

WAVE III | 2022

DATA-POWERED INNOVATION REVIEW



THE RISE OF DATA COLLABORATION



DATA MESH: USING PRODUCT THINKING TO UNLOCK VALUE IN DATA | 06

Yashowardhan Sowale
Marius Nilsen

CREATING A DATA-POWERED CULTURE | 09

Dinand Tinholt

HOW CROSS-INDUSTRY DATA COLLABORATION POWERS INNOVATION | 12

Eve Besant

USING AI AUGMENTATION TO EMPOWER DATABASE ADMINISTRATORS | 16

Arvind Rao

UNLOCKING THE POWER OF AI WITH DATA MANAGEMENT | 19

Jitesh S. Ghai

THE RISE OF TRANSFORMATIONAL AI



GPT-3 ONE-CLICK CODE TRANSLATION | 23

Anitha Madhusudhan

AI DELIVERS SIMPLER SIMULATIONS FOR A COMPLEX WORLD | 26

Sam Mahalingam

AUGMENTED INSURANCE ADVISORY WITH AI | 29

Damien Philippon
Kiran Boosam

A CASE FOR CONTEXT AWARENESS IN AI | 32

Robert Engels

UNBOX BLACK-BOX MODELS WITH AI GLASS BOX | 35

Jatinder Kumar Kautish
Ashish Dhali

THE RISE OF DATA FOR GOOD



LEVERAGING THE VALUE OF DATA FOR SOCIETY 5.0 | 40

Erik Hoorweg
Judith Kennes

DATA AND THE SUSTAINABILITY ECOSYSTEM | 43

Philip Harker
Courtney Holm

BATTLING RIVER BLINDNESS DISEASE WITH THE HELP OF AI | 46

Dr. Daniel Kühlwein
Dr. Janina Kühlwein

IMPROVING LIVES OF FARMERS WITH AI AND AUTOMATION | 49

Marijn Markus
Luc Baardman

AI FOR DYSLEXIA PATIENTS | 52

Dr. Tanu Solanki
AI from Wordtune

FOREWORD



ZHIWEI JIANG
CEO, Capgemini
Insights & Data Global Business Line

I am proud to present you with this third, literally “cool” edition of the *Data-powered Innovation Review*. We need data, analytics, and AI more than ever to address both the business and societal challenges of these unprecedented times. This publication will hopefully yet again inspire you with concepts and cases of data-powered innovation in a range as diverse as data platforms, data sharing and collaboration, AI to augment ourselves in our day-to-day work, and Data and AI for Good.

The latter area never fails to excite me, as it shows there is so much more to achieve with data than just improving organizational performance. We have a mission to create better futures in a better society, with data, analytics, and AI – and I am proud to be in the very middle of it with our Insights & Data global practice.

Having said that – to paraphrase one of the wisest men of modern times – we need to be the change we wish to see in the world. Technology can create brilliant solutions to battle the effects of climate change but that technology itself can consume a great deal of energy as well. Only think what it takes to perform one training cycle for the magnificent AI language transformer models of today (hint: really a lot), and you realize that innovation can come with a high environmental price.

We need to be frugal with the resources that we have, in terms of natural resources, IT assets and – of course – scarce human talent. We might want to buy off the shelf rather than build it ourselves, upcycle existing technology rather than replace it, and share data with others rather than keeping it to ourselves. You’ll hear much more from Capgemini in the

forthcoming months about this crucial topic.

For now, whether you are enjoying summer in Australia or a winter wonderland in Norway – or anything in between – I hope this magazine inspires and amazes you.



EDITOR'S NOTE



RON TOLIDO
CTO, Insights & Data Global
Business Line, Capgemini

WHETHER SERVED
CHILLED OR BOILING
HOT: ENJOY THE
WATERS OF DATA-
POWERED
INNOVATION!

Roll up your sleeves! It's time to get busy with fresh, innovative technologies – and really do something about the societal and business challenges we face. Data, analytics, and AI are at the very center of making innovation real. And we are confident the third edition of the *Data-powered Innovation Review* shows you why this is the case in all sorts of different ways.

We have 15 inspiring articles for you, divided into three topic areas.

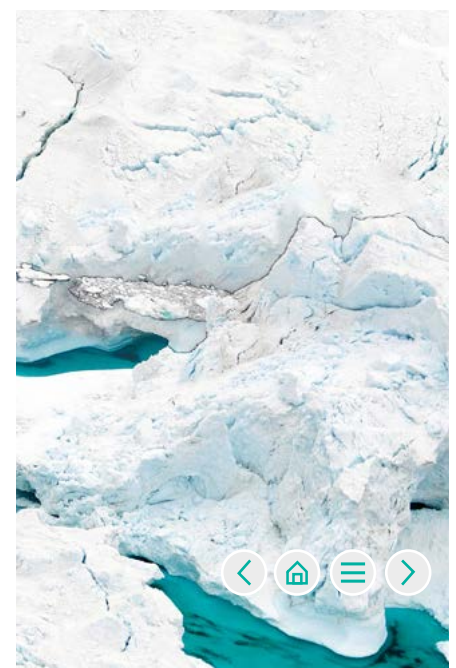
The rise of data collaboration zooms in on that extra-special moment when data tends to become much more valuable: when it is exchanged and shared with others and organizations collaborate closely on data to achieve their objectives. It also comes with a new mindset, as organizations start to realize that data is not “just” an asset but a first-class product that should be owned, seriously managed, and shared by the business domains themselves.

The rise of transformative AI explores many ways in which artificial intelligence augments us in our daily work lives. This is not only crucial from a cost-effectiveness perspective. Scarcity of human talent will be a determining factor for economic success or failure in 2022, and we need all the help we can get from AI and intelligent automation to do the heavy lifting for us. This will transform our perspective on business operations and, just as much, on the way we provide IT and data services.

