BOB1')
or exists (select statement from
bob1.data.allowed_statements where statement=current_statement()
);

alter table bob1.data.my_wealth add row access policy
bob1.data.data_firewall on (wealth);

insert into bob1.data.allowed_statements (statement)
values ('select case
when bob.wealth > alicel.wealth then \'bob is richer\'
when bob.wealth = alicel.wealth then \'neither is richer\'
else \'alice is richer\' end
from bob1.data.my_wealth bob,
alicel.data.my_wealth alicel
where exists (select table_name from
alicel.information_schema.tables where table_schema = \'DATA\'
and table_name = \'MY_WEALTH\' and table_type = \'BASE
TABLE\');');

grant usage on database bob1 to role alicel;
grant usage on schema bob1.data to role alicel;
grant select on view bob1.data.my_wealth_v to role alicel;
```

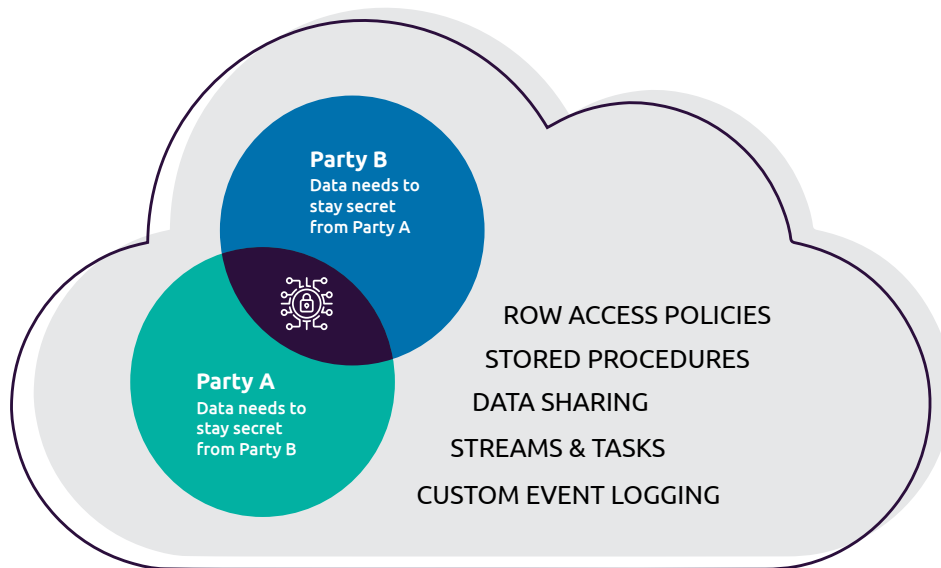
## YAO'S MILLIONAIRES' PROBLEM IN SQL ROW ACCESS POLICY SOLUTION

Now imagine a more realistic business scenario where two companies want to know which customers they have in common – an overlap analysis. They would put the data in tables, establish the statements to compare their customer lists, and specify the information to be returned. Or one company might be interested in finding new prospects among a partner’s customers, and would perform a look-alike analysis comparing customer attributes.

## Data clean rooms transform the ad world

In a real use case, commonly seen in media and advertising these days, brands want to optimize their ad spend through better targeting to specific customers or personas – like the fans watching that exciting game. Media outlets want to offer premium placements by knowing exactly which programming the brand’s customers are watching. Comparing customers is a win-win. However, neither wants to show the

underlying data. The clean room allows them to share without showing. In this case, as illustrated in the diagram, the returned information would include a customer count for each of the media outlet’s programs, but not specific customer data, in order to ensure compliance with privacy regulations. All queries of the data would be monitored and logged for audit purposes.



PROGRAM	HIGH MILEAGE RUNNER	COLLEGE HOCKEY PLAYER	U18 FOOTBALL PLAYERS	65+ TENNIS	HIGH SCHOOL FOOTBALL
WHALE WEEK	2500	2000	500	3000	5500
FRIENDLIEST CATCH	5500	3000	404	220	2000
CYBER RUSH	2200	500	100	8000	500
RULEBUSTERS	3300	<b>7000</b>	850	7100	150
CAKEMAKERS	25	60	8	14	32

REDACTED TO REDUCE RISK OF RE-IDENTIFICATION



# DATA CLEAN ROOMS  
# DATA COLLABORATION  
# DATA ECONOMY  
# DATA ECOSYSTEMS  
# SNOWFLAKE

# Innovation takeaways

## CROSS-INDUSTRY COLLABORATION AND DATA SHARING

A growing trend that's here to stay.

## DATA CLEAN ROOMS FACILITATE JOINT DATA ANALYSIS AND ML

While ensuring that confidential information will stay protected from sharing partners.

## DATA ECOSYSTEMS AND SECURE DATA COLLABORATION

They accelerate value creation.

In the past, this scenario required data to be copied and moved across the AdTech value chain from enrichment to activation to attribution. Not only were there the aforementioned governance concerns, but that data was also immediately stale. With Snowflake, live, near real-time data can be shared where it resides – no copies necessary. Data governance capabilities allow all parties to assign access and use policies that limit both who can query the data and exactly which queries are allowed. Additional capabilities add further security to the clean room. Data can be encrypted, anonymized, tokenized, or pseudonymized with built-in hashing functions, or obfuscated with data masking or by injecting differential privacy.

With today's technology, data clean rooms allow parties across teams, companies, government agencies, and international organizations to collaborate and securely share sensitive or regulated data. As Thomas Edison said, "The value of an idea lies in the use of it." The more data is used, the more value is created. Secure data collaboration accelerates value creation.

# DATA AS A PRODUCT MEETS DATA GOVERNANCE

A match made in  
heaven



ROOSA SÄÄNTTI

Head of Insights & Data Finland,  
Capgemini



MINNA LIND

Head of Data Management,  
Analytics & AI, Insights & Data,  
Finland, Capgemini



The data as a product approach is rapidly becoming popular with organizations elevating their business with data. Data product thinking leverages data as a strategic asset to create innovation and new business value. But without well-functioning data governance as an enabler to these product principles, there is a risk of creating just a data mess. Data products and data governance thus match perfectly for the data-powered enterprise journey.

Enterprise data governance is an important foundation for data ownership and culture. With today's big trend of considering data as a product, data governance is becoming an even more crucial element of success for all organizations. Where data governance sets up the required roles, responsibilities, and operating model enabling and securing high-quality and fit-for-purpose data within an organization, the data product approach creates a new

world view where data can be trusted, built, and served for data consumers to serve their needs. Setting up a robust data foundation, including governance, is a transformation journey shifting the data responsibility and ownership from IT to the business. In simple terms: data not needed by the business is useless data. So, business should be in the driver's seat in this change, taking ownership of their data.

## What is meant by data product thinking?

A data product approach is a paradigm (one of the four principles of the "data mesh") where data products are created to meet business needs. Those products are managed and governed by domain-specific teams. One can think of a data domain in this context as being a high-level grouping or category of data that is important for the organization – for example, customer, product, and location.



## DATA-DOMAIN EXAMPLES

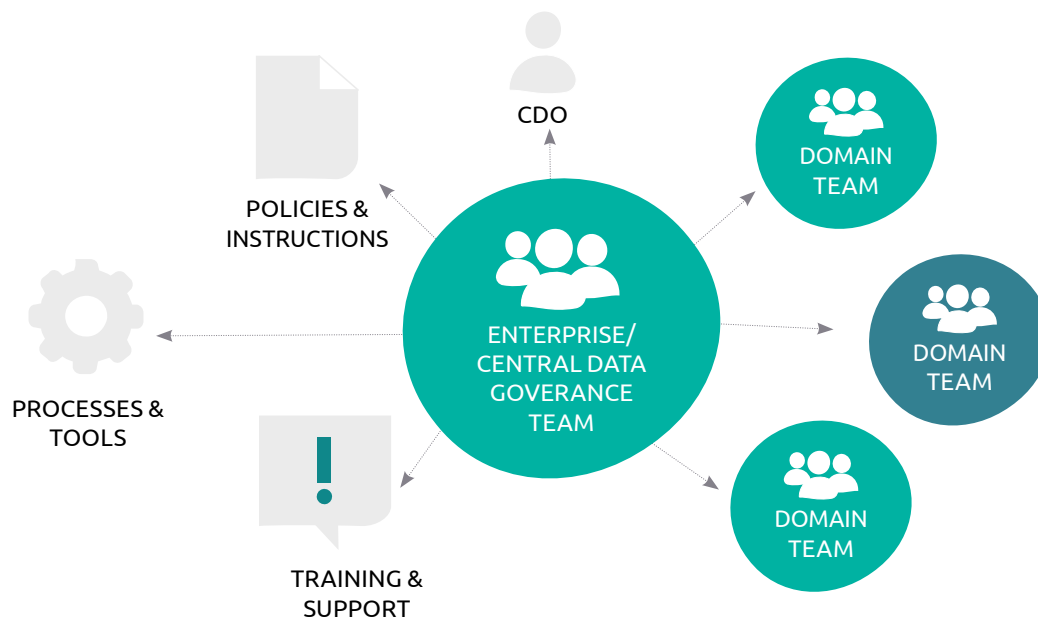
Data product thinking refers to the approach of designing, developing, and delivering products that are built on data and capable of generating insights, predictions, or recommendations that can help solve business problems. By using data product thinking, companies can create products that are more efficient, effective, and personalized, leading to improved customer experience and increased revenue.

How is data governance related to the data product approach?

Data product thinking and data governance are closely related concepts, as both play critical roles in the development and deployment of data-driven products. To be successful, the data product approach requires access to high-quality data and effective data management practices.

Data governance refers to the policies, processes, and standards used to manage the company's data assets. It involves setting up rules for data access, quality, security, and privacy to ensure that data is

correct, reliable, and protected. The data product and domain-driven approach is best supported by a federated data governance operating model, where deep data ability and understanding can be found in the domains creating and maintaining the data. But a common central team is also there to set the policies, instructions, tooling, and more to ensure interoperability of data and aligned ways of working across the domain teams.



## A FEDERATED DATA GOVERNANCE OPERATING MODEL

### How can data product thinking boost data governance implementation?

Many organizations that have already embarked on the data governance journey have realized that implementing governance is a multi-year journey where management and business buy-in is often not easy or straightforward. Business might not yet recognize the necessity of taking ownership of the data as it is often seen as IT's responsibility. The data governance implementation might not yet have permanent funding but only a project budget. While collaboration between business and IT is clearly needed, securing business ownership and continuous funding are the keys to really using the possibilities data has to offer.

Data product thinking can support or even accelerate the data governance implementation journey as it offers a way to create a roadmap where it is

*“Data masters use a hub-and-spoke model for data organization and create dedicated data roles at the leadership level. The operating model for data governance should evolve from a centralized model to a hub-and-spoke model where a chief data officer (CDO)-led central unit handles policy-making and governance, MDM, and data quality, while the operationalization of initiatives is managed by the business units themselves.” Capgemini’s data mastery research*

possible to start small and quickly display progress and business benefits. This will strengthen the business buy-in and commitment to the change journey. Implementation of the domain-specific data governance roles and responsibilities can be done in phases starting from a single or only a few domains with the prioritized data products. This way, instead of an all-at-once approach which can be too heavy to steer and control, the scope is limited for rapid value realization and business buy-in.

### Three important steps to get going

A successful start in a data product and data governance journey requires the following.

1. **The right data governance operating model.** We recommend a federated data governance operating model, as the best understanding of the data is always found at its source. [Capgemini Data Mastery research](#) results show that

#DATAASAPRODUCT  
#DATAGOVERNANCE  
#DATAMASTERS #DATASTRATEGY  
#BUSINESSVALUE  
#DATATRANSFORMATION  
#DATAFORVALUE

# Innovation takeaways

## DATA GOVERNANCE AS A CRUCIAL SUCCESS FACTOR

Data product thinking requires well-managed and fit-for-purpose data to create business value.

## A FEDERATED MODEL TO SUPPORT A DOMAIN-DRIVEN APPROACH

A data product and domain-specific approach is best supported by a federated data governance operating model.

## WELL-MANAGED DATA FOR AGILITY

Good-quality data creates agility to face changes in the market and in business needs.

mature organizations use a federated operating model, in the report referred to as a hub-and-spoke model, where a chief data officer plays a key role in enabling the central team establishment to support the domain-specific teams' data-governance efforts.

2. **Business buy-in** to secure a successful transformation journey and investment in the form of data and product ownership and continuous funding
3. **A data product and governance roadmap based on your data strategy priorities** and agility in implementation to secure business commitment and to display rapid business benefits and value realization

Effective data governance practices can help organizations find and mitigate data quality issues, ensure data privacy and security, and set up data access policies for employees. Governed data will also create many other advantages: decisions based on correct data, speedy development of new products and services – like data as a product – to respond to varying market needs, increased customer and employee satisfaction, lowered data risks, and development of self-service options. High-quality data also serves as a solid foundation for AI, machine learning, and analytics.

Well-managed data will make your organization agile to unexpected changes, new regulations, and emerging new business needs or opportunities.

# DATA IS THE BUSINESS

Driving  
a collaborative  
data ecosystem



**DINAND TINHOLT**

Vice President Insights & Data,  
Capgemini



**RESHMA BHATT**

CP and Retail Industry Lead,  
Insights & Data, Capgemini



**ROB SAKER**

VP Global Retail & Manufacturing,  
Databricks



To drive business value, it is important to leverage all the data from within your organization as well as from partners outside of it. Such a collaborative data ecosystem is an alignment of business goals, data, and technology, among one or more participants, to collectively create value that is greater than each can create individually. It is both combining and collaborating on that data.

## With a little help from your friends

John Lennon and Paul McCartney met by chance in 1957 when Lennon's band The Quarrymen was performing in Liverpool. McCartney then joined The Quarrymen and, after the band had already changed its name to The Beatles, they were by chance discovered by Brian Epstein, at that time a local record store manager who became the band's manager in 1962.

The way we see data ecosystems is similar: it is sometimes about a chance encounter and then bringing various elements together. We could refer to the well-known Beatles song from 1969 Come Together as the unifying theme of this article but instead let's choose another one, namely With a Little Help from My Friends, which was released in 1967. In the context of this story, a little help comes in the form of a little data. Bringing together data from your friends (customers, suppliers, partners, vendors, whoever) is what we would call "organized serendipity."

Imagine you're a retailer operating in a competitive market needing to stay on top of trends, having to make sure your shelves (whether physical or virtual) are filled and are appealing to your customers. As an example, out-of-stocks remain the single largest problem in retail. The challenge with keeping products stocked involves a complex value chain that must anticipate and respond to dynamic market forces. Extreme weather, local events, and even activity from social influencers can quickly alter the demand for a product. In an optimal world, suppliers, distributors, retailers, and other partners would have visibility to changing dynamics and consumption in real-time, enabling them to optimize their operational decisions on-the-fly. And yet supply chains across retail and consumer goods still operate much as they have for decades, making decisions on data that is days or weeks old. It is this delay between changes in demand and our ability to respond that lead to out-of-stocks.

The main sources for retail data are operations by the retailer, data from their ecosystem, competitive data from syndicated sources, and external environmental data from governments and commercial sources.

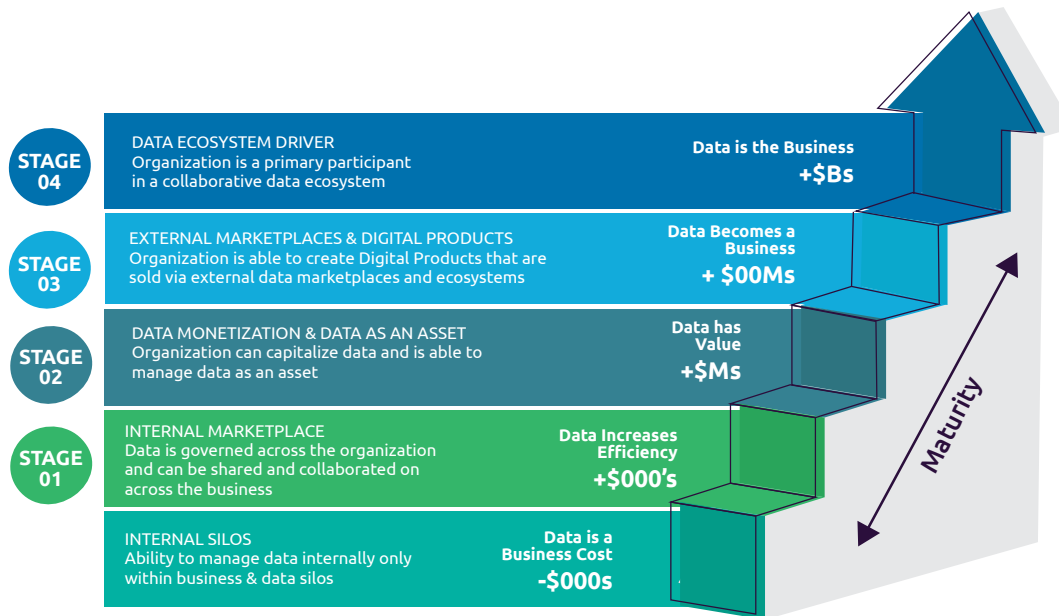
- **Retail operational data** comes as a result of business operations, and includes everything from customer-facing retail sales data, advertising, e-commerce, customer support, reviews, and loyalty to back-of-house data from inventory, distribution, planning, and other management systems.
- **Retailers operate in a complex value chain**, with data coming upstream from suppliers, wholesalers, and distributors, and integrating downstream

with advertising and delivery partners.

- **Competitive data sources** help retailers understand how their key competitors are operating in similar areas. Competitive distribution, assortment, pricing, promotions and advertising, sales, and other sources help retailers index their performance.
- **Environmental data** helps retailers understand the context in which consumers are making decisions. This includes environmental data such as weather, local economic forces, census information, local events and foot traffic data, legal and regulatory changes, social data, keyword searches, and more.

## Finding a cost-effective technology

No two organizations leverage the same data in the same way. The differences in their strategies, operations, competitors, geography, and the systems that support them are designed to help the company succeed. But this means that no two businesses have the same data ecosystem. Companies may exchange data in key areas but increasingly the differences in data between companies is perceived as a competitive advantage. Legacy data-sharing technologies were designed to support the lowest common denominator of collaboration, but have struggled to meet the needs of real-time data sharing, quality, and governance and decisioning. Companies want the flexibility to communicate in real-time with a variety of information and across platforms.



The key to achieving this is to select a cost-effective technology that enables the broadest range of sharing options without proprietary technology or vendor lock-in, facilitates real-time data sharing and collaboration, ensures the control of quality and governance of data, and enables companies to focus on immediately leveraging all types of data to drive better decisions.

## A retail lakehouse simplifies collaboration

A data lakehouse is a modern data-management architecture that combines the features of both data lakes and data warehouses. It is a unified platform for storing, processing, analyzing, and sharing large volumes of data, both structured and unstructured, in its native format, with support for batch and real-time data processing.

Databricks' Lakehouse is built on open-standards and open-source, which avoids proprietary lock-in. This importantly extends to data sharing and collaboration. Databricks

*“Collaborative data ecosystems hold immense potential for retail companies looking to thrive in an increasingly competitive and data-driven industry.”*

introduced Delta Sharing, which is an open-source project started by Databricks that allows companies to share large-scale, real-time data between organizations in a secure and efficient manner.

A Lakehouse is the optimal method for data collaboration as it addresses the critical needs in retail.

- **Real-time collaboration.** Not only can companies share data that is being continuously updated, but Delta Sharing also enables sharing without movement of data.

- **Collaborate on all of your data.** Unlike legacy systems, Delta Sharing enables companies to share images, video, data-science models, structured data, and all other types of data.
- **Centralized data storage.** The Lakehouse architecture makes it easier for different users or groups to access and share data from a single source of truth, eliminating data silos and enabling seamless data sharing across various stakeholders.
- **It supports quality and compliance.** A Lakehouse architecture helps ensure data integrity, traceability, and compliance with regulatory requirements, which are important considerations when sharing data with external users or organizations.
- **It simplifies data management and discovery.** The Lakehouse architecture includes a robust data catalog and metadata management system that helps in documenting and organizing data assets.



#DATAECOSYSTEMS  
#DATA  
#RETAIL  
#COLLABORATION  
#SHARING

# Innovation takeaways

## EMPOWERING COLLABORATION

By leveraging data from within and outside their organization, businesses can create collective value that surpasses individual capabilities, fostering collaboration and innovation.

## BRIDGING THE GAP

Outdated supply chains hinder retailers from effectively responding to dynamic market forces, making real-time data sharing imperative for optimizing operational decisions and reducing out-of-stock issues.

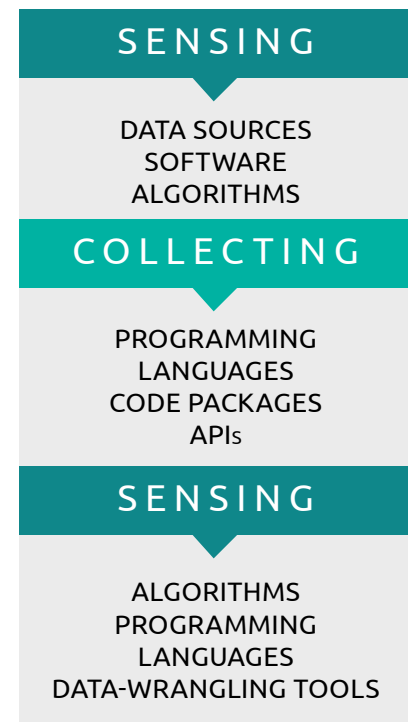
## LAKEHOUSE ARCHITECTURE

A modern data-management approach, the Lakehouse architecture combines data lakes and data warehouses, enabling real-time collaboration, centralized storage, and simplified data management for improved decision-making.

## DELTA SHARING

Delta Sharing, an open-source project, empowers companies to securely share large-scale, real-time data without data movement, unlocking the potential for seamless collaboration, compliance, and valuable insights in the retail industry. enabling real-time collaboration, centralized storage, and simplified data management for improved decision-making.

## COMPONENTS OF A DATA ECOSYSTEM



With Delta Sharing, companies can securely share data with other organizations without having to copy or move data across different systems. Delta Sharing uses a federated model, which means that data remains in the original location and is accessed remotely by the recipient organization. This approach allows organizations to maintain control over their data while still sharing it with others.

Collaborative data ecosystems hold immense potential for retail companies looking to thrive in an increasingly competitive and data-driven industry. By leveraging these ecosystems, retailers can optimize their supply chain, gain valuable customer insights, make informed decisions, foster collaboration, and ensure data security and compliance. As more organizations recognize the value of such ecosystems, we can expect the retail industry to become even more connected, efficient, and customer-centric.

# BECAUSE I WAS INVERTED

Getting data on  
time



HEMAVATHI  
THIRUPPATHI

Senior Enterprise Architect, EA CoE,  
Insights & Data, Capgemini



YASHOWARDHAN  
SOWALE

Architecture head, India domain  
leader for AI, Insights & Data,  
Capgemini



Time is money, and timely insights indeed can make or break a company's success. Enterprises are on a perpetual quest to process data in a jiffy to focus more on extracting golden nuggets of insights from a deluge of data events. But the path to data nirvana is littered with pitfalls. Many companies have stumbled and fallen, thanks to a focus on expensive technology that fails to deliver the desired results in a prudent timespan. Data inversion proposes a radically different way.

Organizations assume digital transformation will solve their traditional, non-digital business problems and drive data-driven decision-making, but many such initiatives result in failure because they lack clarity about overall business objectives and do not ensure tying their business vision and goals to the processes and outcomes.

Data is abundant, but organizations typically wait for it to be conformed and certified. Just like Dr. Edwards Deming once said, "In God we trust; all others must bring data." So why not provide data coming from the core operational systems right when it is created?

The failure to do this is often due to taking a technology-first approach which focuses more on technical processes and pipelines and modelling rather than processing and providing actual data. In such cases, even getting agreement on shared conceptual models becomes a challenge. And in the meantime, nobody gets the data.

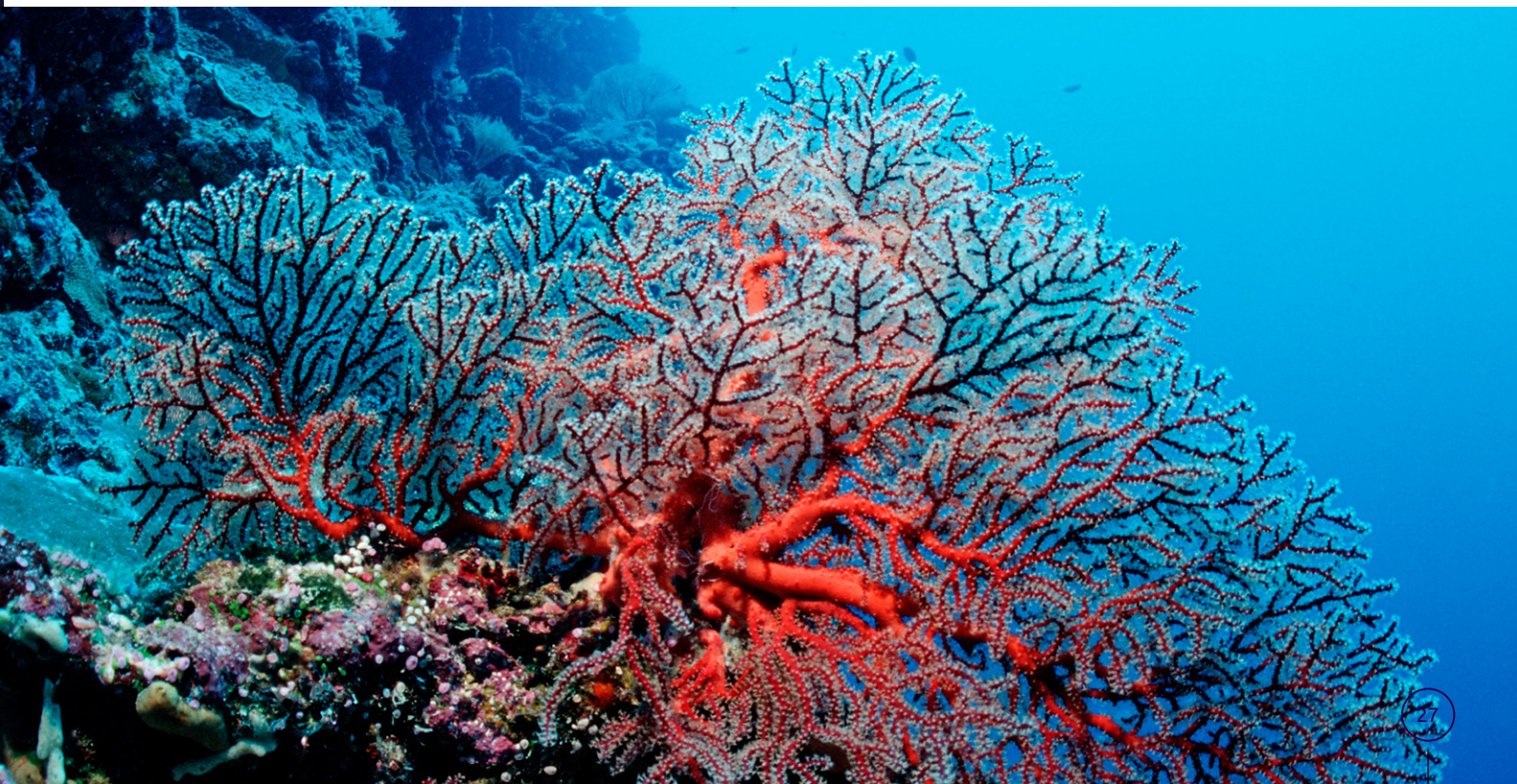
*"Data inversion: Unlocking the power of a data-first approach for timely insights and business success."*

## What is data inversion?

To succeed, the organization must start developing from a data-first perspective instead of a technology-first perspective. Data inversion is the implementation of a data-first approach.

It is a principle which lays emphasis on providing the following.

- Enabling operational speed reporting and insights. Some organizations also employ edge analytics or "listening AI" to drive an outcome.
- Enabling publishing of fundamental data products:
  - o Enterprise-wide master data products
  - o Transactional data products
- History of change/change subscription for point in time interactions. For example, using an SLT stream on SAP to see things changing temporally and actively push insights into processes to drive an iterative and incremental approach. Having a real-time change-data capture approach to shifting data between different layers, and particularly between the core operational systems and transformation platforms is a leading practice required for data inversion.
- Faster reconciliations. Merge standard operational reporting with action-oriented reconciliations to create a unified operational approach which



combines both an operational dashboard (e.g., an audit balance control reconciliation dashboard) with the ability to directly query data pipelines.

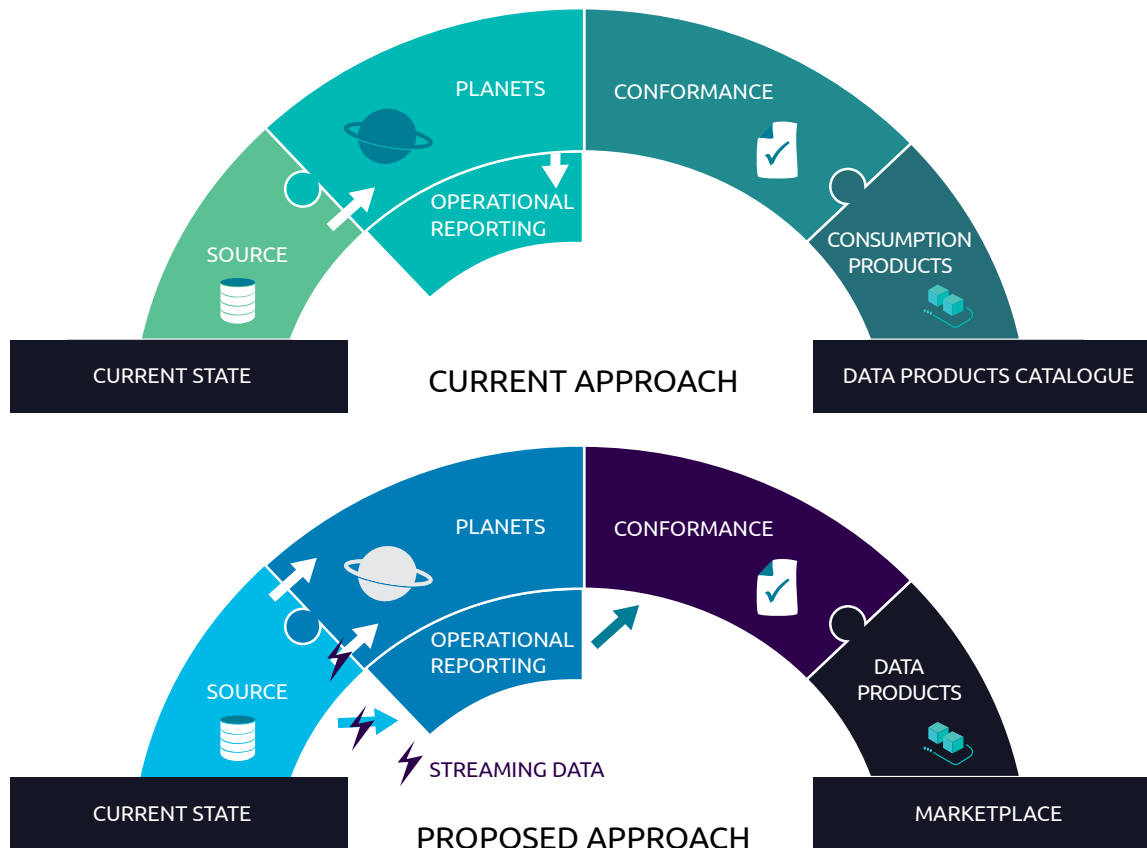
## Why data inversion is critical

Businesses need to undertake initiatives to get to a holistic view of their business processes, key data subjects, business data quality, and governance based on a data

inversion approach. It addresses a few issues which commonly occur in technology-first approaches.

1. Avoids excessive ETL processes which may not be required based on operational insights
2. Avoids silos, as we can harvest same core operational data in one place
3. Speeds up data monetization: enhance value of data as an asset and unlock data's value
4. Empowers data management: unconventional data management (data-product based) provides performance and flexibility
5. Enhances scalability: faster processing right at operational data hub layer provides business agility

## JOURNEY TO DATA ECOSYSTEM DESTINATION BASED ON DATA INVERSION



## DATA INVERSION TO GET TO VALUE QUICKER

## Business benefits

1. **Deliver operational data at speed and scale:** To give an example, a [matcha tea product was launched by HUL](#) in less than nine months based on the People Data Centre which monitors social-media trends in real time and was able to identify a product that was increasing popular among consumers who were interested in healthy products.
2. **Trusted data availability:** Business should be able to trust they are building digital data products on operational data from vehicles. An example is Lyft which analysed the GPS coordinates of its rides and found that 90 percent

overlapped with other rides from nearby locations. This operational insight led to the creation of “Lyft Line” – a pairing service that allows passengers to share a car and receive discounts of up to 50 percent.

## How to adopt data inversion

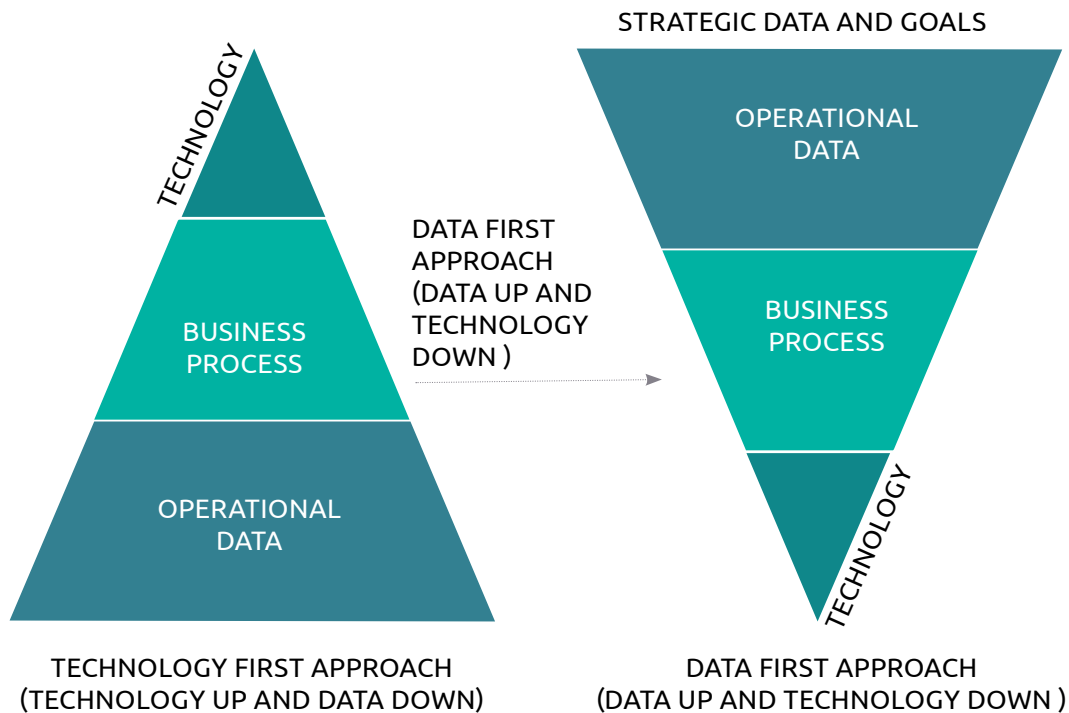
1. **Define a vision and strategy**  
The organization strategy that is supported by objective data, often known as a forward-looking data-driven strategy, is where it all begins. A data strategy which speeds up harvesting of data to become a critical business asset for strategic and operational decision-making is a must.

### 2. Build a data-powered culture

A data-driven culture is essential for implementing a data-first approach. The organization needs to adopt a new way of thinking. A culture of data-driven decision-making must be established and propagated.

### 3. Apply data-mesh architecture

The architectural design pattern known as “data mesh” enables companies to scale quickly, deal with continuously altering data sources, and get quick insights. The ownership of the data has changed from a single, centralized data architecture to a network of domain teams that manage it as a product.



## DATA FIRST FOR A PARADIGM SHIFT

# DATA  
# DATAINVERSION  
# INVERTEDDATA  
# DATAMESH  
# DATAINSIGHT  
# CAPGEMINI

# Innovation takeaways

## A DATA-FIRST APPROACH STARTS AT THE TOP

In today's data-powered landscape, a data-first approach is the key to unlocking untold treasures of insights and opportunities. But, as with all things worthwhile, the change must start from the top.

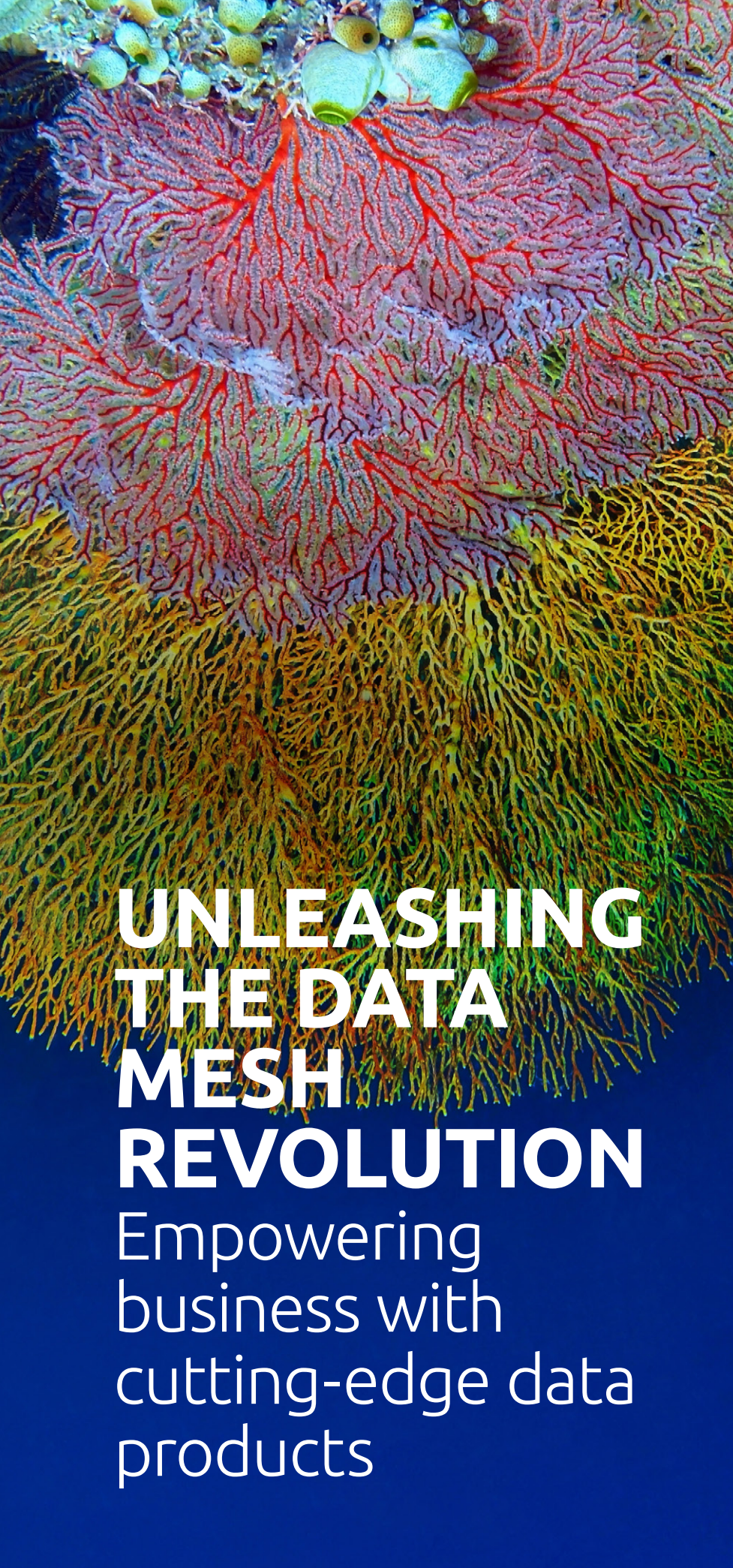
## DATA MESH IMPLEMENTS DATA INVERSION

Data mesh provides a powerful architectural approach to transform data into a valuable resource. It has taken the world of data by storm, playing a vital role in enabling a data-first culture.

## DATA INVERSION IS CRUCIAL

With a focus on data inversion and data quality, AI and analytics efforts will yield accurate and meaningful results. But that's not all: being data-driven can also minimize regulatory and compliance issues, which ensures staying on the right side of the law. And let's not forget about the impact on sustainability: by using data more effectively, the carbon footprint can be reduced.

Data inversion offers a paradigm shift in harnessing the power of data for timely insights and business success. By prioritizing a data-first approach over a technology-centric mindset, organizations can drive data-powered decision-making. It emphasizes early access to core operational data, enabling real-time decision metrics and operational speed reporting. Data inversion also enables the publishing of fundamental data products, facilitates faster reconciliations, and eliminates excessive ETL processes. Embracing data inversion delivers business benefits such as delivering operational data at speed and scale, ensuring trusted data availability, and empowering data management. To adopt data inversion, organizations must define a data-driven vision, cultivate a data-driven culture, and consider architectural patterns like data mesh.