The rise of AI for good brings back our favorite, recurring topic: the use of data and artificial intelligence to create better, more sustainable futures. By now, many organizations

see the tremendous potential of data and algorithms in battling the impact of climate change. But AI can change personal lives as well, as is evident from several compelling articles about the application of AI to improve health and to achieve real, demonstrable inclusion.

The articles have again been crafted by leading Capgemini experts and several also by – or in collaboration with – key technology partners such as Google, Snowflake, Informatica, Altair, and Zelros. Don't hesitate to contact the contributors. They'll be most happy to help you on your innovation journey. Also, look out for a series of follow-up activities, such as additional in-depth articles and live events featuring contributors and guest panelists.



THE RISE OF DATA COLLABORATION



DATA MESH: USING PRODUCT THINKING TO UNLOCK VALUE IN DATA | 06

Yashowardhan Sowale
Marius Nilsen

CREATING A DATA-POWERED CULTURE | 09

Dinand Tinholt

HOW CROSS-INDUSTRY DATA COLLABORATION POWERS INNOVATION | 12

Eve Besant

USING AI AUGMENTATION TO EMPOWER DATABASE ADMINISTRATORS | 16

Arvind Rao

UNLOCKING THE POWER OF AI WITH DATA MANAGEMENT | 19

Jitesh S. Ghai





DATA MESH: USING PRODUCT THINKING TO UNLOCK VALUE IN DATA

**YASHOWARDHAN
SOWALE**

VP, Enterprise Architecture,
Capgemini Insights and Data, India

MARIUS NILSEN

Architect, Capgemini Norway

It's the riddle of our times: enabling all domains of an enterprise to leverage data to optimize services without creating an unmanageably complex data platform. Solving it means outperforming the competition. An answer worth trying is to use product thinking, creating an ecosystem of data products in which owners and creators of data products act with autonomy. This Data Mesh is based on a set of principles for enabling such autonomy without creating data silos.

Organizations now have access to internal and external data, computing power, and the bright minds who can extract insights out of that. However, bringing quality insights at scale and shoring up trust in data does not happen on its own. Many have tried to gather data in a centralized platform to find a working solution, but as this

scales it becomes complex and hard to manage. Scaling here is about adding a variety of sources and bringing value to additional domains of data consumers. Building an ecosystem of data products solves the challenge of scaling and furthermore enables ownership of analytical data products close to the people who will benefit from it.

To deliver on the grand expectations of what analytical data can bring to an organization, we look at the concepts of product thinking and Data Mesh.

Product thinking is about treating the users of analytical data and insights as customers (business users or fellow developers), being emphatic about their point of view, and designing products that create a insightful experience by using a dataset, dashboard, report, or any other analytical product. The principles of Data Mesh ensure we can create an ecosystem of business managed data products as a lever to optimize organizations and a technology to scale for trusted insights and form data marketplaces.

Core principles

There are four principles to the Data Mesh which balance each other for scaled insights and agility, even as a variety of sources and consumers are introduced and continual changes are needed in a complex business landscape. These are as follows.

- **Domain-oriented ownership**
Decompose ownership of analytical data in the same manner as the business decomposes its functions and areas. Enable agility by reducing the need for coordination with a centralized data owner.
- **Federated computational data governance**
This is a data-governance model based on federated decision-making and accountability. That is, some things must be agreed globally to get the added value of interconnecting data products and domains, however we must enable engagement and agility from different domains without having to synchronize with everyone.
- **Self-serve data platform**
Provide infrastructure with the correct policies needed to create data products (without having to be an infrastructure expert).
- **Data as a product**
One can think of the data product's output as an analytical dataset. The data product introduces a new unit of logical architecture – containing data, code, policies, and infrastructure needed to share the data as a

product in the ecosystem. The products belong to a domain, they have sources, and they have creators and consumers.

To be a valid data product it needs to be valuable, usable, and feasible. Value is decided by the domain owning the product and the characteristics it uses to measure success (e.g., increased efficiency, number of users). For it to be feasible means the product can be created with a self-serve data platform. For it to be usable means it is discoverable, understandable, addressable, secure, interoperable, trustworthy, accessible, and valuable on its own.

These characteristics of the data product put demands on the platform where they are in operation, on the data product creators, and on the resources governing the data products.

Recommended best practices for setting up an ecosystem of data products

Domain-oriented ownership

Business domain users will be responsible for validating trusted and authoritative sources of data, defining and approving metadata, and reviewing data contracts and service contracts.

Federated data governance

- As data moves through its lifecycle, data-catalogue integration into ingestion/preparation/provisioning (for specific use-case) is the responsibility of the enterprise data platform.
- To enable proper classification and cataloguing of data and products, it is important to employ early data discovery/profiling during the discovery phase and data-based hypothesis building. Data architects also need to design for personal data regionalization for country specific restrictions.



THE JOURNEY TOWARDS A VIBRANT DATA ECOSYSTEM MUST BE A SYSTEMATIC JOURNEY FROM AN INTERNAL MARKETPLACE, TO DATA MONETIZATION, TO EXTERNAL MARKETPLACES."

Self-serve data platform

Provisioning of data is the functional responsibility of data requesters, but it is critical that organizations avoid unnecessary data copies and take transformation closer to the data.

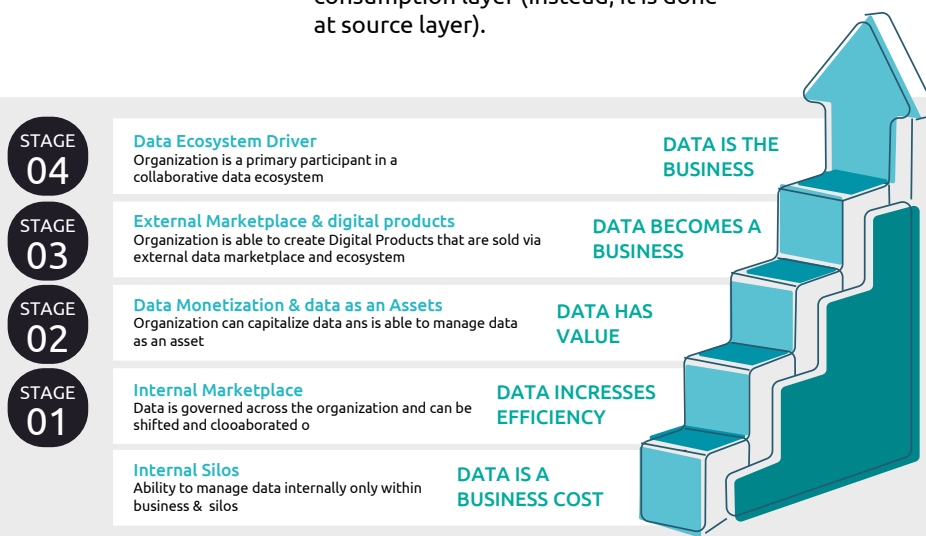
The platform team also needs to ensure data availability for quick access (fast and slow paths) and shareability (via API/microservices/query/files/virtualization, with appropriate security) and enforce technical data quality based on an ingestion layer data-quality engine.

Data as a product

- Data products are key pieces of the data ecosystem puzzle. Architects need to template creation of data-product blueprints for easy creation and to make product easily addressable via APIs. To automate data product maintenance and to support a highly agile data culture, architects should utilize “DataOps” pipelines for creation of new products and enable formalized data products.
- For all products, it is important that data quality initiatives are handled at the source and “data surgery” should not be performed at the consumption layer (instead, it is done at source layer).

- An automated Audit Balance Control and Reconciliation framework (ABCR) is a critical backbone of data orchestration and lifecycle, and it must be employed for reconciliation purposes as well. ABCR needs to cover count, value, volume, and KPI reconciliations.

The journey towards a vibrant data ecosystem must be a systematic one, from an internal marketplace to data monetization to external marketplaces, as shown in the picture.



#DATA #DATAMESH #DATAPOWERED #DATAECOSYSTEM

INNOVATION TAKEAWAYS

DOMAIN-ORIENTED OWNERSHIP

Shifting the ownership of, but also responsibilities for, data products to the business domains that are closest to it.

FEDERATED GOVERNANCE

Properly balancing local ownership and responsibilities with enterprise-scale central platform services.

SELF SERVICE

Enabling domain users to access, use, and provide data without the need for specialized integration and engineering skills.

PRODUCT THINKING

Applying product management to data, making and managing it as a first-class product, like any other product or service from the business.



CREATING A DATA-POWERED CULTURE

DINAND TINHOLT
VP Insights & Data, CPRD Lead
North America, Capgemini

Becoming a data-driven organization isn't as much about data or technology as it is about culture. An organizational culture is like different layers of an onion, with values at its core and practices permeating each layer. To become a data-driven organization, it is important to drive advocacy and adoption of data-driven decision-making and to ingrain data habits in every part of the business.



DRIVING A DATA-DRIVEN CULTURE WITHIN AN ORGANIZATION STARTS WITH THE VALUES."

Data provides answers, but people drive change

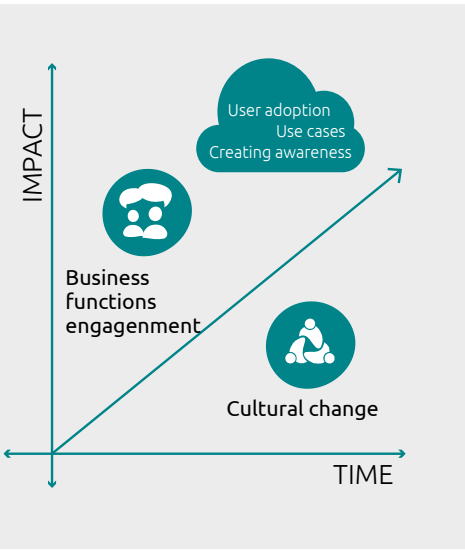
"Culture eats strategy for breakfast." The famous quote by Peter Drucker is a perfect call out here as no matter how detailed and solid your data-powered vision and strategy are, if the people executing it don't nurture the data culture then your journey is likely to fail. Becoming a data-driven organization isn't just about data or technology, it is about transforming the way decisions are made based on deep analysis of facts rather than intuitions and emotions.

Organizations are made up of individuals and the people working there determine the success of a data-driven transformation. Therefore, any transformation journey needs to have organizational culture at the root of any change it wants to affect.

The Capgemini Research Institute report on [The data-powered enterprise](#) found that a majority (75%) of data masters invest in a collaborative and innovation-driven data culture building a data-first culture."

Companies that follow a data-powered culture not only stay resilient but thrive in disruptions. Data forms the lynchpin of their flywheel model of operations that drives customer centricity, innovation, and adoption of advanced technologies. Data is ingrained in their DNA, guiding all decision making and helping them be nimble, agile, and able to adapt to adversities.

But how do you successfully do that?



To build a data-first culture, we have adopted a structured approach of advocacy and adoption

In our AI & Data Activate programs for clients across the globe and industries, there is usually a technology stream to modernize a data platform, a business stream to implement use cases, an organizational/process stream to streamline data governance, and a change-management stream to support the people dimension. It is about winning the hearts and minds of all stakeholders within (and often even outside) organizations to create and/or strengthen a data-driven mindset.