# UNLEASHING THE DATA MESH REVOLUTION

Empowering  
business with  
cutting-edge data  
products



DAN O'RIORDAN

VP AI & Data Engineering,  
Capgemini

---



ANDY MOTT

Partner Solution Architect,  
Starburst

---



The principles of data mesh have moved beyond being just theoretical concepts for data architects and forward-thinking executives. It's time to start delivering on data mesh's promise of exceptional data products. Data mesh principles can help us uncover the valuable insights that businesses need.

## Feedback Fusion: The Power of Continuous Iteration for Product Success

When building a product, it's crucial to understand the utility of the product and how any changes to the product will impact its utility over time.

If we consider building a mobile phone or any other product, the cost of building a phone that is unusable will be significant. Therefore, conducting thorough research in the beginning to understand what the market wants is critical before beginning the product-development process.

Once we have built and distributed a phone we need to continually consider feedback from different channels, including social media and online reviewers, to continuously iterate and improve the phone.

This feedback loop is also imperative for data, however in the past data developers have typically waited for feedback from data consumers and then reacted. This has introduced time delays and ultimately frustration for data consumers.

With product thinking this approach is turned on its head, data product developers are continuously

monitoring both quantitative and qualitative feedback from consumers.

This feedback allows data product teams to proactively evolve the data product to ensure that as data consumers need new capabilities they are being built into the data product, thus avoiding delays and frustration and enabling better outcomes for the organization.

## Data mesh dilemma: Embracing innovation amidst fear and uncertainty

Data mesh principles, which focus on the notion of first-class data products and other factors, have gained an unprecedented amount of interest in the past eighteen months. The conversation in the data mesh community has largely focused on the principles data mesh and what they mean for each organization. Most organizations have invested heavily in cloud but are still struggling to keep up to the pace that the business requires. "Why does it take me three to six months to get a new or modified dataset? Who's responsible for the data governance? How can I trust that the dataset can be trusted?" and the list of questions goes on.

What we discovered during these conversations with clients is there is

an overall acceptance that data mesh and its principles make good sense, but there is the fear factor on the pain an organization needs to go through to get to the promised land of a truly federated data estate of quality, secured, discoverable data products. So, most organizations have kicked the can down the road.

## Start small, think big, and design for industrialization

Here are useful guidelines to help reduce this fear of failure.

1. To effectively build data products, it's crucial to identify the problem you're trying to solve and determine why a data product is the appropriate solution from the beginning of the process. Taking the time to clarify the reasons behind your approach will ultimately save you a great deal of time, money, and effort. This fundamental step is applicable to any product-development process, and it's no different when building data products.

A simple data product canvas together with the business and domain experts need to be committed to this phase. Note: Forget about all technology during this phase.

PROBLEM	DATA	SOLUTION	KPIS	ACTION
<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What is the problem?</li> <li>Why is it a problem?</li> <li>Whose problem is it?</li> <li>Why, why and why...?</li> </ul>	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>Source</li> <li>Quality</li> <li>Access vs Availability</li> <li>Process/Transformation</li> <li>Outputs</li> <li>Test/Training/Validation</li> </ul>	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>Type (Analytics, ML, AI...)</li> <li>What will be the solution</li> <li>Output expected</li> </ul>	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>How to evaluate the model?</li> <li>Which metrics should be used?</li> <li>How much uncertainty can we handle?</li> <li>AB Testing - How?</li> </ul>	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What is the problem?</li> <li>Why is it a problem?</li> <li>Whose problem is it?</li> <li>Why, why and why...?</li> </ul>
	<b>HYPOTHESIS</b>		<b>ACTORS</b>	
	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What will be tested?</li> <li>What are the expected responses for each of them?</li> <li>What should we do from each answer? What strategy should we follow</li> </ul>		<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What is the problem?</li> <li>Why is it a problem?</li> <li>Whose problem is it?</li> <li>Why, why and why...?</li> </ul>	<ul style="list-style-type: none"> <li><b>PRODUCT VISION</b></li> <li><b>STRATEGY VISION</b></li> <li><b>BUSINESS VISION</b></li> </ul>
<b>PROBLEM</b>	<b>RISKS</b>	<b>PERFORMANCE/IMPACT</b>		
<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What is the size of the problem?</li> <li>Why is the baseline?</li> <li>What is the uplift / savings?</li> </ul>	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What are the risks?</li> <li>What these risks might block?</li> </ul>	<b>Ask the right questions:</b> <ul style="list-style-type: none"> <li>What is the impact? How to measure it?</li> <li>Where can you see this improvement/performance?</li> </ul> <b>Examples:</b> <ul style="list-style-type: none"> <li>Increase the customer base</li> <li>Reduce customer churn</li> <li>saving last revenue</li> </ul>		

Each domain deals with a key area for the correct planning and development of the product, providing a 360 view that goes from problem determination to strategic execution

DATA PRODUCT CANVAS IS DIVIDED INTO 10 BLOCKS SEPARATED INTO 3 DOMAIN AREAS



2. Many organizations have not changed their approach to data management in the last 30 years. It is commonly believed that all data must be centralized into a data warehouse or data lake before it can be analyzed, which is both difficult and costly in terms of human resources and technology. Today decision makers wait for data to be made available before it can be used. This means waiting for data pipelines to be specified and built, however this is typically done in the absence of the complete knowledge of the value of the data to a particular use case. This unnecessarily elongated process is fragile and has a negative impact on an organization's ability to compete using data.

Fortunately, solutions like Starburst/Trino offer intelligent

connectors and a highly optimized federated MPP SQL engine that enables the creation of data products by analysts in the lines of business (domains) with no need for intimate knowledge of the source technology. Lines of business can quickly access data and determine its applicability to a use case without having to rely on central data teams.

If we consider this in the context of cloud-data migrations, solutions like Starburst/Trino enable these data products to be created, managed, and retired while the underlying data platforms are migrated. The system administrators only need to update the connector to ensure uninterrupted service for business users. With Starburst we want to give the data-product teams the option to decide on

what works best for them to deliver the best data product that will satisfy the requirements as outlined by the data product canvas.

3. Finally, to ensure that the quality of data products is maintained over time as business needs change, a continuous monitoring and feedback loop is key. Data-product producers need to understand who, how, and for what purpose their data product is being used, so they can proactively manage the data product. This management requires technology capabilities to provide this insight as well as an agile approach to streamline the pipeline from ideation to production and constantly improve efficiency. We look at this as the building of a factory-like model for data products.

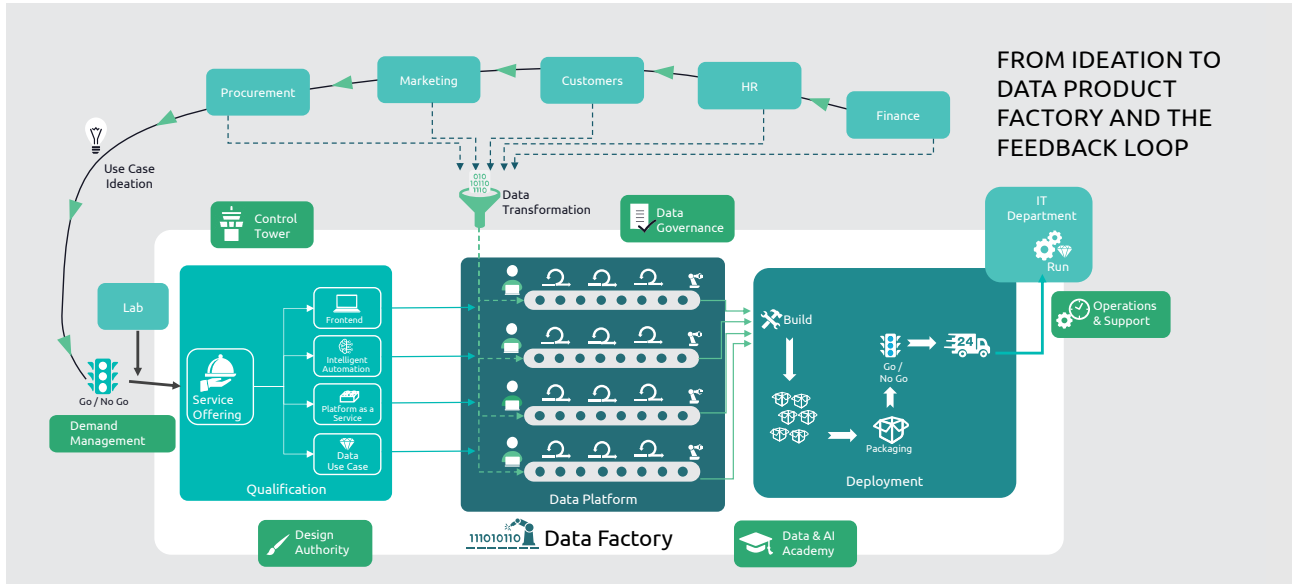


FIGURE: 2

### Data mesh in action

At online fashion retailer Zalando, various lines of business independently utilize Amazon S3 for storing and managing datasets, eliminating the need for a central data team. A central data “enabling

team” oversees data-governance standards and identifies reuse opportunities, while a dedicated platform team supplies compute services including a distributed SQL Engine (Starburst) for analytics. This clear division of responsibilities

– lines of business managing data, the enabling team governing it, and the platform team providing technology – prevents bottlenecks and centralization, fostering agility in leveraging data to maintain a competitive edge.

#DATAVALUECHAIN #DATADIAMONDS  
#DATAMESHPRINCIPLES  
#DATADRIVENSUCCESS #STARBURST  
#FEDERATEDDATAESTATE  
#DATAPRODUCTS #DATA MESH

# Innovation takeaways

## OVERCOMING ADOPTION HURDLES IN A FEDERATED DATA ESTATE

Data mesh principles enhance data-product creation, driving valuable insights and competitiveness, but adoption is slowed by perceived challenges in achieving a federated data estate.

## THE THREE PILLARS OF EFFECTIVE DATA MESH IMPLEMENTATION

Implementing data mesh effectively involves problem identification, utilizing modern data-management solutions, and establishing continuous monitoring and feedback loops.

## DATA MESH IN ACTION

Success stories like Zalando and a large French state organization showcase the benefits of data mesh, including improved efficiency, agility, and competitiveness.

## BRIDGING THE GAP, PRACTICAL STEPS TO DATA MESH SUCCESS

Moving from theory to practice in data-mesh implementation allows organizations to better harness data-product power and succeed in a data-powered world.

A prominent French state organization has been devising its data-estate roadmap for 2025 over the past year. Its current extensive data platform comprises batch processing, streaming processing, AI, and use cases, with concerns about cloud readiness. With a complex data estate plagued by performance and monitoring issues, its goal is to streamline operations using a new data platform based on Starburst and Apache Iceberg. The primary objective is simplification and reduced complexity, achieved by focusing on business outcomes and scaling with data-mesh principles.

*“Start small, think big  
and design for  
industrialization.”*

## Dawn of a new era

The rise of data mesh and its principles plus the technical offerings from Starburst marks the dawn of a new era for data products. As businesses embrace the principles of data mesh, it's essential to address the fear factor associated with adopting this approach. By following the guidelines outlined in this article – focusing on identifying the problem to be solved, leveraging modern solutions like Starburst/Trino for data management, and implementing continuous monitoring and feedback loops – organizations can confidently embark on their journey towards a truly federated data estate.

Success stories like Zalando and the large French state organization demonstrate the transformative power of data mesh in improving efficiency, agility, and competitiveness. As we move forward, it's crucial for businesses to embrace the promise of data mesh, shifting from theoretical discussions to real-world implementation. Only then will they be able to harness the full potential of exceptional data products and uncover the valuable insights needed for sustained success in an increasingly data-powered world.

# THE RACE TO PRODUCTIZE DATA

The slow and  
steady tortoise  
won't win



ARJUN SUD

Principal, Insights & Data,  
Capgemini



Most organizations are still in the early stages of the productization of data. Even though consensus exists for the need to productize data (especially with the concepts of data mesh gaining interest), guidelines to do so are often scant or too basic for a meaningful blueprint. Hence, organizations regularly gravitate to a wait-and-watch approach, underestimating their readiness for a bold transformation.

In this article, we list the guiding principles and the operating model for rapid productization of data, to aid and accelerate a data organization's existing efforts to achieve full productization of data. We start with key guiding principles, followed by a series of characteristics of a suitable operating model.

## Guiding principles

1. **Providing tangible value:** Data products must address a known issue, opportunity, or be necessary to run a business function. Its value is indisputable and is validated by users. The traditional role of the data organizations to just make data available through migration and

ingestion is now obsolete. Data-runs-the-business or data-is-the-business are the new mantras. Also, the value of the data product increases by an order of magnitude as the product evolves from just providing information to providing insights and unified entity views.



2. **Discoverable, available, and interoperable:** Data products should be catalogued effectively for easy discovery through basic keyword-based search. They should be programmatically available, and utilize open standards, formats, and APIs, allowing for interoperability.
3. **Trusted:** Establishing trust in the data product is obvious, but the level of trust needed varies across products and consumers. For example, a downstream consumer generating invoices requires 100 percent accuracy, but an analyst examining long-term trends may have some tolerance to volatility in data

quality. It's best to publish a data-quality health scorecard and let consumers decide if and how to consume the data products.

4. **Monetized:** The productization of data enables its monetization. Even if the enterprise chooses not to monetize externally, the opportunity to leverage "showbacks" or chargebacks for internal consumption should be considered. Leaders who dismiss the idea of internal monetization as an unnecessary accounting exercise miss the opportunity to evolve the data organization from a cost to a value center. Instead of constantly requesting

budgets to support the business programs, a data organization must provide products that are priced for the value they deliver.

## Operating model

Here are the characteristics of a suitable operating model, divided over the dimensions of technology, process, and people.

### Operating model: Technology

1. **Data mesh:** Centralization of the data organization slows development and inhibits creativity. A data mesh architecture, logically defined by domains, allows business

stakeholders to define their own products, and make them available for consumption through a shared layer. Rather than constantly striving for consensus amongst disparate brands and functions with heterogenous needs in a complex enterprise, it is usually preferred to allow each to run at their own pace and foster healthy competition.

2. **Semantic views:** Business users should be empowered to create consumer-friendly views to underlying disparate and complex data sets. Data virtualization allows for the rapid realization of tremendous value stored within the enterprise data. Convert that potential energy to kinetic energy for business value.
3. **Feature stores:** As AI scales, the evolving and increasing needs of the data scientists must be understood. Feature stores are a great way to provide data for model training and production deployment.
4. **Third-party data:** Third parties can enrich the available enterprise data to enable a new set of products. The data platform must allow for the easy discovery, hypothesis testing, and eventual onboarding of third-party data.
5. **Marketplace:** A marketplace for the discovery, exchange, consumption, and subscription of data product is necessary. This should have the same UI/UX parameters that we expect from any compelling e-commerce site today.
6. **Analytical sandbox:** Allow users to Bring-Your-Own-Data to enable the creation of new products and insights. An analytics sandbox should provide

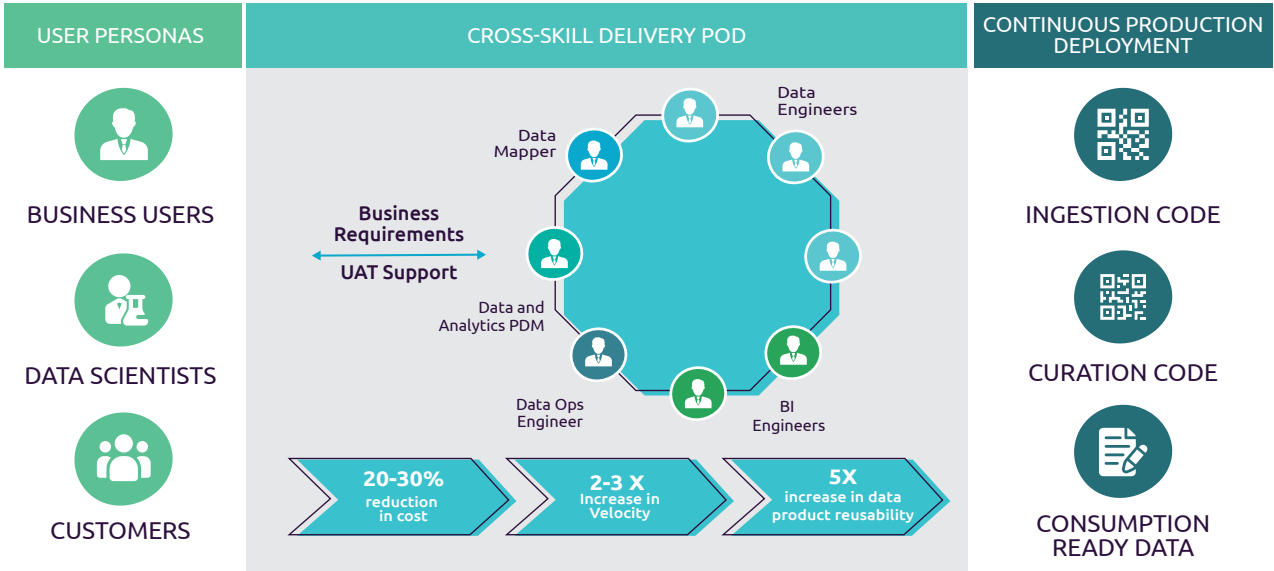
the data, tools, and ML models for users to innovate in a safe space.

#### Operating model: Process

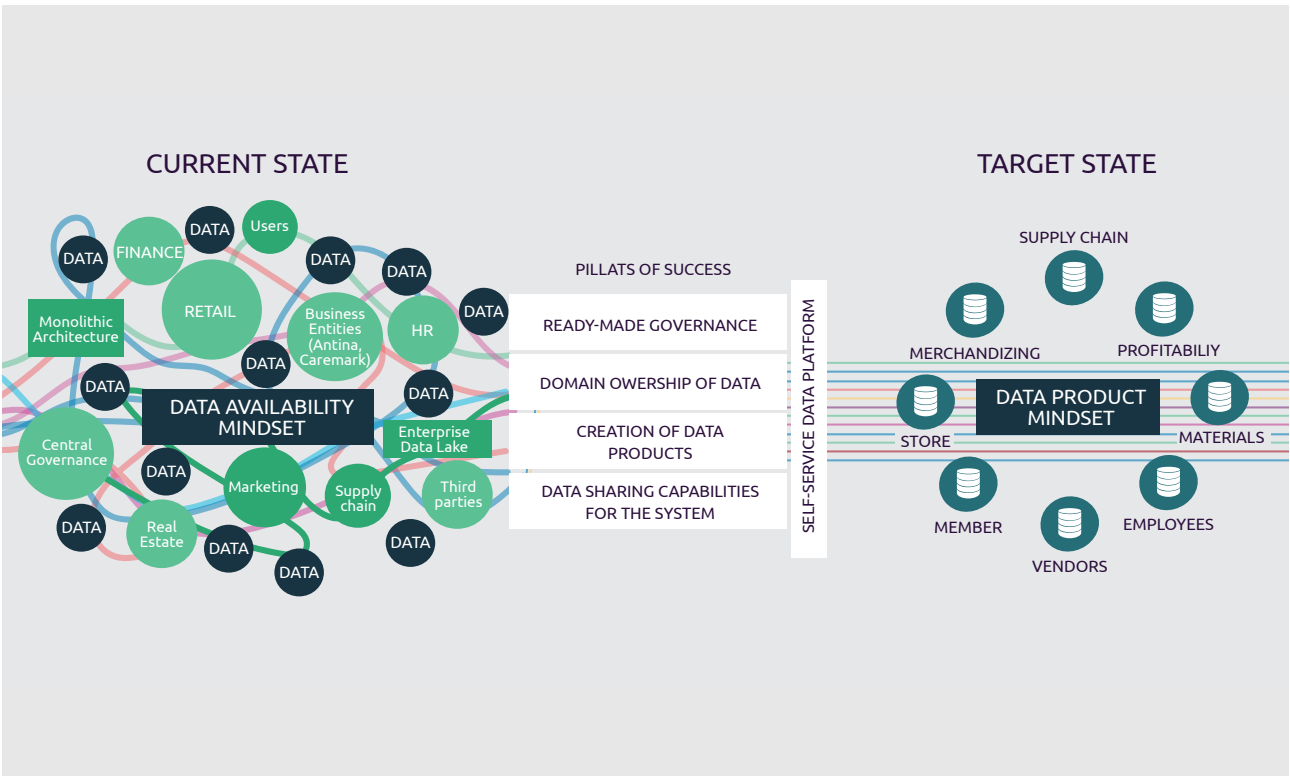
1. **Product onboarding:** ROI based on an effort-vs-reward selection method should be used to create a backlog for the data teams. No engineering or analytical effort must commence unless this analysis is understood and agreed upon by the stakeholders. Upon agreement, the data teams must have a streamlined and automated delivery mechanism to make the product available for use.