[Hofstede depicts an organizational culture](#) like different layers of an onion, with values at its core and practices permeating each layer.

Driving a data-driven culture within an organization starts with the values. This means that data-driven decision-making within an organization should

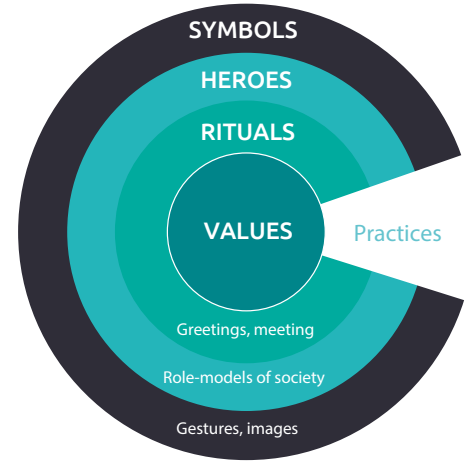
be seen as the default. This requires commitment at all levels in the organization, amplified through communication, training, and management attention. At a global consumer-products company that we worked with for many years, executives at the CxO level would not even consider proposals not underpinned by a thorough data-driven assessment, and they considered their analytical prowess a unique differentiator in the market.

After values, it is important to embed a data-driven culture in rituals which are mostly processes, meetings, and ways of working. The next layer is heroes, or having champions within an organization that have adopted the new data-driven way of working. They are the key advocates that exemplify the new way of working through leading by example. These examples are supported by success stories. Hence, PR and communication play a crucial role to amplify these stories. At the outer layer of culture are symbols. Marketing and expanding the reach of a data-driven transformation means creating brand awareness within an organization, through PR management and publishing assets, images, and stories.

Finally, every layer of the culture model is permeated by practices or, simply put, what you do. Without a strong data-powered culture, evidence-based decision making gets relegated to only a few areas of operations, and organizations fall back on tried-and-tested strategies for important decision making.

The power of habit: Cultivating data-driven behaviors across the enterprise

For the best data-powered organizations, data has become a habit. In "The Power of Habit," Charles Duhigg writes about why companies and people do what they do. He looks at habits from a scientific perspective and identifies a three-step habit loop with Cues, Routines, and Rewards. Changing a habit to support culture change often focuses on changing the routines ingrained within an



organization to promote a data-driven culture. Instilling new data-driven habits means creating new cues with new routines and associated rewards. To make data the cornerstone of their organization, companies need to invest across four operational pillars: people, platforms, partners, and processes. A relationship that is less transactional and more strategic between business groups and IT and BI teams would enable data to permeate the organization and become an enterprise-wide priority.

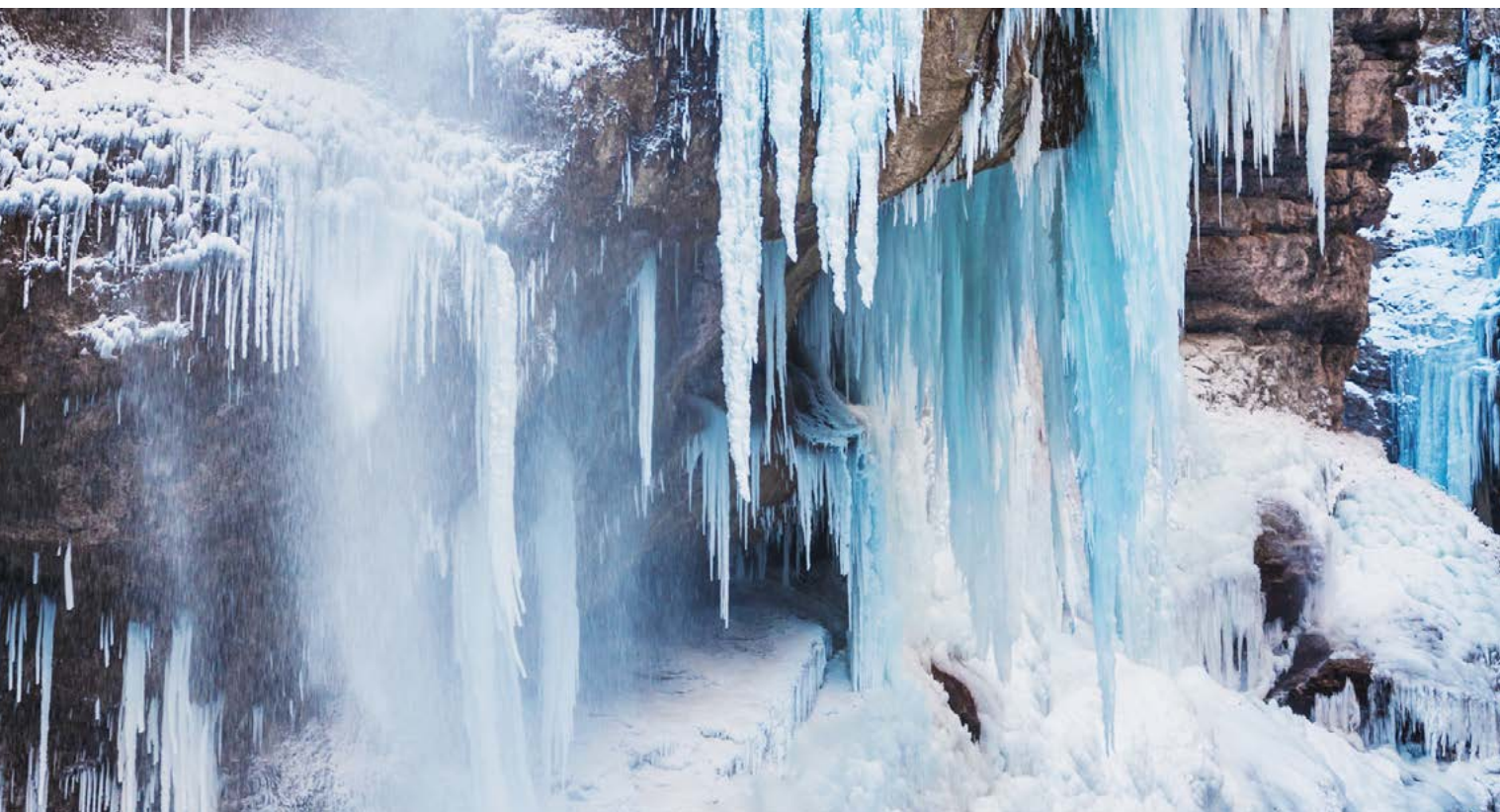
Finally, when making any change in an organization, make sure user centricity is at the core of everything you do. Every design of a new tool or process needs to be ruthlessly user centric. We're all marketers and behavioral scientists. If you desire to augment decision-making with data, cognition is key. Less is more, with simplicity being the ultimate sophistication – to quote Leonardo Da Vinci.

Data success

An international consumer goods manufacturer inculcated data-powered thinking and used meaningful and relevant data from across the organization to connect closely with its one billion customers. A data incubator was set up and launched to initiate a transformation that put information and insights at the heart of all decision making. The needs of the customers were defined. The insights and

analytics supported the human decisions to improve desired outcomes, leading to demonstrable business benefits by surfacing opportunity, supporting human creativity, and increasing penetration, effectiveness, and revenues. These success stories were actively shared, champions were promoted, stories were told, and new ways of working were designed with a pure user-centric lens. Becoming more data-driven became powerful, simple to do, and a daily habit.

While every company aspires to utilize data to make better decisions consistently, many fall short due to old habits or not having a clear approach. To infuse data-powered culture in an organization's DNA, business leaders must take a step-by-step approach to win the hearts and minds of everyone.



#DATAPOWERED #DATACULTURE #USERCENTRIC #DATAHABITS

INNOVATION TAKEAWAYS

VALUES ARE AT THE CORE

Data-driven decision-making within an organization should be seen as the default, accepting nothing less through all the ranks.

WE COULD BE HEROES

Recognize, profile, and support cultural role models that show a data-powered mindset in their daily work.

SYMBOLS LEAD THE WAY

PR management and publishing assets, images, and stories all help to brand and market a data-powered culture.



HOW CROSS-INDUSTRY DATA COLLABORATION POWERS INNOVATION

EVE BESANT

SVP, Worldwide Sales Engineering,
Snowflake



WE'VE SEEN AN INCREASE IN THE NUMBER OF CUSTOMERS WHO WANT TO COLLABORATE ON DATA FROM OTHER INDUSTRIES TO SPUR NEW IDEAS."

Cross-industry data collaboration increasingly fuels innovation across many industries. To unlock the potential of innovation through mass data collaboration, however, organizations must make sure their data management and sharing capabilities are up to date. A robust modern data platform, such as Snowflake's Data Cloud, can go a long way.

Innovation doesn't happen in a vacuum. The development of new products, services, and solutions involves input and information from a multitude of sources. Increasingly, many of these sources are not only beyond an organization's borders but also beyond the organization's industry. According to a [2020 research paper on cross-sector partnerships](#), "Cross-industry innovation is becoming more relevant for firms, as this approach often results in radical innovations."

But developing innovations through cross-industry partnerships must involve coordinated data collaboration. "Firms can only benefit from cross-industry innovation if they are open to external knowledge sources and understand how to explore, transform, and exploit cross-industry knowledge," the paper's authors noted. "Firms must establish certain structures and processes to facilitate and operationalize organizational learning across industry boundaries."

Examples of cross-industry data collaboration

There are a multitude of examples of how organizations across industries have spurred innovation through collaboration.

- **In financial** services, institutions that must prevent and detect fraud use cross-industry data sharing to better understand the profile of fraudsters and fraudulent transaction patterns.
- **In manufacturing**, companies are using AI to manage supply-chain disruptions. Using data from external sources on weather, strikes, civil unrest, and other factors, they can acquire a full view of supply-chain issues to mitigate risks early.
- **In energy**, smart meters in individual homes open new doors for data collaboration, transmitting information about energy consumption.
- **In education**, school systems, local governments, businesses, and community organizations work

together to improve educational outcomes for students.

- **In healthcare**, during the COVID-19 pandemic, hospitals relied on information from health agencies and drug companies regarding the progression and transmission behavior of diseases. Governments followed data from scientists and healthcare professionals to create guidance for the public. Retailers heeded guidance from the public and healthcare sectors to create new in-store policies and shift much of their business online.