2. **Accounting:** Based on the ROI analysis for product onboarding, the value of the data product should be tracked and reported to the business stakeholders on a continuous basis. The data organization must be compensated in proportion to the value delivered. The effort portion of the analysis determines the cost of delivering and maintaining the product. All accounting components are now available to manage the data organization as a P&L-based value center.
3. **User feedback:** The usage of the product and customer satisfaction with it must be continually measured. A product roadmap is refined based on this feedback and, when necessary, products must be decommissioned if not meeting user needs.
4. **Marketing:** Data organizations must not shy away from marketing the wonderful work they do. Product roadmap awareness sessions help obtain early feedback and alignment with business users.

#### Operating model: People

1. **Product managers:** This is the most important, but often misunderstood, aspect of the operating model. The product data managers (PDMs) set the tone and pace for the productization of data. Often organizations graduate traditional business analysts to PDMs, without investing in the necessary training and setting a department vision. Hence, instead of developing and owning product roadmaps, the new PDMs continue to collate technical and business "requirements." Also, sometimes when domain experts are considered for this role they do provide good expertise on product roadmaps but often lack the understanding of how data is ingested, curated, and consumed. Hence, they struggle to develop a pragmatic and ROI-positive roadmap for the portfolio of products. The right talent strategy for PDMs involves domain experts that have some background (or at least training) in fundamental data concepts.
2. **User personas:** User persona capture and its continuous refinement helps maintain product alignment with user expectations. Third-party data on demographical segmentation can also be leveraged.
3. **Delivery PODs:** The delivery, maintenance, and support of the product(s) must be managed by a fully empowered cross-functional "two-pizza team" that has minimal dependencies. Product development, support, and roadmap must be managed by the same team.



The productization of data is within reach of most data organization. The right blueprint and product mindset can elevate the data organization from a cost to a value center that continuously delivers tangible benefits to the business. In a fast-paced data landscape that constantly evolves, the slow-and-steady tortoise will always be outmatched by more nimble rivals.



# DATAPRODUCTS  
# DATA MODERNIZATION  
# DATAENGINEERING  
# ANALYTICS  
# AI

# Innovation takeaways

## MONETIZATION

External or internal monetization of data products enables the evolution of the data organization from a cost to a value center.

## DATA AND ANALYTICS PRODUCT MANAGER

Brings not only deep-domain expertise, but also reasonable understanding of fundamental data concepts.

## PRODUCT ROADMAP

Start managing to a data-product roadmap, versus the traditional approach to gathering and delivering business "requirements."

*"The data organization must be compensated in proportion to the value delivered. All accounting components are now available to manage the data organization as a P&L based value center."*



# DATA FOR AUGMENTATION

## GENERATIVE AI FOR THE ENTERPRISE: 41

Beyond the hype

*Mark Oost, Marijn Markus,  
Dr. Sergey Patsko, Capgemini*

## THE GENERATIVE AI ARMS RACE: 45

Why responsible AI matters now more than ever

*Lee Hickin, Microsoft &  
Aruna Pattam, Capgemini*

## BIGGER, SMALLER, SMARTER, SAFER: 48

The competing pressures disrupting AI

*Mark Roberts & WeiWei Feng, Capgemini*

## DISRUPTION MANAGED: 52

Bringing "gen AI-in-the-room"

*Christopher Scheefer, Capgemini*

## I SEE YOU: 56

Opening the AI black box

*James Wilson, Pantelis Hadjipantelis, Capgemini*

## GENERATIVE AI: 60

A powerful tool, with security risks

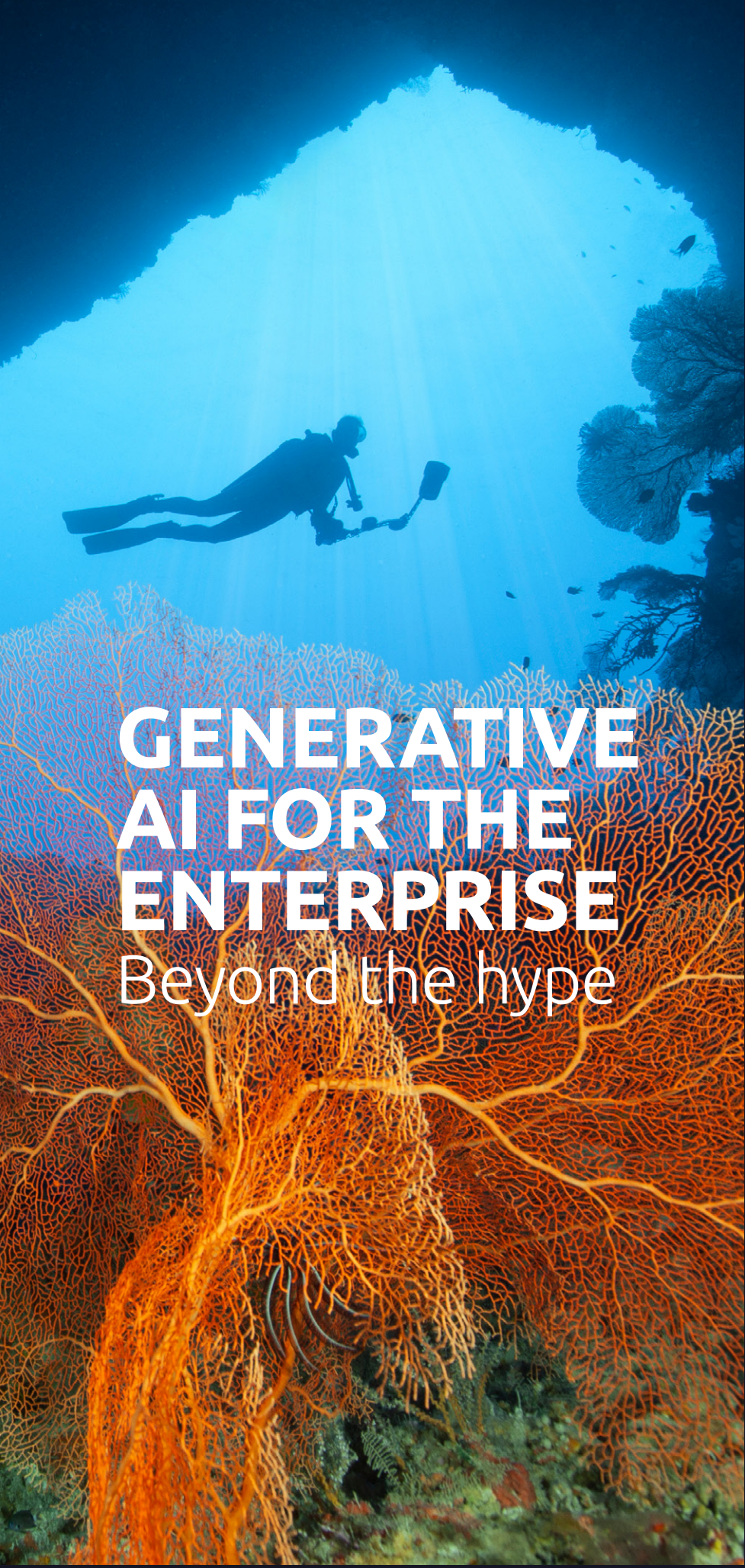
*Matthew O'Connor, Google Cloud*

## SERENDIPITY SYSTEMS: 63

Design for AI

*Charles Aubert &  
Chloe Cheau, Capgemini*





# GENERATIVE AI FOR THE ENTERPRISE

Beyond the hype



**MARK OOST**

Global Offer Leader AI, Analytics & Data science, Capgemini



**MARIJN MARKUS**

AI Lead, Managing Data Scientist, Capgemini



**DR. SERGEY PATSKO**

Data & AI Deputy Group Offer Leader, Capgemini



Attention enterprises: to stay ahead, one must be quick to adopt generative AI. There's no room for half-measures. Learn and adopt, or risk becoming a dinosaur.

## Progress is inevitable

Society is still reeling from ChatGPT's impact. Students are celebrating, teachers are protesting, and governments are struggling to pass legislation. How should your enterprise respond?

Previous tech breakthroughs like the typewriter, the calculator, the smartphone, and even Microsoft Word all had one thing in common: they drastically changed the way we work.

Did you know that in the 1970s, teachers tried to ban calculators in

class? They argued children would rely on them too much, without understanding the math underneath. Today, similar concerns are raised about the impact of social media and smartphones on health and development. And yet, society adopted the technology anyway, finding ways to address concerns.

As history shows, technological progress cannot be stopped. Only regulated.

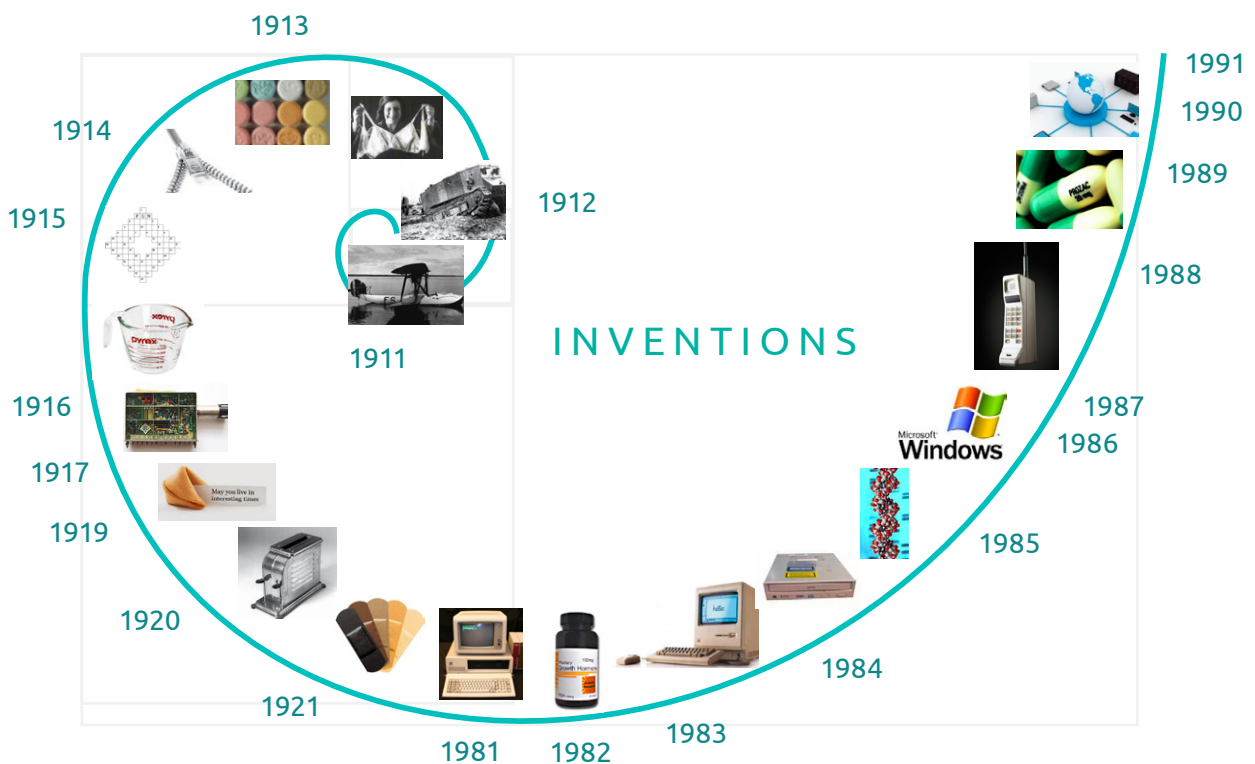
It is up to us to figure out how to best adapt to it.

Radio, telephones, and smartphones changed the way we

communicate, and search engines like Google changed the way we seek information. Generative AI is now changing the way we generate or retrieve text, music, and visuals in any combination.

Of course, this raises many concerns about ethics, data privacy, and intellectual property.

But to address these concerns, we must first properly understand the technology of generative AI: What is under the hood?



## TECHNOLOGICAL ADVANCEMENT THROUGH THE YEARS

## The technology of ChatGPT

ChatGPT is a Large Language Model (LLM), trained on massive amounts of text language data. Its goal is surprisingly simple: to predict the most likely next word in a sentence, based on the context of previous words.

The model is trained on books, articles, and online content from sources like Wikipedia and Reddit, learning to recognize patterns and relationships between words. This knowledge is then used to calculate the probability of each possible next word in a sentence. For example:

Given the input “I love to eat” the model might predict the next word is “pizza” rather than “shoes.” This makes the technology a perfect fit for generic pizza marketing but not necessarily for niche edible-shoes salespeople.

Accuracy depends largely on the quality and diversity of training data. A model trained only on medical journals will struggle to predict words about sports or music, but these improve as their data, methods, and computational power improves.

But the model is guessing words, in the end. These are very educated guesses based on massive amounts of data but it does get facts wrong, or “hallucinates” falsehoods, and crucially it cannot discern between truths and lies.

This is the core weakness of LLM technology, and it is up to humans

*“Kids are already adopting generative AI into their daily lives. Will your organization adopt as well? Or go the way of the dinosaurs?”*

to decide how to deal with it as we use it.

We’ve all heard the doomsday scenarios: AI “replacing” humans. Yet no technological revolution so far has replaced us. Technologies have enhanced our abilities, like the computer and calculator, enabling us to focus on highly intellectual tasks like reasoning and strategic planning, critical thinking, decision-making, and soft skills.

Routine work is more likely to be left to generative AI, like searching, typing, and spelling of the past. White-collar jobs are more likely to be affected than blue-collar jobs, as LLMs are great at automating and replicating implicit knowledge gained through experience and repetition. Because of this, fields like legal and accounting are likely to be affected in a big way. Most fields that require several years of on-the-job training will as well.

And while individuals will be upskilling themselves to adapt, enterprises must ready themselves to harness this technology and empower employees with it.

## Enterprise enabled

To prepare for a generative AI future, enterprises should focus on these six key steps.

1. **Identify opportunities.** Identify areas of benefit within the organization, where gen AI can enhance productivity or generate new insights. For example, corporate information retrieval for R&D, or for marketing content generation.
2. **Leverage proprietary data.** The goal shouldn’t be to adapt the same AI technology as everyone else, but to create custom generative AI that your enterprise alone can leverage.  
  
Focus on cases where enterprise-specific proprietary data can give you a competitive edge.
3. **Invest in Infrastructure.** Invest in the necessary IT infrastructure to scale your own generative AI models, data platforms, and generative AI portals, and upgrade hardware and utilize new cloud services. Focus on the leaders Microsoft, OpenAI, Google, AWS, and Adobe, as well as challengers like Cohere, Anthropic, Stability AI, and Hugging Face.
4. **Invest in people.** Create skilled teams with in-house experience in Large Language Models, not just to build new LLM models from scratch (a costly endeavor), but to leverage data and models available on the market, and to combine them with proprietary data to create unique solutions.

#CHATGPT  
#LLM  
#AI  
#GENERATIVEAI  
#OPENAI  
#MICROSOFT

# Innovation takeaways

## EMBRACE PROGRESS

History shows technological progress cannot be stopped. It is up to us to adapt or fall behind.

## NOT YET INTELLIGENT

Generative AI is not “intelligent” like humans. It can only reproduce data it was trained on. It cannot reason, only guess. This risk can be minimized, but never removed outright.

## PURSUE UNIQUE IMPLEMENTATIONS

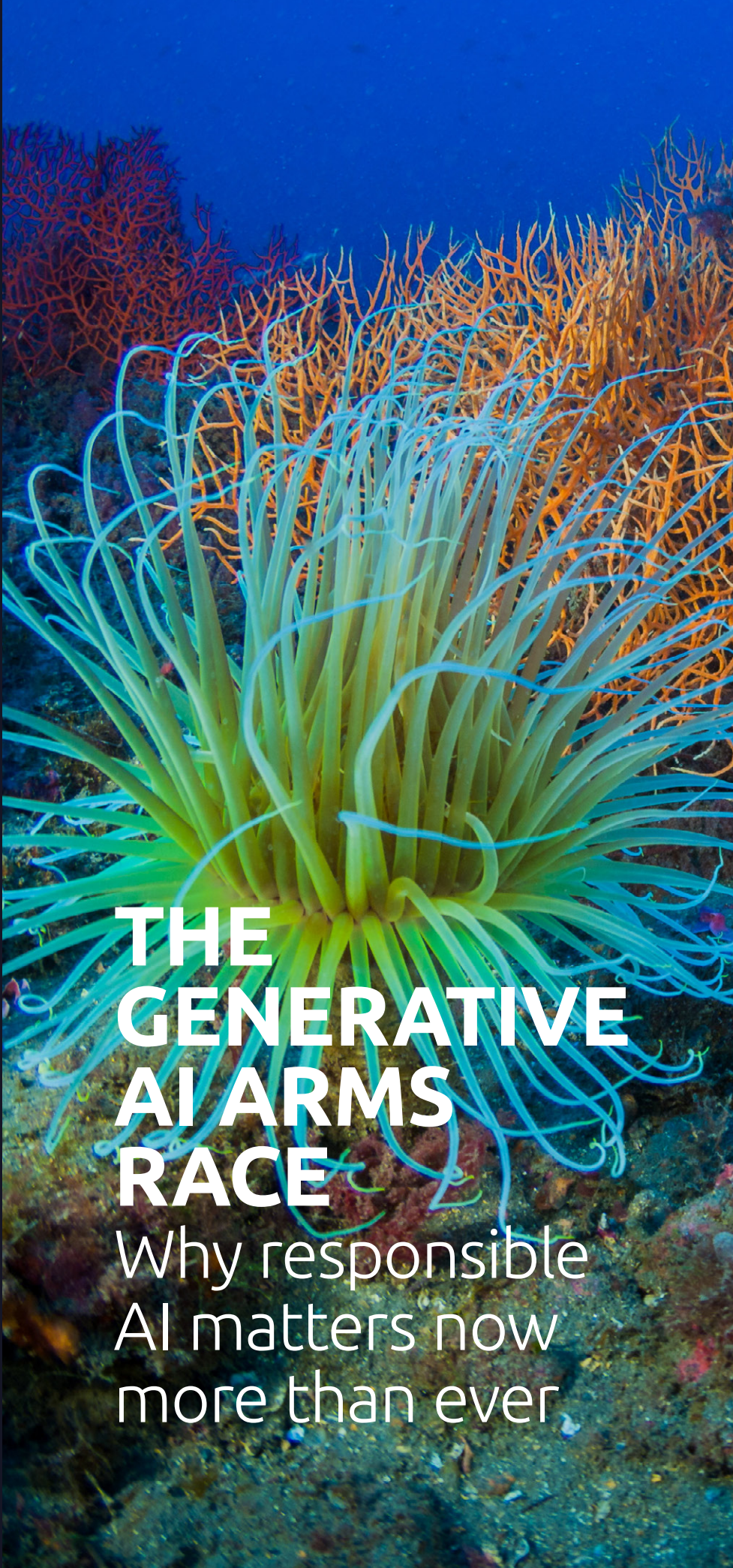
To stay ahead, enterprises must be ready to adapt. Build platforms to leverage proprietary data and gain custom, differentiating generative AI models, as enterprises who won't adapt will surely be left behind.

5. **Partner up.** Partner with industry leaders to create collaborative data ecosystems to train or fine-tune custom generative AI models for the use cases of your choice, as technology will inevitably progress faster than the knowledge of any single organization.
6. **Implement safeguards.** To ensure the reliability of your models, it is crucial to continuously assess and enhance them. Establish guardrails to monitor and minimize unreliable (wrongly guessed) output of your models, as well as to address ethical, privacy, and regulatory concerns.

Have a look at [Capgemini's custom generative AI](#) services for more information.

Of course, the biggest threat of all in an ever-changing world is inaction. A failure to address these six points brings a significant risk of missing out on the future. Generative AI doesn't just challenge enterprises to adopt new technology, it also challenges many to upgrade and adapt their old tech stacks and personnel to thrive in this brave new generative AI world.

After all, new technology is rarely the issue. Most often, legacy is.



# THE GENERATIVE AI ARMS RACE

Why responsible  
AI matters now  
more than ever



LEE HICKIN

National Chief Technology Officer  
ANZ, Microsoft



ARUNA PATTAM

Head – AI, Analytics and Data  
Science, Insights & Data, Capgemini



As we race to unlock generative AI's potential, let's prioritize responsible AI practices, ensuring ethical and inclusive solutions that revolutionize industries while safeguarding society's well-being for a better, collective future.

In our fast-paced digital world, businesses are always searching for innovative ways to stay ahead of the competition. And one of the most potent tools at their disposal is artificial intelligence (AI) – specifically, generative AI. As companies strive to harness the

power of generative AI, there is an ever-growing demand for responsible AI practices to ensure ethical and fair use of this technology.