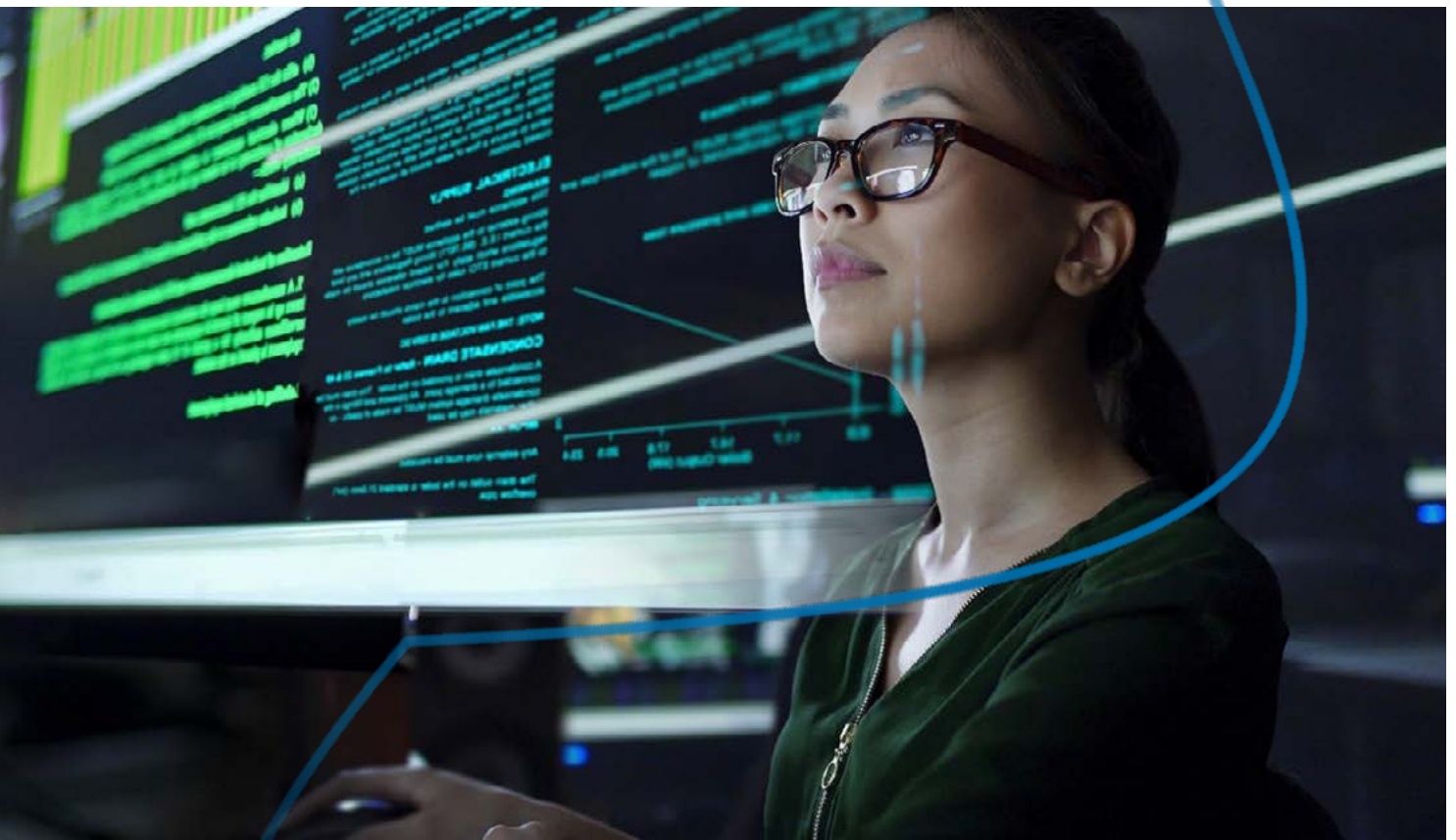
The role of cross-industry data collaboration in innovation during the pandemic is perhaps nowhere better exemplified than in the [COVID-19 Research Database](#), involving a cross-industry consortium of organizations. The database, which can be accessed by academic, scientific, and medical researchers, holds billions of de-identified records including unique patient claims data, electronic health records, and mortality data. This has enabled academic researchers in medical and scientific fields as well as

public health and policy researchers to use real-world data to combat the COVID-19 pandemic in novel ways.

Best practices for cross-industry collaboration

As the examples above show, organizations that have developed cross-industry data collaboration capabilities can more easily foster innovation, leading to a competitive advantage. Here are some of the considerations and best practices that enable sharing and collaborating on knowledge across industries.

- **A single, governed source for all data:** Each industry – and indeed, each company – stores and formats its data in different ways and places. Housing data in one governed location makes it easier to gather, organize, and share semi-structured and structured data easily and securely.



- **Simplified data sharing:** The relevant data must be easily accessible and shareable by all partners. Data is stored in different formats and types, and it can be structured, semi-structured, or unstructured. It can be siloed in specific departments and difficult or slow to move, or inaccessible to the outside world. What processes and tools are in place to transform cross-industry knowledge into a shareable, usable format?
- **Secure data sharing:** Data privacy is of the utmost importance in today's society. Data must be shareable securely and in compliance with privacy regulations. Cross-industry data sharing often involves copying and moving data, which immediately opens up security risks. There may also be different data protection and privacy regulations in different industries.
- **Inexpensive data management:** Data must be shareable, and budgets kept in mind. Centralizing, organizing, securing, and sharing data is often resource-intensive, so organizations need to find ways to manage and share their data more efficiently.
- **Democratized data:** While data security and privacy are paramount, companies must "democratize" data so that it is accessible and shareable in a way that allows non-technical users in both internal and external parties to use it easily.
- **Advanced analytics:** Technologies such as AI and machine learning can help companies glean deeper insights from data. This requires a data foundation and tools that can analyze all types of data.

Technological tools are making it easier for organizations to follow and gain ROI from these best practices. For example, [Snowflake's Data Cloud](#) enables the seamless mobilization of data across public clouds and regions, empowering organizations to share live, governed, structured, semi-structured, and unstructured data (in public preview) externally without the need for copying or moving. Snowflake enables compliance with government and industry regulations, and organizations can store near-unlimited amounts of data and process it with exceptional performance using a "pay only for what you use" model. They can also use Snowflake's robust partner ecosystem to analyze the data for deeper insights and augment their analysis with external data sets.

"We've seen an increase in the number of customers who want to collaborate on data from other industries to spur new ideas," Snowflake's Co-Founder and President of Products Benoit Dageville said, "to foster innovation, to be able to freely collaborate within and outside of their organization, without added complexity or cost."