In this article, we'll delve into this Generative AI arms race, and why responsible AI practices are more crucial now than ever. Our goal is to inspire the AI community to recognize the importance of responsible AI practices in their organizations and personal applications.

## The generative AI arms race: A brief overview

Let's dive into the exciting world of generative AI. This ground-breaking technology can create new content that's never been seen before, whether it's text, music, or even visuals. The potential applications of generative AI are limitless, from transforming marketing and entertainment to revolutionizing healthcare and finance. Augmenting human creativity and automating repetitive tasks can lead to huge efficiency gains, cost savings, and new business opportunities. But let's not forget the importance of responsible AI practices.

The generative AI arms race is heating up, with companies and nations striving to develop cutting-edge artificial-intelligence systems. Along with this there are different factions shaping the conversation, from those pushing for rapid development to those warning of existential risks.

- Enthusiasts are optimistic about the benefits, even of achieving artificial general intelligence (AGI), which could lead to revolutionary advancements in various fields.

- On the other hand, AI doomers worry that powerful AI could lead to humanity's extinction.
- Meanwhile, the AI safety alignment camp aims to ensure AI systems obey programmer intentions to prevent power-seeking AI from causing harm.
- Then there are AI ethicists, focusing on responsible development and deployment of AI technology.

With these differing views, we must all recognize that this technology has enormous potential to change the world.

That's why it's essential to strike a balance between pushing the boundaries of what's possible with AI while also ensuring its responsible development and use.

By embracing responsible AI practices, we can achieve a brighter future for everyone.

## Why Responsible AI Matters Now More Than Ever

The democratization of AI, through the likes of ChatGPT, is making powerful tools available to everyone, fueling a fundamental shift in their usage and impact.

With AI transitioning from controlled, reactive applications to proactive, creative solutions, the need for responsible AI has become more crucial than ever to ensure ethical, inclusive, and secure outcomes in this widely accessible landscape.

## What responsible AI looks like, and why it is important

At its core, responsible AI is about recognizing the potential impact of AI on society and taking steps to ensure that it is used in a way that is

*“Embracing responsible AI practices is the key to unlocking the true potential of generative AI, striking a balance between pushing the boundaries of innovation and ensuring its ethical, inclusive, and secure development for a brighter future.”*

safe, fair, and transparent. This includes everything from protecting people's privacy to ensuring that AI systems don't perpetuate harmful biases, mistruths, or outright lies!

One of the key challenges with generative AI is that it can be very difficult to monitor for accuracy and fairness. However, one of the strengths of large language models (LLMs) is that they are designed to model language, not the intent behind it. This means that they don't inherently contain biases, but rather may (if prompted to do so) reflect the biases that exist in the data they are trained on.

It's up to us as humans to ask good questions, prompt with good intentions, and be aware of the potential biases in the data.

If we continue to believe that AI will only lead to a negative dystopian future, it surely will. But if we can elevate our thinking and handling of this technology to be aligned to the amplification and empowering agency it brings to humanity – we may all one day look back on this time as the moment we chose the right path and developed the skills

#GENERATIVEAI  
#RESPONSIBLEAI  
#AIETHICS  
#INNOVATION  
#AIDEMOCRATIZATION  
#AIFUTURE

# Innovation takeaways

## NAVIGATING THE GENERATIVE AI ARMS RACE

As companies and nations compete to develop advanced AI systems, it is essential to acknowledge the differing viewpoints on AI's potential risks and benefits.

## DEMOCRATIZATION AND IMPACT OF AI

As powerful AI tools become more accessible to a broader audience, the need for responsible AI practices intensifies to maintain ethical, inclusive, and secure outcomes in this transformative landscape.

## EMBRACING RESPONSIBLE AI PRACTICES

Ensuring the ethical, fair, and transparent use of generative AI is crucial in maintaining a balance between technological advancements and responsible development, ultimately creating a brighter future for everyone.

and mindset to prompt our AI co-pilots towards helping us. An example comes from Microsoft, where the ethical AI framework helps developers create AI systems that are safe and trustworthy. This framework includes a set of guiding principles, such as ensuring that AI is designed to enhance human capabilities and that it is transparent and accountable.

The age of generative AI brings both opportunities and challenges for responsible AI. On the one hand, these technologies have the potential to transform the way we interact with the world and solve some of society's most pressing problems. On the other hand, they can also be used in ways that are harmful or perpetuate biases – but only if we as the operators of this AI don't take responsibility for our contributions, our inputs, and our prompts.

By recognizing our responsibility to ask good questions, prompt with good intentions, and be aware of the potential biases in the data, we can ensure that AI is used in a responsible way and develop AI systems that are not only powerful, but also ethical and trustworthy.

## Prioritize ethics

The potential of AI is awe-inspiring, but it comes with great responsibility.

Responsible AI practices are no longer an option; they're a must-have for businesses to thrive and for society to progress. Core principles such as sustainability, fairness, and transparency to develop inclusive, eco-friendly AI systems should be embraced.

By prioritizing ethical considerations, promoting diverse teams, and utilizing reliable data, we can harness AI's power to create a better world for all.

Let's make responsible AI the norm and join forces to ensure its positive impact on society and our collective future. Act now and contribute to the responsible AI movement.



# BIGGER, SMALLER, SMARTER, SAFER

The competing pressures disrupting AI



MARK ROBERTS

Principal Technologist, Capgemini Engineering



WEIWEI FENG

AI Nordic Technique Lead, I&D Capgemini



Artificial intelligence is not a new field; in fact, it has been around for a lot longer than most people realize. But it's certainly sitting at a hugely important juncture, with the next decade possibly being one of the most significant in its long history.



Recent impressive achievements in Generative AI such as ChatGPT have opened the world's eyes to the fact that we are in the midst of a technological revolution that will continue to transform our world in profound ways. However, there are competing pressures and rumbling feuds in the field that are exposing new issues, and opportunities.

From watching the media over the last few years, you'd be forgiven for thinking that AI was now a solved problem, as there have been so many stories of impressive AI systems outperforming humans. If that's true though, why are there not robots among us? Where is the super-intelligent AI that will solve all of the world's problems?

The truth is that AI has solved some very narrow problems in very narrow ways. These are extremely impressive feats of engineering, but for the most part they are not "intelligent" at all, and rarely survive first contact with the infinite complexity of the real world. This has led to a lackluster adoption of AI across industries, with these competent but fragile AI solutions being deployed only in narrowly defined low-risk use-cases. Even the seemingly impressive feats of Large Language Models lack the rigor and trustworthiness to be let loose in situations that require unsupervised dependability and accuracy.

## Scale versus symbols

This commercial reality contrasts starkly with the high-profile picture painted in the popular media, which seems to show impressive feats of intelligence, from apparently sentient AI language models to gallery-worthy generative artworks. However, these impressive feats are just the end-result of a decade of rapid progress in the use of artificial neural networks. In recent years, we have mastered the art of creating huge machine learning models which can so deeply and completely identify the patterns present in certain datasets that they can predict their outputs or manipulate the relationships in the data at will to produce endless creative variants from it.

This realization that scale alone can produce intelligent-looking results has been one of the driving factors in AI over the last decade. In fact, this philosophy is behind one of the two major factions emerging within AI, where some people believe that true intelligence will just emerge if only we could make the models big enough (sometimes called "scaling maximalism"). Others, however, believe that these large models are nothing more than glorified curve fitting – a statistical mimicry of intelligence that can't extrapolate beyond their trained experience. This second faction believes that true intelligence can only come from systems that explicitly reason about real-world concepts (so called "symbol manipulation"), developing a true general intellect rather than just copying patterns of behavior from other intelligent beings.

The debate rages on between the two camps, but the reality is of course that both are probably right to some extent, and elements of both approaches will be part of an

ultimate, hybrid solution. The massive deep-learning models and their amazing abilities to comprehend and manipulate complex data will likely work in tandem with higher-level symbolic reasoning to produce some truly astonishing leaps forward.

## Environmental cost

External pressures are also driving the field of AI in a different direction, from the very big, to the very small. The environmental cost of modern AI is hard to comprehend. The energy requirements to train just one of these large models equates to a carbon footprint equivalent to the lifetime emissions of five family cars, and this is forcing the field to question its obsession with ever-increasing scale.

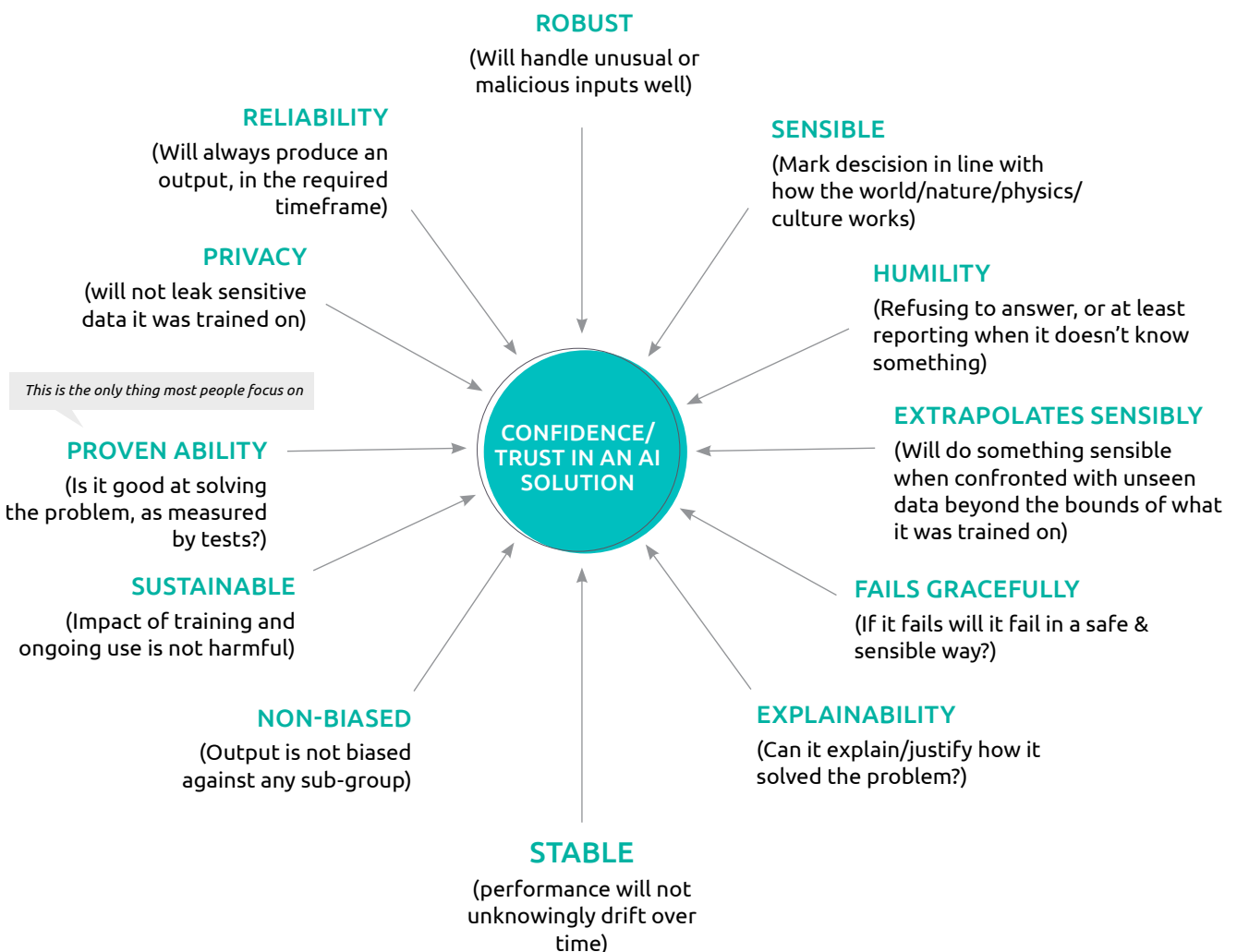
At the same time, consumer trends are driving a demand for AI deployments on smaller edge devices, both in a quest for independence from cloud connectivity and due to well-founded privacy fears. We are entering an "AI Everywhere" era where most devices around us and most digital services will incorporate AI in some way. Being everywhere will likely make us *less* conscious of it, in an ironic twist that we have seen with other technologies. **AI will disappear into ubiquity.** As it becomes less visible, but more deeply intertwined with the fabric of our lives, we must put more effort into the neglected area of AI governance and trust to ensure we understand the implications of such a symbiotic relationship.

## Aiming for success

For many years, the industry has used extremely narrow definitions of what “good” looks like for AI, focusing on narrow measures of performance on unseen benchmark data. This narrow focus has led to

many failed projects, and many others to be doomed to a life of perpetual proof-of-concept. To have true confidence in a deployed AI system requires a much more holistic view of the critical success factors. While many have thankfully recognized the importance of

explainability as a second important factor, this is just one of a dozen facets of accomplishment in AI which all need to be at least considered to ensure a successful outcome.



#AIEVERYWHERE  
#DISAPPEARINTOUBIQUITY  
#SCALEMAXIMALISM

# Innovation takeaways

## THE FUTURE IS HYBRID

Despite the war between the two camps in AI, it is likely that the future will incorporate elements of both large-scale statistical learning and more focused reasoning of explicit knowledge.

## AI EVERYWHERE

As AI pushes out to the edge, it will be everywhere around us, and we need appropriate safeguards in place.

## TRUST IS MULTI-FACETED

Confidently deploying an AI system requires more than just headline accuracy. Many factors need to be considered together to ensure long-term adoption.

One thing is clear though: the improvements in AI's abilities are on an incredibly fast exponential curve, accelerating at a pace that is surprising even experts in the field. AI is both simultaneously over-hyped and under-hyped, with some of the short-term successes blown out of proportion, while the majority of people simply cannot comprehend the magnitude of change that is coming in the medium term. There is every reason to believe that AI will be truly transformative for humanity, society, and the planet, although the path will not be easy or risk-free.

Whatever happens though, it is clear it will come surprisingly fast. Like all exponential curves, the rapid growth phase has a habit of creeping up and then suddenly overtaking. For those who see it coming, there will be massive opportunities.

*"AI is both simultaneously over-hyped and under-hyped, with some of the short-term successes blown out of proportion, while the majority of people simply cannot comprehend the magnitude of change that is coming in the medium term. Being everywhere will likely make us less conscious of it, in an ironic twist that we have seen with other technologies. AI will disappear into ubiquity."*



# DISRUPTION MANAGED: Bringing “gen AI- in-the-room”



**CHRISTOPHER  
SCHEEFER**

Intelligent Industry Lead, Insights &  
Data North America, Capgemini



Generative AI has the potential to revolutionize and disrupt the landscape of the business world. We are bringing “gen AI-in-the-room” by integrating AI and large language models (LLMs) into how we design, create, and innovate with our clients. Experiencing generative AI firsthand, as in-room advisors, AI assistants, and co-participants is an eye-opening endeavor that is helping our clients rethink, reimagine, and reshape their business strategies.

## Generative AI is the new blue with massive potential

As one of the hottest technological innovations in the marketplace today, generative AI has become the dinner-table topic of dystopian futures or the future savior of collegiate essay writers. What may be true is that generative AI is the future of business. It has the potential to revolutionize the way we work by automating tasks that were once thought impossible for machines to do. This technology is capable of creating content, developing strategies, and even making decisions. With generative AI, businesses can streamline their operations and optimize their workflows, making them more efficient and effective.

Through many of our AI-powered innovation sessions, we are redefining how our clients ideate, capture, share, analyze, and innovate using generative AI. Within these sessions we are actively experimenting with our clients around removing traditional constraints and barriers on how work gets done, accelerating analysis, insights, and applied learnings that have restricted rapid decision making and strategic-to-tactical execution.

## What is generative AI?

The emergence of generative artificial intelligence (AI) is changing the landscape of the business world, and its impact is already being felt. This advanced technology has the potential to revolutionize the way work is performed in a business environment. From automating mundane tasks to enhancing creativity, generative AI can

*“Experiencing generative AI firsthand, as in-room advisors, AI assistants, and co-participants is an eye-opening endeavor that is helping our clients rethink, reimagine, and reshape their business strategies.”*

streamline operations, improve productivity, and create new opportunities for growth.

Generative AI uses deep-learning algorithms and LLMs to create new data that can be used to train other machine-learning models and provide a step-change in the maturity of natural language understanding. This means that generative AI can produce original and high-quality data that can be used for various applications, such as generating human-like contextualized responses, unique imagery, and even videos.

## How will generative AI impact how we work?

One of the key ways generative AI will affect the way work is performed in a business environment is by automating repetitive and tedious tasks. In many industries, such as manufacturing and logistics, there are many routine tasks that need to be performed every day. These can be time consuming and lead to

burnout and decreased productivity among employees.

Through integrating AI, we can now automate these tasks, freeing up employees to focus on more complex and creative work. For example, in manufacturing, generative AI can be used to simulate, predict, and automate quality-control processes, such as inspecting products for defects. This can save time and increase efficiency, allowing employees to focus on more critical tasks such as design and product development.

Another way that generative AI will affect the way work is performed in a business environment is by enhancing creativity. Traditionally, creativity has been considered a human-only attribute, and machines were not considered capable of replicating it. However, generative AI is changing this perception through LLMs capable of creating essays and fictional stories, editing documents, and synthesizing large amounts of corporate information into tangible insights.

## Bridging cyber/physical and AI/humans together

We have started the process of integrating and applying generative AI into many aspects of our core work structures with our clients. One key example is putting the AI/human interaction front and center within our Innovation and Strategy Workshops and “AI Unbound” discovery workshops with our clients. This cutting-edge integration of generative AI is helping clients both understand and assess the value of AI when applied in everyday work and then quantify the potential value it possesses for the workplace.

How do we do it? During these workshops we take an integrated approach utilizing speech recognition front ends with text to speech to allow AI to be a participant “in the room” and present with attendees. Within these sessions, AI is always present and can be conversed with, tasked, and consulted for answers. This gen AI-in-the-room approach helps clients see the usage patterns, efficiencies, and roles an AI advisor and collaborator may provide within their organizations.

Utilizing this gen AI-in-the-room approach, our generative AI performs as a scribe, listener, and team anchor, providing summarized conversation details, synthesis during ideation, and clustering activities and cold-eye analysis of thematic areas. During many of these sessions, generative AI also provides creative services by providing artistic elements of graphic art and drawing services to visualize ideas and concepts. Key roles that AI plays include real-time workshop co-facilitator and near real-time playback, synthesis, analysis of ideas, decisions,

conclusions, and improvements. It is important to note that, given the evolving nature of LLMs, we currently preclude any sharing of client confidential content, information, and intellectual property to ensure we are protecting and restricting any client property and data from being shared publicly.

Within the workshops, one evolving usage pattern is how Generative AI can provide a creative spark in a “design-storming session,” providing potential options for product logos, team names, and product/innovation prototypes, not only shaping the content but also evaluating the application and addressable market and constructing a limited non-financial business case. This evolution is happening quickly; as more generative AI models are being marketed and refined with tighter security and privacy guardrails, the expected uptake and disruption of these tools will be tremendous.

## Beware the hallucinations or ghost in the machine

It is important to note that the current AI models we utilized are not trained for specific workshop tasks as of yet and are being used at the current basis of generative AI “chat” knowledge, with a combinatorial series of interfaces allowing for a conversational interactive model. There are, as always, a series of drawbacks and continual improvement activities that happen in real time during workshops. There are several technical limitations around speech recognition, memory/history, and “hallucinations” that are always present. In many cases, these hallucinations stem from idiomatic and colloquial misinterpretations or speech-to-text inaccuracies that led down some strange and amusing paths. As this is novel technology, clients generally accept it as “ghosts in the machine.” However, to trust AI means it must provide unquestionably valid analysis or it will become yet another Siri or



#DATAPOWERED  
#AI4GOOD  
#DATAMASTERS  
#GENERATIVEAI  
#LLMS  
#APPLIEDAI  
#AIINTHEROOM

# Innovation takeaways

## GEN AI AND THE EVOLUTION OF WORK

Generative AI can perform basic knowledge-worker tasks with incredible speed and efficiency. This will force corporations to rethink and reimagine how work will get done and the human/AI relationship in the workplace.

## GEN AI AND ACCELERATED CREATIVITY

Generative AI can create unique and original content. This will accelerate employee creativity, helping them create new product ideas, market analysis, and design validation.

## GEN AI AND ENTERPRISE KNOWLEDGE

Generative AI can be used to synthesize content on a massive scale, making it ideal for organizations to democratize valuable information and insights.

## GEN AI IN-THE-ROOM

Using generative AI as in-room advisors, AI assistants, and co-participants during workshops is an eye-opening endeavor that helps people absorb and appreciate the potential.

Alexa equivalent, losing the collective trust of its human operators.

## Where is it headed? Building the corporate brain and redefining work

Some of our key development areas are currently around internal and private corporate information LLMs, becoming essentially the “corporate brain,” enabling next-generation game-changing knowledge management, through utilizing the entirety of corporate enterprise internal data and content. This will enable a step change in efficiency of work by providing accessibility to the entirety of internal corporate knowledge with the smoothness of human interfaces and interaction, like ChatGPT and Bard. The critical element of this activity is the creation of privatized and secured AI models, which will dramatically increase uptake and trust in the usage and consumption of these models within businesses.

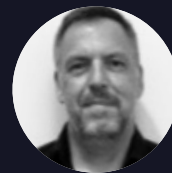
## The first mover opportunity is now

The possibilities with generative AI are almost endless and the acceleration of adoption is taking place at a pace unlike anything before. We are helping shape how our clients integrate AI to accelerate creativity, productivity, and ultimately deliver value through innovative methods to evolve traditional ways of working. Whether through an AI-powered engagement such as an AI-Unbound workshop or developing next-generation private corporate brains, we are always actively working to remove traditional barriers to information and rapid decision making. These new generative AI and LLMs offer tremendous opportunities to rethink and reimagine not just processes, but the entirety of our concept of work, as we integrate AI agents, AI assistants, and digital co-pilots into everyday work constructs.



# I SEE YOU

## Opening the AI black box



JAMES WILSON

I&D Advisory, Insights & Data,  
Capgemini

---



PANTELIS  
HADJIPANTELIS

Lead Data Scientist, Insights & Data,  
Capgemini

---



Your customer needs to be able to see you just as clearly as you see them. Ultimately, the strength of their relationship with you relies on how well you can explain the operations of your AI.



The data scientists were baffled. They had been training the algorithm for weeks, watching it reach new heights of accuracy when identifying dog breeds. Eighty-five percent, 95 percent, eventually it peaked at 98 percent accuracy. However, every now and then, it seemed to throw the same error, identifying a multitude of very distinct breeds as the same thing: a husky. Repeatedly, the model would return husky when the picture was clearly of a poodle, a dalmatian, even a bulldog. Frustrated, they reverse engineered explainability into the model.

The result was shocking.

The algorithm was completely ignoring the features of the animals in the pictures. Instead, to the data scientists' collective surprise, it was using other attributes of the image that happened to be common to every husky picture they had used to train it. It was



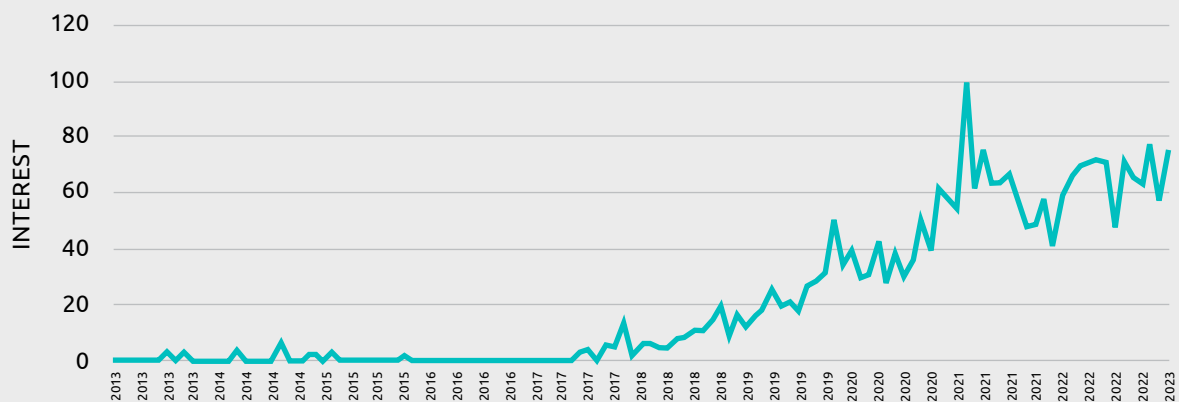
identifying the picture because of the snow and the trees in the background.

Imagine if the machine learning (ML) algorithm that processed your

mortgage application was using equally irrelevant attributes to make a critical decision for your future, such as the gender or ethnic origin you declared on the application form.

There is no question that machine learning-based models are critical enablers for the complex processes inherent in the modern world, however organizations are becoming increasingly aware of the impacts from getting their implementation even slightly wrong. And high-profile examples, including the well-publicized reputational impact suffered by Apple as a result of the [gender bias in the credit limits](#) offered during its credit card launch and the [\\$100 billion market value impact](#) suffered by Google after the launch of its GPT offering, Bard, have ensured that the public are too.

### WORLDWIDE SEARCH INTEREST IN "EXPLAINABLE AI" DATA BY GOOGLE TRENDS



SOURCE: [GOOGLE TRENDS](#), RETRIEVED: 2023-FEB-17

## Can you really see me?

Customer intimacy is a gauge of your alignment with your customers' needs and values. It's more than just talking to your customers; it's about understanding them and understanding their perception of your organization. Being customer-centric relies on cultivating customer intimacy, and that relationship is built on trust. Why am I being offered that particular promotion? Why have I been quoted a particular price for my online purchase or had a particular video suggested to me? In the early days of AI mass-adoption, from the mid-2010s (sometimes termed the "AI summer"), consumers and customers were largely willing to accept the "magic" of artificial intelligence on face value. However, as machine-learning algorithms have become progressively integrated into our interactions with technical systems, service providers, and even other people, terms like "interpretable ML," "explainable AI," and "glass-box model" have become increasingly pertinent to this key group. This demand for

*"Being customer-centric relies on cultivating customer intimacy, and that relationship is built on trust."*

transparency from customers, consumers, and citizens alike is only going to be further amplified as their digital dexterity and AI literacy increases, supported by the increasing focus on customer experience and the simplification introduced through automation and AI.

## 20:20 vision

Legislation is catching up; just consider the explainability clauses proposed in the EU and UK's forthcoming AI strategies, or even just the tenets of GDPR, but irrespective of how much an organization fears falling afoul of some current or future law, good corporate governance is a critical discipline; it is an enabler and liberator for an organization when it is done properly and, conversely, an

inhibitor for corporate growth when it isn't. Having clear control of the organization's processes is at the core of getting governance right, so understanding the AI that is powering them should not be considered optional.

Investigating and validating modern end-to-end ML is akin to opening the hood of a modern car. With some basic guidance, we can identify the major components but that's about it for most people. Knowing how all those components propel your vehicle is as inscrutable as understanding just why a hotel room in Croydon suddenly costs so much more on a particular day on that travel booking site. And while, in the case of a car, there are well-understood laws, rigorous tests, and certifications that ensure the vehicle's road worthiness, those tests and certifications are not yet formally in place for AI systems. Yet. For the most part, ML models remain unregulated, custom-designed information processing systems that we have little understanding of and are too often referred to as "black boxes."



#DATAPOWERED  
#AI4GOOD  
#DATAMASTERS  
#KNOWLEDGEGRAPHS  
#RESPONSIBLEAI  
#EXPLAINABLEAI  
#CUSTOMERINTIMACY

# Innovation takeaways

## EXPLAINABLE AI IS IN DEMAND

The public are already demanding that the decisions organizations make with AI are explainable and appropriately transparent.

## CONSUMER SENTIMENT IS CHANGING

As digital dexterity and AI literacy increases, people are no longer willing to accept the “magic” of algorithms on face value.

## THE TIME TO ACT IS NOW

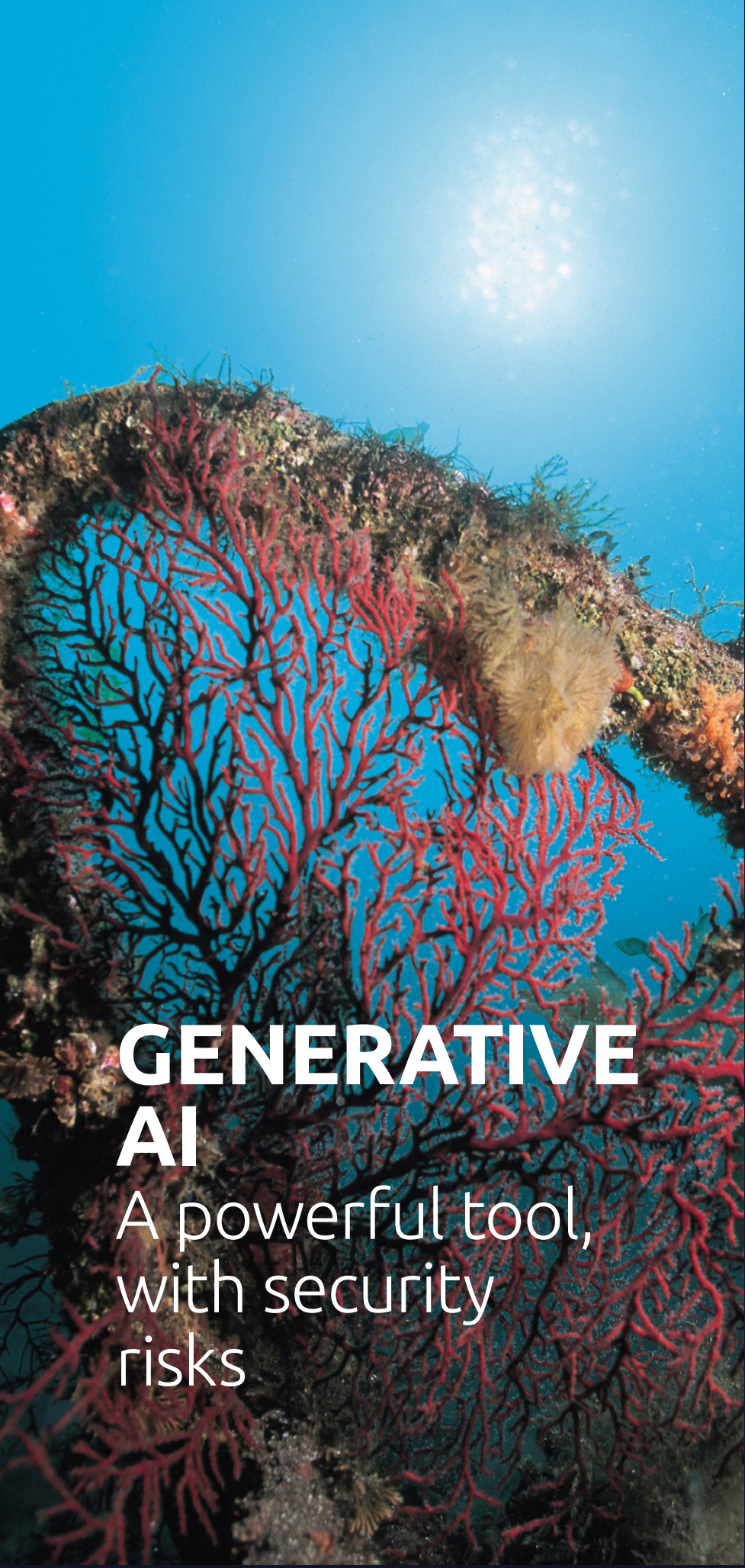
Implementing explainable practices in AI and ML will provide much-needed trust and customer intimacy.

## Opening the black box

For those reasons, a black-box AI system is no longer adequate for a competitive organization, especially when there are increasingly compelling requirements for proactive responsibility, compliance, and sustainability targets. Major technology industry players have themselves introduced open-source toolkits for assessing biases and explaining model behavior, with [Microsoft’s Fairlearn](#) and [IBM’s AI Fairness 360 being](#) the most prominent examples.

Almost every ML project we undertake involves work on explainability, because it allows us to rationalize model behavior when presenting results to our clients. And thanks to the advances in explainable AI, we have seen a resurgence of interest in previously niche research areas such as causal ML, fair ML, ethical AI, and sustainable AI. Taking the latter as an example, using explainable AI techniques to better understand what a model is actually doing will lift the lid on whether the additional power consumption required, that “carbon investment,” is really worthwhile if it only improves the accuracy of the model by a few percentage points.

It is true that AI is currently allowing us to explore information use that we never previously thought possible (one of the authors is really looking forward to AI-generated unit tests). Nevertheless, that exploration needs to entail the introspection and explainability of the AI systems themselves, so that we can be certain that everyone benefits from them. In the very near future, as AI becomes increasingly pervasive across all facets of society, people are going to want to know exactly what you think you know about them, and exactly how you are using that information to make decisions on their behalf. Expect an increasing number of questions like this and be prepared to be able to answer them with good explainable AI practices.



# GENERATIVE AI

A powerful tool, with security risks



**MATTHEW O'CONNOR**

Technical Director, Office of the CTO, Google Cloud



Generative AI is a powerful technology that can be used to create new content, improve customer service, automate tasks, and generate new ideas. However, generative AI also poses some security risks, such as data security, model security, bias and fairness, explainability, monitoring and auditing, and privacy. Organizations can mitigate these risks by following best practices to ensure that generative AI is used in a safe and responsible manner.

Generative AI is a rapidly emerging technology that has the potential to revolutionize many aspects of our lives. Generative AI can create new data, such as text, images, or audio, from scratch. This is in contrast to discriminative AI, which can only identify patterns in existing data.

Generative AI is made possible by deep learning, a type of machine learning that allows computers to learn from large amounts of data. Deep learning has been used to train generative AI systems to create realistic-looking images, generate human-quality text, and even compose music.

There are many potential benefits to using generative AI.

- **Create new content:** Generative AI can create new content, such as articles, blog posts, or even books. This can be a valuable tool for businesses that need to produce a lot of content regularly. The technology can also support the reduction in time it takes to generate work, enabling a steady stream of fresh content for marketing purposes.
- **Improve customer service:** Generative AI can improve customer service by providing personalized assistance. Generative AI can create chatbots that can answer customer questions or resolve issues. These types of uses can support both an enterprise's employees and customers.
- **Automate tasks:** The technology can be used to

*“Generative AI is a powerful technology that can be used for good or evil. It is important to be aware of the potential risks and to take steps to mitigate them.”*

automate tasks that are currently done by humans. This can free up human workers to focus on more creative or strategic work. The technology has the potential to eliminate a lot of toil in many standard business practices, such as data entry and workflow.

- **Generate new ideas:** Generative AI can be used to generate new ideas for products, services, or marketing campaigns. This can help businesses stay ahead of the competition.

Generative AI provides a lot of potential to change the way businesses operate. Organizations are just beginning to leverage this power to improve their businesses. This is a very new area, and the market potential is just starting to reveal itself. Most of the current market is focused on startups introducing novel applications of generative AI technology.

Enterprises are thus starting to dip their toes into this space but the growing use of generative AI also presents security risks. Some of these risks are new for AI, some risks are common to IT security. Here are some considerations for securing AI systems.

**Data security:** AI systems rely on large amounts of data to learn and make decisions. The privacy and security of this data is essential. Protect against unauthorized access to the data and ensure it is not used for malicious purposes.

**Model security:** AI models are vulnerable to attacks. One example is adversarial attacks. An attacker manipulates the inputs to the model to produce incorrect outputs. This can lead to incorrect decisions, which can have significant consequences. It is important to design and develop secure models that can resist this.

**Bias and fairness:** If the training data in the models contains biased information, the resulting AI systems may have bias in their decision-making. This can produce discriminatory decisions, which can have serious legal and ethical implications. It is important to consider fairness to ensure that AI and ML system designs reduce bias.

**Explainability:** AI systems are sometimes opaque in their decision-making processes. This makes it difficult to understand how and why decisions are being made. Lack of transparency leads to mistrust and challenges the credibility of the technology. It is important to

#AI #ARTIFICIALINTELLIGENCE  
#MACHINELEARNING #CYBERSECURITY  
#INFORMATIONSECURITY #ITSECURITY  
#SECURITY #HACKING #CYBERATTACK  
#DATABREACH #AISECURITY #AIFORSECURITY  
#MLSECURITY #CYBERSECURITYAI #INFOSECAI  
#ITSECAI #SECURITYAI #HACKINGAI  
#CYBERATTACKAI #DATABREACHAI

# Innovation takeaways

## GENERATIVE AI IS INNOVATIVE

It is a powerful technology that can be used to create new content, improve customer service, automate tasks, and generate new ideas.

## THERE ARE RISKS WITH THE USE OF GENERATIVE AI

Generative AI also poses some security risks, such as data security, model security, bias and fairness, explainability, monitoring and auditing, and privacy.

## COMMON SENSE CAN HELP COMPANIES LEVERAGE GENERATIVE AI

Organizations can mitigate these risks by following best practices, such as protecting data privacy and security, developing secure models, reducing bias in decision-making, making AI systems more explainable, monitoring and auditing AI systems, and considering privacy implications.

develop explainable AI systems that provide clear and transparent explanations for their decision-making processes.

**Monitoring and auditing:** Track and audit AI performance to detect and prevent malicious activities. Include logging and auditing of data inputs and outputs of the systems. Watch the behavior of the algorithms themselves.

**Privacy:** Private data in model building and/or usage should be avoided as much as possible with artificial-intelligence models. This avoids unintended consequences. [Google's Secure AI Framework](#) provides a guide to securing AI for the enterprise.

Securing AI systems is critical to effective deployment in various applications. Considering these issues, organizations can develop secure and trustworthy AI and ML systems. These deliver the desired outcomes and avoid unintended consequences.

In addition to security risks, there are also ethical concerns related to the use of generative AI. For example, some people worry that generative AI could be used to create fake news or propaganda, or to generate deep fakes that could damage someone's reputation. It is important to be aware of these ethical concerns and to take steps to mitigate them when using generative AI. Organizations will want to enact policies on acceptable use of generative AI which appropriately support their business objectives.

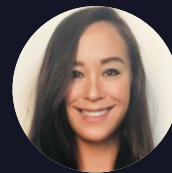
Overall, generative AI is a powerful technology with the potential to revolutionize many aspects of our lives. However, it is important to be aware of the security risks and ethical concerns associated with this technology and to use this technology responsibly. By taking steps to mitigate these risks, we can help to ensure that generative AI is used in a safe and responsible manner and supports your future business goals.

# SERENDIPITY SYSTEMS: Design for AI



CHARLES AUBERT

Technology and Innovation Lead,  
Applied Innovation Exchange,  
Capgemini



CHLOE CHEAU

Head of Experience Engineering,  
Customer First, Insights & Data  
North America, Capgemini



Creating serendipitous experiences is the new frontier of personalization. By embracing experimentation and cross-functional collaboration, businesses can leverage data, intelligence, design, and orchestration to surprise and captivate customers. This article explores how to take the first step towards serendipitous experiences through continuous experimentation and an innovation culture.

Imagine you are browsing an online bookstore for a new novel. You see a list of recommendations based on your previous purchases and ratings. You scroll through the titles, but nothing catches your eye. You feel bored and uninterested.

Now imagine a different scenario. You see a list of books that surprise you with their diversity and relevance. You discover books that you never knew existed, but that match your interests and preferences perfectly. You feel curious and excited.

You have just experienced serendipitous personalization, the art and science of delivering experiences that are both relevant and unexpected to your customers. It is not just about showing them what they want, but also what they didn't know they wanted. The question is, how do you achieve serendipity?

## Achieving serendipity

The foundation for serendipity begins with four dimensions.

1. **Data:** A 360-degree view of the customer and the enterprise is integral to providing consistent and relevant experiences.
2. **Intelligence:** Data-powered culture is key. Linking analytics and outcomes and identifying a holistic set of KPIs ensures that you are optimizing business strategy to accomplish business goals.
3. **Design:** Design is driven by experimentation and is an end-to-end cross-functional exercise from qualitative research and data and analytics

*“The foundation for achieving serendipity begins with four dimensions: data, intelligence, design, and orchestration. Utilizing these dimensions to achieve serendipity requires an organizational culture that rewards experimentation, fosters cross-functional teams, and is customer-centric.”*

to creative development and UI/UX development.

4. **Orchestration:** Orchestrating intelligence and actions for interventions across the customer journey is essential to create consistent experiences. Journey visualization, data, and analytics are critical in empowering employees to improve customer experiences.

Utilizing these dimensions to achieve serendipity requires an organizational culture that rewards experimentation, fosters cross-functional teams, and is customer-centric. Without the right type of environment or governance, it is likely that any attempt at creating serendipitous

personalization will fail or be unsuccessful. While this article's purpose isn't to focus on organizational culture, we will look at how to take the first step in providing serendipitous experiences: how to start the flywheel of continuous experimentation within the business, and pave the way for a new kind of innovation culture.

## Approach creating serendipitous experiences like you would develop a product

Cross-team collaboration including design, research, UI/UX, data, and analytics teams work together to rapidly experiment with new user experiences. Smaller pod-based teams enable agility and agency, improving velocity in providing new innovative user experiences. A holistic, user-driven approach involving a cross-functional pod representing all areas of the business is critical in ensuring your initiative is customer-centric and aligned with business goals.