The future of mass collaboration

In the future, cross-sector data collaboration will only play a larger role in innovation as technology becomes more ubiquitous and the public grows more comfortable with sharing data. We could see worldwide consortiums that collaborate on data to solve some of humanity's biggest problems: utilizing medical and scientific information to tackle global health crises, enabling more-efficient use of resources to fight poverty and climate change, and combating misinformation. Organizations such as the World Bank are already working on such initiatives. Its Data Innovation Fund is working to help countries benefit from new tools and approaches to produce, manage, and use data. According to a recent [World Bank blog post](#), "Collaboration between private organizations and government entities is both possible and critical for data innovation. National and international organizations must adopt innovative technologies in their statistical processes to stay current and meet the challenges ahead."

To unlock the potential of innovation through data collaboration, organizations must make sure their data management and sharing capabilities are up to date. A robust, modern data platform can go a long way. But what's also needed is an audit of internal processes and tools to ensure that barriers to data sharing and analysis are not impeding innovation and growth.

#DATACOLLABORATION #DATASHARING #INNOVATIONTHROUGHCOLLABORATION
#CROSSINDUSTRYINNOVATION #CROSSECTORINNOVATION
#CROSSINDUSTRYCOLLABORATION #CROSSECTORCOLLABORATION

INNOVATION TAKEAWAYS

COLLABORATION NEEDS BEST PRACTICES

Organizations that implement best practices in cross-industry data collaboration can foster innovation, leading to a competitive advantage.

DATA CAPABILITIES MUST BE UP TO DATE

Organizations must make sure their data management and sharing capabilities are current, to unlock the potential of innovation through data collaboration.

TECHNOLOGY AND PLATFORMS TO THE RESCUE

Dedicated tools and data platforms make it easier for organizations to gain cross-sector data-collaboration capabilities much quicker.



USING AI AUGMENTATION TO EMPOWER DATABASE ADMINISTRATORS

As data platforms rapidly evolve and become more powerful, DBAs are the important link between data scientists and business users.

ARVIND RAO
Partner Architect Advisor, Google



ORGANIZATIONS UNDERSTAND IT'S IMPERATIVE TO MODERNIZE THEIR DATA AND HARNESS THE FULL POWER OF AI WHILE HUMAN RESOURCES POSE A MAJOR CHALLENGE.”

Most enterprises already have the talent in-house to start using AI to unlock the full potential of their data. They are the database administrators. They know the data, they know the organization, and they are trusted advisors – they just need a little help from data-platform vendors.

The world's largest organizations generally understand that to continue to succeed in today's competitive environment, they need to become data-powered enterprises. They acknowledge that it's imperative to modernize their data and harness the full power of tools such as AI to derive actionable insights. However, many of these companies have also learned that human resources are a major challenge in making this transformation.

In short, there are not enough data scientists – those who create the solutions that leverage state-of-the-art technologies such as AI. Based on my experience in data analytics over the past couple of decades, in an ideal world data scientists would account for 10 to 15 percent of the staff at a data-powered organization. Yet the majority of organizations – including most successful technology enterprises – has not achieved that ideal/goal.

DBAs to the rescue

The good news is most enterprises already have the talent to successfully make this transformation.

Database administrators (DBAs) – those who manage a company’s data warehouses and similar data platforms – are the backbone of most IT operations. These professionals understand the data an enterprise has collected, where it’s stored, and how to use it. They ensure authorized people have access to the data they need. And since data is sensitive and valuable, they control who has access to it to keep it safe from misuse or theft.

Knowledge and trust

As a result, database administrators know more about their company’s data than anyone else in their organization. They certainly know more than the data scientists who work for the technology vendors that develop the data platforms upon which modern enterprises rely.

At the same time, database administrators are trusted advisors within their enterprise. They’re the go-to source for help when a business user needs to derive insights – whether that’s a salesperson looking to improve lead generation, a service manager trying to spot potential customer-satisfaction issues, or an executive seeking market predictions for the coming year.

It therefore makes sense to ensure database administrators can leverage the insights and capabilities of AI-augmented data platforms.

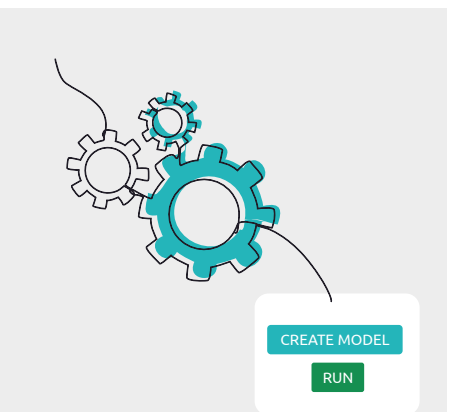
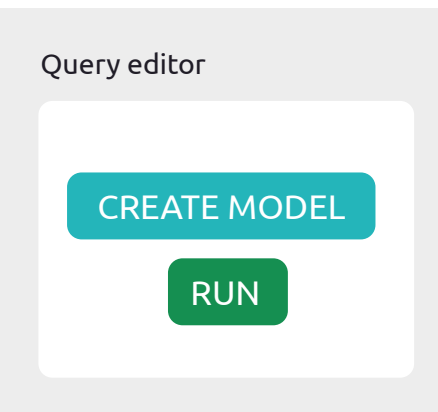
The Lake House

The majority of data-platform vendors have been working towards the concept of a Lake House – a convergence of databases, data warehouses, and data lakes – that

makes the platform usable and accessible to everyone and everywhere. With data scientists increasingly focused on creating these new platforms, vendors have fewer resources to dedicate to building, managing, and maintaining the – often highly customized – tools required by business users.