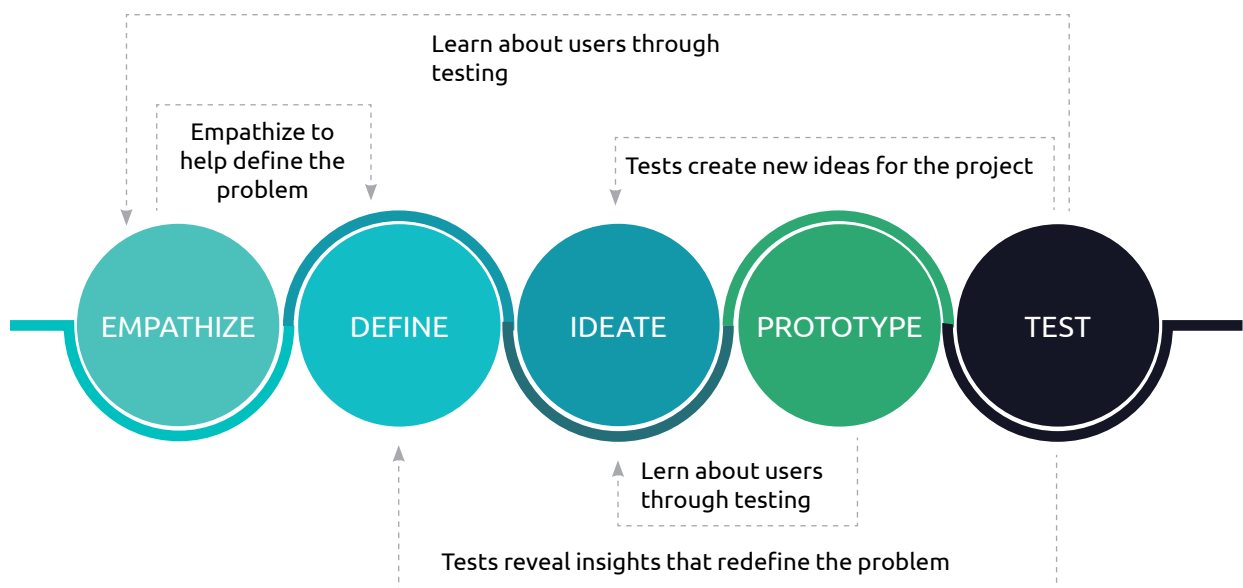
Assembling cross-functional teams is sometimes complex, especially when tackling technical initiatives like serendipitous personalization. Therefore, upskilling and educating are often recommended to reduce the variance in knowledge amongst the group, facilitating alignment in the later stages.

Now that the right organizational structure and culture have been put in place, it is time to take the first step.



## Taking the first step

By using design thinking, we can bring together the four dimensions required to deliver serendipitous experiences. Design thinking is a customer-centric and iterative methodology that leverages empathy, user research, ideation, solutions prototyping, in-situ testing, and feedback collection from real users. As seen in the diagram below, design thinking fosters purposeful iterations through a user-driven feedback loop, instantly measuring success or failure, paving the way for each cycle.



When applying this framework in the context of serendipitous personalization, the following steps are derived.

**Empathize:** Discover your customers' pain points, goals, the channels they use, and the way they engage with your brand. How satisfied are they with your current recommendations? Are you retaining them over time?

**Define:** Translate the insights from the first step into a problem. Validate this problem using data from the business, consumers, or market research. At the end of this step, frame the problem into an opportunity, again supporting business goals and customer needs.

**Ideate:** This is the fun part. Ideate and brainstorm possible solutions, diverging as much as possible with blue-sky ideation before converging, using insights from the previous steps as prioritization lenses. Horizon scoping may be required to identify the first step in building toward an end-state solution.

**Prototype:** Build low-fidelity versions of your solution that can be quickly and cheaply tested against your previously defined success metrics. Each build iteration should have a clear purpose of either trying to prove desirability, viability, or feasibility.

**Test:** Validate your solutions with real customer feedback, production data, and metrics. The key is to

place the user in the center of this step. In fact, insights derived from their feedback will shape the goals of further iterations and ensure their success.

**Iterate:** By following this process iteratively, you can solution, action, and ship personalized experiences that are not only effective but also constantly evolving, responding to the new demands of customers.

Velocity is the key. Leveraging an ecosystem of startups, partners, and hyper-scalers can accelerate your iterations, and amplify the magnitude and effectiveness of your initiative. Startups can help reduce the barrier to entry during the first iterations while hyper scalers can help scale the solutions in the long run.

#DESIGN4AI #INNOVATION  
#SERENDIPITY#DESIGNTHINKING  
#PERSONALIZATION  
#SERENDIPITYSYSTEMS  
#OMNICHANNEL  
#DIGITALANDPHYSICAL  
#DATACULTURE  
#DATAPOWERED  
#USERFIRST

# Innovation takeaways

SERENDIPITOUS EXPERIENCES  
REFLECT AN ORGANIZATION'S  
AGILITY/EFFICIENCY AND  
DEDICATION TO THEIR CUSTOMERS

When organizations deliver seamless experiences, it is a result of cross-functional team synergy and a culture of innovation.

PERSONALIZATION SHOULD NOT  
ONLY BE OMNI-CHANNEL BUT ALSO  
BRIDGE THE DIGITAL AND PHYSICAL  
DIVIDE

The approach to identifying and forming these experiences should be replicated across all interactions with a customer, advancing an organization's personalization strategy: immersive/engaging experiences, leveraging emerging technology, etc.

EMPOWER TEAMS BY PROVIDING  
THEM WITH AGENCY DELIVERS  
TRANSFORMATIVE EXPERIENCES

Create cross-functional pods which support end-to-end design, apply design thinking to facilitate rapid delivery and experimentation, and provide them with the tools and data needed to support their initiatives.

## Digital goes physical

Throughout this story, we have talked a lot about assembling cross-functional teams, leveraging 360-degree customer data views, and fostering an experimentation culture to create serendipitous and personalized digital experiences. What if you could apply the same framework to bridge the gap between in-store and online experiences?

Coming back to our online bookstore scenario, imagine if you could bring the same kind of serendipitous personalization to your local brick-and-mortar bookstore. With recent advancements in AI, data treatment, telecommunication infrastructure, screens, and a wide variety of other technologies, this reality is closer than ever before. A few minutes of design thinking ideation yields some exciting examples.

Digital people on large touchscreen kiosks could greet customers as they enter stores. Leveraging massive amounts of already available customer data, a digital store associate could provide the same serendipitous and personalized recommendations already available on your online channels.

Products themselves could even become vectors for conveying data or unlocking augmented experiences. Using a smartphone app, customers could point their phone at shelves, to get a real-time highlight and customer reviews of books they're most likely to enjoy.

These are just two examples that illustrate how an end-to-end customer-centric approach leveraging cross-functional teams from all practices can unlock experiences never possible before. While being end-state solutions, a roadmap can be uncovered by simply applying design thinking and fostering this innovation/experimentation culture.



# DATA FOR POSITIVE FUTURES

## FROM BIG DATA TO RIGHT DATA: 68

Data platforms in the age of eco-responsibility

*Nicolas Ydder & Nicolas Claudon, Capgemini*

## GLOBAL DATA SCIENCE CHALLENGE FOR A SUSTAINABLE FUTURE: 71

The Biodiversity Buzz

*Dr. Dan Stowell & Bart Braun, Naturalis Biodiversity Center*

## DATA MASTERS DRIVE THE FUTURE OF INNOVATION: 74

Forward-thinking companies scale and transform with data and AI

*Ivar Aune, Thomas Svahn & Magnus Carlsson Capgemini*

## CREATING URBAN FORESTS: 77

With a little help from data analytics and visualization

*Anil Kandpal, Douglas Hoover & Daniel Freytag, Capgemini*

# FROM BIG DATA TO RIGHT DATA

## Data platforms in the age of eco-responsibility



**NICOLAS YDDER**

Managing Data Architect, Insights & Data France, Capgemini



**NICOLAS CLAUDON**

CTO, Insights & Data, South Central Europe, Capgemini



The current practice of collecting and storing large amounts of data is often counterproductive and environmentally damaging. Adopting a frugal data approach, where quality is prioritized over quantity, can help reduce the amount of data collected and improve environmental impact. This requires a cultural and strategic shift, including establishing a governance body and implementing operational measures such as optimizing storage and

differentiating storage modes. A lifecycle approach and focusing on information needed, rather than data, can also lead to data-saving choices. This reinforces control over data assets, reduces costs and risks, and aligns with a company's CSR strategy.

When it comes to data, it has long been considered that the more data you collect, the better off you are. After all, it could always be used. And with the rise of big data and the cloud, it was believed there was no reason to limit ourselves, since the infrastructures could support gigantic volumes without strain. However, we are aware that masses of data, measured in petabytes, are rarely useful or are even counterproductive, as it becomes impossible to find your way around. As a result, up to 80 percent of the data is not used.

But even unused data requires transfers, handling, processing, storage, security measures, and more. And since all this happens on a very large scale, the resulting environmental footprint is not negligible. For example, in the European Union, data centers accounted for [2.7 percent of electricity demand in 2018](#) and will reach 3.21 percent in 2030 at current pace. At a time when IT is being criticized for its growing ecological impact, fighting against this "infobesity" is a priority in terms of eco-responsibility.

## Cultural and strategic issue

Data frugality means breaking with the habit and ease of keeping

everything. From exhaustive big data, we need to move to right data: prioritize quality over quantity by collecting and keeping only what is useful. But choosing and selecting is an effort and a change, and therefore above all a cultural issue. Everyone must be aware that, despite its "immaterial" nature, the accumulation of data is a waste that is detrimental to the company and the environment and that it is everyone's responsibility to reduce it: IT, in charge of infrastructures; the data organization, in charge of data assets; and the business lines, the only ones able to evaluate the value and interest of datasets.

## Necessary governance

Data frugality is thus becoming a common goal. However, not everyone will have the same point of view on what should be kept, for what purpose and under what conditions. It is therefore necessary to set up a governance body to define policy, guidelines, and roles, and to arbitrate differences of opinion. At the technical level, it can be supported by a "design authority," embodied by architects and business decision-makers who will issue rules, manage their deployment, and ensure they are applied rigorously and consistently.

One of the reasons for the inflation of data is that no one is responsible for controlling volume. In the framework of governance, it is therefore essential that someone takes on this role. It will be up to that person to ensure data has optimal business impact and veto if its environmental impact outweighs the benefits. To make such decisions, it will be necessary to put in place indicators more detailed than the volume of data and storage cost. All of this should be

managed at the project portfolio level.

## Operational measures

In practical terms, data frugality involves several operational measures, many of which can begin to be implemented without waiting for the governance framework to be established. These actions are the priority to obtain the first significant gains and to initiate a change in perception.

- **Storage:** A lot of data is not up for debate and could be rationalized by simply leveraging platform features such as deduplication to optimize storage devices, with positive returns in line with a "FinOps" approach.
- **Access:** Another priority area that is relatively easy to implement is the differentiation of storage modes over the data lifecycle, with "hot" access for the most immediately needed data, "warm" access for less frequent analysis and reporting, "cold" for precautionary archiving, and finally archiving for purely historical purposes, on tape.
- **The form:** At the level of the data architecture, several technical levers can help limit volumes, such as compression (provided that the gains are not absorbed by too frequent decompression), binary serialization (optimization of object storage), data virtualization (to avoid unnecessary replication), and data sharing.

## Adopt a lifecycle approach

But the most important change must occur at the level of data

#DATA #DATAFRUGALITY  
#DATASHARING  
#DATACOLLABORATIVEECOSYSTEM  
#ENVIRONMENTALIMPACT

# Innovation takeaways

## PRIORITIZE QUALITY OVER QUANTITY

By collecting and keeping only what is useful, which requires a cultural and strategic shift towards a data frugality approach.

## ESTABLISH A GOVERNANCE BODY

To define a policy, guidelines, roles, and to arbitrate differences of opinion. Implement operational measures such as optimizing storage and differentiating storage modes.

## WORK THE DATA ARCHITECTURE

Several technical levers can help limit volumes, such as compression (provided that the gains are not absorbed by too frequent decompression), binary serialization (optimization of object storage), data virtualization (to avoid unnecessary replication), and data sharing.

projects and products, where we must now focus on information, on what we need to know, and not on the data that enables us to know it.

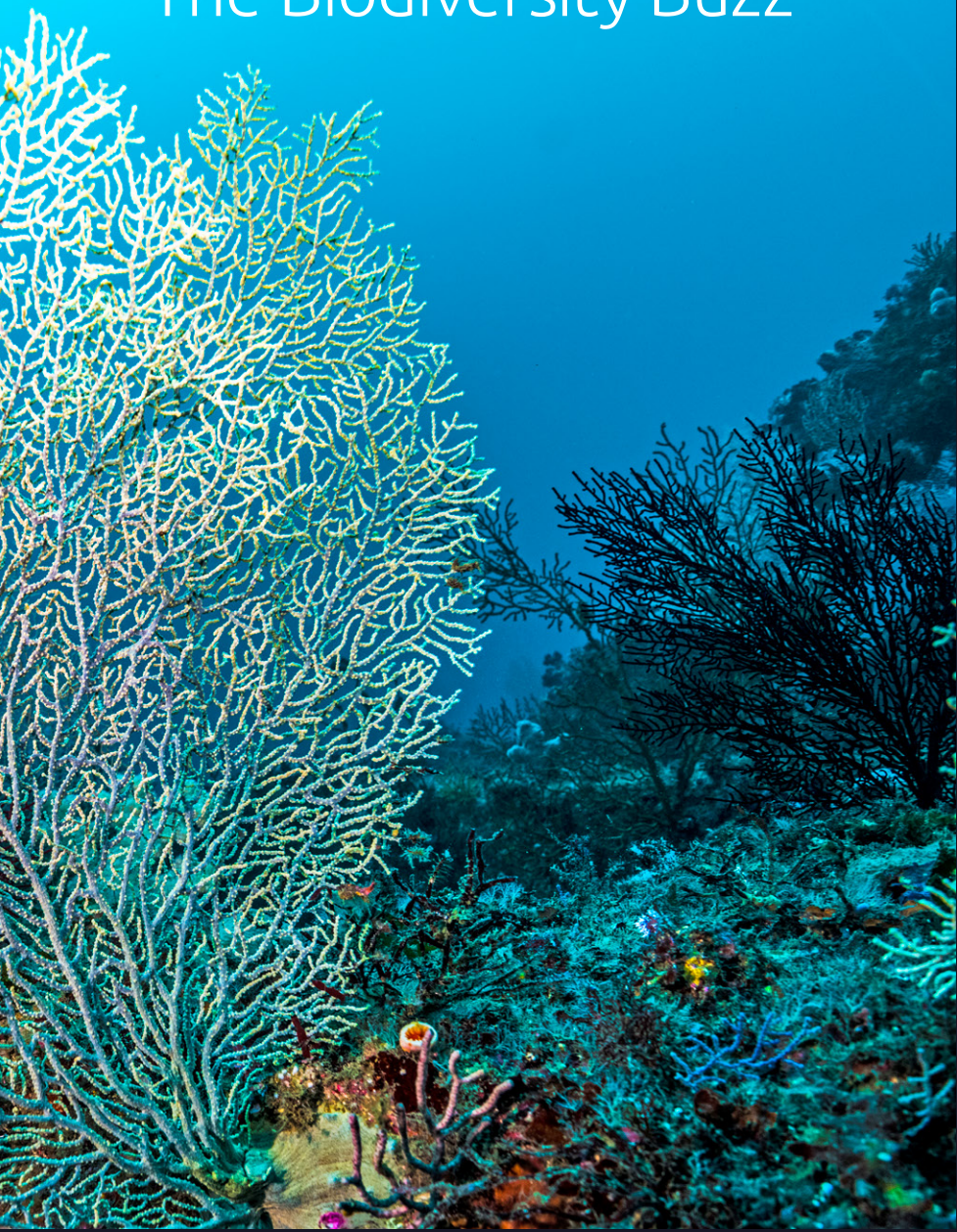
In this way, it is possible to make data-saving choices throughout the lifecycle without detriment to business results, by using, if available, third-party data rather than collecting and owning one's own; filtering data at the source and pre-processing it to bring up only what makes sense; choosing pre-trained or data-saving algorithms (few-shot learning, zero-shot learning); determine the threshold of precision/relevance that is just necessary and do not extend the calculations beyond that threshold; keep only the results and not the raw data that made it possible to obtain them (or keep only representative samples); and be satisfied, when possible and relevant, with aggregated results and averages, rather than detailed figures. Note that all these measures need to be documented and traceable in case there is a need to account for the various sorts and deletions made.

If they coordinate their efforts, business, IT, and data organizations have many levers to reduce data volumes and their environmental footprint. Above all, by considering the reflex of constantly questioning the meaning and usefulness of what is collected, this policy of frugality reinforces the control of data assets, which, in turn, also contributes to reducing costs and associated risks. For the time being, there are no regulations that require companies to moderate the amount of data they collect but, from all points of view, it is in their best interest to anticipate this.

Data frugality is a concern that transcends the scope of green IT alone and intersects with the broader framework of a company's CSR strategy. This approach to data collection and storage aligns with the transformation towards a more data-driven organization. Therefore, driving the change towards data frugality must be a top-level decision and included in the strategic objectives for the environment.

# GLOBAL DATA SCIENCE CHALLENGE FOR A SUSTAINABLE FUTURE

## The Biodiversity Buzz



DR. DAN STOWELL

Associate Professor of AI & Biodiversity, Naturalis Biodiversity Center



BART BRAUN

Science Communication Officer, Naturalis Biodiversity Center



Being able to recognize animals from their sounds will make a crucial difference to biology and nature conservation. It also presents a great challenge for data engineers and scientists. "It's a whole new field for people to do impactful innovation in."

Biologists have a problem that is rare among the sciences. The physicists are not running out of particles and the rate of star formation in the universe is slowing down, but not on a timescale that is bothersome for astronomers. But those who study nature see that their subject is in trouble. Some call it the "biodiversity crisis" or even

“the Sixth Extinction,” with capitals to emphasize how bad things are.

There are silver linings, but the general trends are all going down. This is obviously a bad thing for animals, plants, fungi, and other living things. It is also bad news for humans, who depend on these species for pollinating crops and many other natural services, some not even properly understood. To protect what’s left, the first order of business is to find out what is out there. Which species occur in each area? If we take measures to protect them, are these measures helping? This work is called biodiversity monitoring, and help is urgently needed. Help from data engineers and data scientists, to be precise.

## AI species identification

The true biodiversity specialists are slowly going extinct themselves, as it takes many years to learn the skills of a taxonomist. And even if they weren’t, there are simply too many different species. There are millions of species on earth and most of them are not even properly described by science. Getting people to go out and count them by hand is not even conceivable, let alone doable. Which is why

*“Digital insect recognition is an unexplored field, with lots of different types of deep learning that could be used”*

biodiversity research institutes have been hiring data scientists to help them with digital species identification. If you could put up a camera and get an AI to recognize all the animals that pass in front of it, that will save a lot of work. The same goes for sounds: a microphone that helps recognize all the bats, frogs, birds, and crickets in an area is highly preferable to going into a moss-filled swamp and trying to catch a moss-colored cricket the size of a jellybean.

Sound recognition in a swamp? “For birdsong, this works,” says **Dr. Dan Stowell** of [Naturalis Biodiversity Center](#), the natural history museum of the Netherlands. Stowell is one of the members of the fast-growing club of computer scientists who work in biodiversity research. In

fact, his work is part of the reason that digital identification of birdsong is now possible. But birds are just the first step: “Everyone recognizes that insects are the most important for ecosystems. Birdsong recognition works by adapting techniques already in use for human speech. But insect sounds are more different from speech than birdsong is. The techniques don’t work as well. It’s currently not clear what is the best algorithmic approach for digital insect recognition. That means it’s a whole new field for people to do impactful innovation in.”

## Insect sounds are weird

In the upcoming [Global Data Science Challenge \(#GDSC6\)](#), organized by the Naturalis Biodiversity Center with Capgemini and AWS, participants are challenged to help identify 66 European species of crickets, cicadas, and grasshoppers based on their mating sounds, Stowell explains. “This is basically an unexplored area of research, with lots of different types of deep learning that could be used. For example, when analyzing complex sounds like speech or birdsong, you





#BIODIVERSITY  
#DIGITALSPECIESID  
#AIFORGOOD  
#GDSC6

# Innovation takeaways

## TO RECOGNIZE IS TO CONSERVE

Digital species recognition is crucial to biodiversity and nature conservation.

## RECOGNIZING BY HAND DOES NOT CUT IT

There are too many species and too few people that can do it.

## DATA SCIENCE COMES TO THE RESCUE

The next Global Data Science Challenge will explore ways to apply data engineering and AI to automatically identify 66 European species of insects.

turn the sound into a spectrogram and then compare spectrograms. But insect sounds are a bit weirder; maybe it's better to skip the spectrograms altogether? Does a new kind of neural network fit the task better? Another problem is that the dataset is not so big; so how are we going to treat it?"

The advances in digital species recognition go hand in hand with developments in the devices that gather data. If a camera or microphone recognizes the sounds itself, on-device, and this equipment is small and affordable, you can put up intelligent cameras everywhere and get data on an unprecedented scale and resolution. "It used to be that you had to pay thousands of dollars for a big box that you could put in the field. Now, you see much smaller devices, and lots of projects are heading off to monitor all kinds of things. The vision is to make a tiny device that is smart." Instead of streaming its recordings, it just sends a list of all the species it identified itself. Imagine birdhouses that don't just tell you that the nest inside is built by the same great tit couple that lived there last year, but also informs you which species of insects they are feeding to their chicks. Imagine waterproof microphones that provide new insights into the calls made by fish and sea mammals. Imagine having such devices everywhere, and connected to each other, so you could have almost real-time biodiversity monitoring.

Both national governments and the EU are keen to have the data such devices could provide, and large research infrastructures such as ARISE in the Netherlands are already being built to connect all the data such devices gather. "This is not just a quaint little hobby for biologists – the EU is paying us to do it," Stowell emphasizes. "And once we've done it here, we can roll it out among many countries all over the world. In India and Africa, the need for such monitoring and the gaps in our knowledge are even greater. Together, a little bit of data science could have a big multiplier of our effect on conservation."

# DATA MASTERS DRIVE THE FUTURE OF INNOVATION

Forward-thinking  
companies scale  
and transform with  
data and AI



IVAR AUNE

Head of Insights & Data, Nordics,  
Capgemini

---



THOMAS SVAHN

Head of Insights & Data, Sweden,  
Capgemini

---



MAGNUS CARLSSON

Head of Innovation Insights & Data,  
Nordics, Capgemini

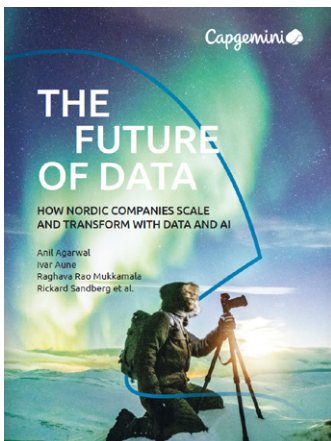
---



Today's forward-thinking companies are using data as a regenerative resource to scale swiftly and increase profits. Insights drawn from data can help transform businesses, enabling them to predict market trends, anticipate customer behavior, manage risks, and

drive innovation. The creation of a data-driven culture is the future-focused way companies can put this power to work.

Using data to derive business intelligence is not new. But a tectonic shift is coming in which organizations that have data engineered into their DNA will take the lead in an increasingly fast-paced and competitive business landscape. Those who don't keep up will fall to the back of the pack – or disappear altogether.



The newly released book [The Future of Data: How Nordic Companies Scale and Transform with Data and AI](#) looks forward 10 years to predict what will happen in the field of data and analytics from a business perspective. It provides real-world examples and success stories from companies such as the LEGO Group and Wireless Car, among others, which are already set up to reap the rewards of becoming data masters. Although these companies are headquartered in the Nordic nations, their insights are universal and can be adopted by businesses worldwide.

“Over the course of the next decade, the winners will be those organizations who can make sense of their data and provide data infrastructure and services to scale the value across the enterprise,”

said Valeri Voev, Principal Data Engineer, The LEGO Group. Voev is one of many leaders who share insights in *The Future of Data*, a collaboration between Capgemini, Copenhagen Business School, and the Stockholm School of Economics.

## A foundation built on business and technology

In the coming years, new tools and technologies will amplify the opportunity to leverage data in revolutionary ways. Adoption of the metaverse, 5G, and Web3 will drive much of the development but, for companies to succeed, they must focus on business processes, people, and culture. Data is the glue that will join these worlds to create a roadmap to mastery.

But what differentiates data masters from everyone else? The answer lies in how business leaders use data to scale, transform and, ultimately, profit. It distills down to basic areas – which are often overlooked – that require deliberate planning, followed by execution, analysis, and iteration.

## Digital transformation and sustainability

Whether organizations are in the automotive sector, telecommunications, or consumer products and retail, success starts with planning. Data masters don't just set up a data strategy, they ensure it's aligned with their business strategy. Both technical and business teams work on the same data in real-time and have a forward-looking and sustainable strategy for collecting data. This cohesive approach enables organizations to use this information quickly and effectively across business units.

Trust is everything when it comes to driving a business forward. It is critical that data is collected in a mindful way, prioritizing quality

*“Over the course of the next decade, the winners will be those organizations who can make sense of their data and provide data infrastructure and services to scale the value across the enterprise.” – Valeri Voev, Principal Data Engineer, The LEGO Group*

over quantity. For instance, data masters must invest in secured and measured data quality and fair and transparent artificial intelligence (AI). Data must be reliable and meet strict obligations with respect to privacy regulations whether used within an organization or shared with external partners.

Future-focused organizations have moved to a modern data landscape. Legacy architecture has shifted to cloud adoption, with data masters prioritizing the ease and accessibility of data. More and more organizations will implement a data center infrastructure sustainability program in the forthcoming years. Sustainable data helps companies use such intelligence for more effective decision-making overall that extends to every area of the business, from technology to finance, marketing, and human resources.

## Activate data to stay ahead

Transforming data and insights into action drives businesses forward. Initiatives have resulted in tangible use cases that demonstrate improved efficiency and the opportunity to create new revenue streams.

#BIGDATA  
#DATAMANAGEMENT  
#INNOVATION  
#THEFUTUREOFDATABOOK

# Innovation takeaways

## DON'T WAIT TO GET STARTED

Looking back five to 10 years, many of the most successful organizations of today are the ones that started early. It will never be the perfect time, but companies can benefit from learning and adapting along the way.

## DEVELOP A STRATEGY

Data masters create a cohesive plan, integrating and aligning their data and business strategies.

## FOCUS ON SUSTAINABILITY

Quality data is key. Bring together technical and business teams to develop and implement a forward-looking and sustainable strategy for collecting data.

Wireless Car, a data-driven company in the Swedish automotive sector, realized the critical value of trustworthy data. Established in 1999, the company developed long-term relationships with its clients. However, data volume has increased as the industry transformation has sped up with electric vehicles, autonomous driving, and the software-defined car. As seen globally with data breaches, even what appears to be a small erosion of trust can have fast and far-reaching consequences. Wireless Car's goal is to build the most trusted automotive data ecosystem to support revenue creation. Opportunities include unlocking the business value in vehicle-use insights and building products and services to expand its digital relationship with customers.

In comparison, the LEGO Group is training its focus on its data-native customer base. The toy company's name is an abbreviation of the Danish words "leg godt," meaning "play well," and it's also playing smart, piecing together a data platform that includes data discoverability and intelligence. This infrastructure will support a new generation of data analysts, helping them create innovations within a supportive environment and culture that lets them excel at what they do.

When data is used effectively, all the pieces shift into place, especially for business leaders seeking a lift in profitability. In 2020, the Capgemini Research Institute undertook a survey of approximately 1,000 organizations across 10 sectors globally. It studied how companies derived value from the increasing volumes of data, isolating a group of what it termed "data masters" to measure results. Compared to their peers, data masters outperformed significantly on financial metrics, on average realizing 70 percent higher revenue per employee, 245 percent higher fixed asset turnover, and 22 percent higher profitability.

Data is a defining feature of the future. An organization's ability to use all its data and participate in strategic data ecosystems could potentially lead to a huge competitive advantage and have a big and positive impact on organizations, businesses, and our everyday life.

# CREATING URBAN FORESTS

With a little help  
from data  
analytics and  
visualization



**ANIL KANDPAL**

Delivery Architect Director &  
Sustainability Analytics Lead  
– Insights & Data, Capgemini

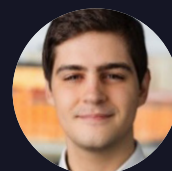
---



**DOUGLAS HOOVER**

Senior Consultant – Insights & Data,  
Capgemini

---



**DANIEL FREYTAG**

Business Analyst – Insights & Data,  
Capgemini

---



The loss of biodiversity is a global challenge with impacts on natural ecosystems around the world. The creation of urban forests is an important movement to tackle biodiversity loss, rebuilding the connection between nature and citizens in built-up areas. The democratization and ease of use of technologies can empower these citizens to work with data-powered solutions, which accelerates the growth and reach of urban forests.

Deforestation and loss of biodiversity are two of the major causes of climate change leading to extreme weather events that affect millions of lives globally.

The Miyawaki Method is a way of creating and growing small urban forests in very limited space, in very little time, while also generating significantly more biodiversity and capturing more CO<sub>2</sub> per square meter than normal forests.

Despite its incredible potential, implementation of the Miyawaki Method has been limited around the world due to a lack of data analytics and visualization being applied to it.

This is where the power of data and AI, applied to this unique method, comes into play. With an advanced data and analytics approach, we aim to raise awareness in society, enable faster implementation of such reforestation projects, and thus jump start an important mass movement.

## Restoring biodiversity

Climate change has often been branded as the challenge of our generation. But a parallel problem has just as much relevance today: biodiversity loss. Biodiversity is essential for the maintenance of ecosystems, and thus fundamental for nature to support a healthy environment. Since the 1970s, the world has seen a great decline in the population of wild animals. Today in Europe, [81 percent of natural habitats](#) have a poor conservation status. This global problem is intertwined with climate change and needs to be addressed at the same level.

Tackling biodiversity is a complex challenge that requires a multi-sided approach. There are global challenges that require international collaboration, such as protecting endangered species and stopping illegal trade, protecting rainforests, etc., but there are also many local biodiversity challenges, such as the drop in the population of pollinators, that are worrisome and could lead to catastrophic damage to local ecosystems.

To address biodiversity loss, we need to also look at local ecosystems and how to create a healthy and sustainable relation between humans and nature.

Restoring biodiversity in urban areas is an important initiative to begin tackling these issues. Urban centers often lack green areas, which can affect quality of life for citizens, create heat islands, and limit biodiversity in local ecosystems. But in such confined spaces, and with a limited area to work in, what kind of impact could that really have?

*“We cannot address biodiversity loss without tackling climate change, but it is equally impossible to tackle climate change without addressing biodiversity loss.”*  
— European Commission

## The Miyawaki Method

Here is where the Miyawaki Method comes in. Using it, one can create biodiversity and regenerate urban areas by planting pocket forests.

Developed by the Japanese botanist and ecology expert Professor Akira Miyawaki, its afforestation principles – in which trees are planted where there were previously no trees – are based on how native tree species would grow and interact in a natural forest in local soil and weather conditions.

Native tree species are planted closely together, thus making them shoot upwards instead of sideways to compete for light. This creates a multi-story structure, where different layers of vegetation appear, creating a healthy, climate-resistant, and biodiverse habitat for many organisms.

CANOPY LAYER

TREE LAYER

SUB-TREE LAYER

SHRUB LAYER



## THE DIFFERENT LAYERS IN A DENSE NATIVE FOREST

This enables the creation of highly effective urban forests, which can grow much faster and denser than a conventional forest, store more carbon, and hold more biodiversity.

But perhaps the most important aspect of all, the Miyawaki method is participative. It looks for local communities to engage with planting and caring for the forests, truly empowering citizens in urban settings to reestablish their contact with nature.

### Environmental stewardship

Humans are meant to be in nature, to relate to the environment and interact constructively with it. Instead, we have devastated the environment, created massive urban areas, and distanced ourselves from nature. In the journey to restoring biodiversity, society must once again learn to live in constructive collaboration with the environment. We must step up to the role of



### CITIZEN SCIENTIST

stewards of this small planet we live on, looking to take care of the nature around us and enable biodiversity to thrive around the planet.

We can look to create urban forests and engage the local communities. This enables people to act and contribute to sustainability with hands-on activities, while also rebuilding their connection to nature. Citizens can look to plant,

grow, and care for the forests in their neighborhoods.

A central task of an urban forest is to ensure its maintenance by engaging the community to explore the conditions of the forest, how it is developing, and how it is affecting the region. Citizens can take up the role of citizen scientists, collecting data about biodiversity, carbon storage, microclimates, and many

#BIODIVERSITY  
#AI4GOOD  
#4POSITIVEFUTURES

# Innovation takeaways

## NEED FOR URBAN GREENING

Urban areas often suffer from a lack of green spaces, distancing citizens from nature.

## ENVIRONMENTAL STEWARDSHIP

It's necessary to recreate a connection between society and nature, where society can become nature's steward.

## CITIZEN SCIENTIST MEETS CITIZEN PROGRAMMER

Technology can enable citizen scientists to work with data-driven solutions and drive change.

other topics, creating a bond with nature and being a steward of the environment.

## Technology can do wonders

Technology is ever more accessible to the public and there's a need to take advantage of innovation, enabling people to use technology to drive sustainable changes around the world. This is no different for urban forests. Without data and data-powered solutions, there is little room for analysis and for coordinating efforts and results on a larger scale.

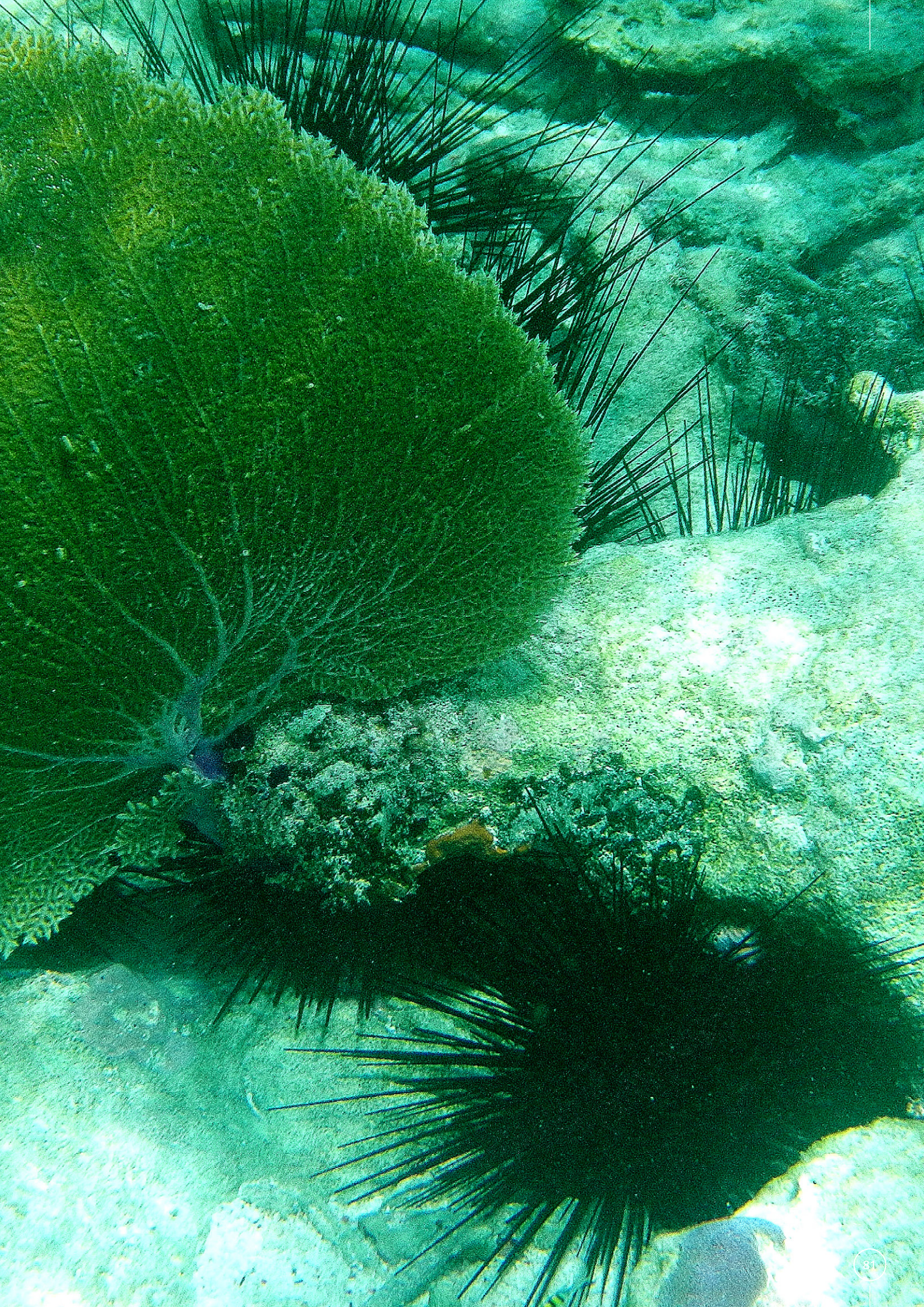
Citizen scientists can become the pillar for environmental stewardship, engaging with forests, collecting data, and helping to restore and keep biodiversity. But the citizen developer can enable this data to be collected and stored in a structured way, using no-code technology to work with data-driven solutions that maximize the work of the citizen scientist.

This is the place where the citizen scientist and the citizen developer meet: combining the democratization and ease of use of technologies with the environmental stewardship and engagement of communities to preserve biodiversity.

## A data-powered solution

By supporting and enabling the Miyawaki Method with a data-driven solution, we are building a framework for enabling citizens to take up the role of scientists and developers. This can help coordinate the data collection and maximize the value of the data collected, enabling more visibility of this innovative method to the global audience. But most importantly, this can engage people to become environmental stewards and work locally to restore biodiversity and find synergy with the environment.





# INNOVATION MOVERS AND SHAKERS



**DR. AKHAURI  
PRAKASH KUMAR**

Chief Strategy Advisor, Insights & Data,  
North Central Europe, Capgemini



**ARNE  
ROSSMANN**

Chief Architect Data & AI for  
Intelligent Industry, Capgemini



**ARUNA  
PATTAM**

Head AI Analytics & Data Science,  
Insights & Data, APAC, Capgemini



**CHRISTOPHER  
SCHEEFER**

Intelligent Industry Lead, Insights & Data  
North America, Capgemini



**DINAND  
TINHOLT**

VP, Insights & Data, North America  
Capgemini



**JAMES  
WILSON**

I&D Advisory, Insights & Data,  
Capgemini



**LIZ HENDERSON**

Executive Advisor,  
Capgemini



**MARIJN MARKUS**

AI Lead, Managing Data Scientist,  
Insights & Data, Capgemini



**MARK ROBERTS**

Principal Technologist,  
Capgemini Engineering



**MARK OOST**

Global Offer Leader AI, Analytics & Data  
science, Insights & Data, Capgemini



# INNOVATION MOVERS AND SHAKERS



## MONISH SURI

Global Innovation Lead  
Insights & Data, Capgemini



## NICOLAS CLAUDON

CTO, Insights & Data, South Central  
Europe, Capgemini



## OLIVER PERHIRIN

CTO Insights & Data  
France, Capgemini



## ROBERT ENGELS

CTO Insights & Data North Central  
Europe, Capgemini



## RON TOLIDO

CTO Insights & Data  
Capgemini



## ROOSA SÄÄNTTI

Head of Insights & Data  
Finland, Capgemini



## DR. SERGEY PATSKO

Data & AI Deputy Group Offer Leader  
Capgemini



## URBAN EKERÖTH

Head of Portfolio, Insights & Data  
Nordics, Capgemini



## YASHWARDHAN SOWALE

Architecture Head, India Domain Leader  
for AI, Insights & Data, Capgemini



## VINOD JAIN

Head of Delivery, Insights & Data  
India, Capgemini



## WEIWEI FENG

AI Nordic Technique Lead, Insights & Data,  
Capgemini





## About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided every day by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of over 360,000 team members in more than 50 countries. With its strong 55-year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fueled by the fast-evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering, and platforms. The Group reported in 2022 global revenues of €22 billion.

Get the Future You Want | [www.capgemini.com](http://www.capgemini.com)

For more details contact us:

**RON TOLIDO**

CTO  
Insights & Data, Capgemini



**MONISH SURI**

Global Innovation Lead  
Insights & Data, Capgemini



Interested in learning more?

[Download our previous Waves here](#)