That’s why it’s important that data-platform vendors augment their solutions with AI. It’s also why these AI augmentations must be easy to use in the DBA’s day-to-day role: They should not have to invest huge amounts of time learning data science to take advantage of these tools.

Enterprises are increasingly demanding this simplicity of their suppliers – whether they are vendors of databases, data platforms, analytics, or cloud-based solutions.



One should not worry about what model to use, or how to create the tools required.

At Google, we’ve developed a number of solutions that help bridge the gap and create data warehouses infused with AI/ML that work for all users – not just data scientists.

- Vertex AI brings together Google Cloud services for building Machine Learning in a unified user interface and API. With Vertex AI, a database administrator can easily train and compare models [using AutoML](#) or custom code training. All models are stored in one central repository and can be deployed in ways that allow DBAs and other non-data scientists to start using AI/ML in their day-to-day work, with very little training.

Model creation should be as simple as clicking “Create model.”

- Dataplex is an intelligent data fabric that breaks apart silos. It provides a single pane of glass that allows database administrators to centrally manage, monitor, and govern an organization’s data – including ingestion, storage, analytics, AI/ML, and reporting. It does this across any type of platform – including data lakes, data warehouses, and data marts – with consistent controls that provide access to trusted data and power analytics at scale.
- BigQuery is a serverless, cost-effective, multi-cloud data warehouse designed for business agility. BigQuery democratizes

DBAs should be able to predict and analyze using their dataset from a single tool set.

insights with a secure and scalable platform to perform functions such as anomaly detection, customer segmentation, product recommendation, and predictive forecasting. It features built-in Machine Learning to derive business insights using a flexible, multi-cloud analytics solution and adapts to data at any scale from bytes to petabytes with zero operational overhead. Most importantly, database administrators can learn BigQuery and easily incorporate it into their tasks.

The smart data-warehouse platform

Looking ahead, I envision a future in which most successful organizations deploy a smart data-warehouse platform that provides a number of important benefits. These include:

- Easy access to the organization’s data, public data, and other business data – without worrying about what kind it is or where it’s stored
- Serverless tools to access data in real time, to mine and infuse AI/ML capabilities. These would be scalable on demand, set a strong foundation for building AI models, and be cost effective.
- Reporting tools that showcase analytics in real time – in a safe, secure, and scalable way
- Modern data warehouse capabilities that equip all users with the tools and resources they need to do their jobs efficiently and effectively, and that provide CXOs with the tools they need to keep their staff motivated.

As enterprises work to achieve this goal, leveraging AI to empower database administrators in their day-to-day work is something they can do now, and do cost effectively. They just need the right tools from their vendors.

Giving DBAs easy-to-learn AI-powered tools will enhance the value they already provide to the enterprise. It can also help keep these knowledgeable team members – the organization’s trusted advisors on all matters IT-related – relevant as the enterprise embraces a new, more powerful, and innovative data-powered future.

#DATALAKEHOUSE #AIMLFORDBA #FUTUREDBA #AIMLFORALL #DATAFORALL

INNOVATION TAKEAWAYS

A VALUABLE RESOURCE

Database administrators know the company’s data and are trusted by its people. They have important roles to play in an organization’s transformation into a data-powered enterprise.

SHARE THE LOAD

Database administrators bridge the gap between the data scientists who are creating the next generation of AI-powered analytics tools and the business users who will benefit from the insights such tools provide.

VENDORS MUST HELP

Data-platform vendors must incorporate easy-to-learn AI tools into their products so database administrators can take full advantage of these state-of-the-art solutions



UNLOCKING THE POWER OF AI WITH DATA MANAGEMENT

JITESH S. GHAI
Chief Product Officer, Informatica



AI IS MOST EFFECTIVE WHEN YOU THINK ABOUT HOW IT CAN HELP YOU ACCELERATE END-TO-END PROCESSES ACROSS YOUR ENTIRE DATA ENVIRONMENT."

Artificial intelligence is crucial to innovation and business growth in today's digital world but, without data management, AI can be a black box that has unintended consequences.

In today's data-driven economy, artificial intelligence (AI) and machine learning (ML) are powering digital transformation in every industry around the world. According to a [2021 World Economic Forum report](#), more than 80 percent of CEOs say the pandemic has accelerated digital transformation.

AI is top of mind for boardroom executives as a strategy to transform their businesses. AI and ML are critical to discovering new therapies in life sciences, reducing fraud and risk in financial services, and delivering personalized digital healthcare experiences, to name just a few examples that have helped the world as it emerges from the pandemic.

For business leaders, AI and ML may seem a bit like magic: its potential impact is clear but they may not quite understand how best to wield these powerful innovations. AI and ML are the underpinning technology for many new business solutions, be it for next-best actions, improved customer experience, efficient operations, or innovative products.

Machine learning in general, and especially deep learning, is data hungry. For effective AI, we need to tap into a wide variety of data from inside and outside the organization. Doing AI and ML right requires answers to the following questions:

- Is the data being used to train the model coming from the right systems?
- Have we removed personally identifiable information and adhered to all regulations?
- Are we transparent, and can we prove the lineage of the data that the model is using?
- Can we document and be ready to show regulators or investigators that there is no bias in the data?

The answers require a foundation of intelligent data management. Without it, AI can be a black box that has unintended consequences

AI needs data management

The success of AI is dependent on the effectiveness of the models designed by data scientists to train and scale it. And the success of those models is dependent on the availability of trusted and timely data. If data is missing, incomplete, or inaccurate, the model's behavior will be adversely affected during both training and deployment, which could lead to incorrect or biased predictions and reduce the value of the entire effort.

AI also needs intelligent data management to quickly find all the features for the model; transform and prepare data to meet the needs of the AI model (feature scaling, standardization, etc.); deduplicate data and provide trusted master data about customers, patients, partners, and products; and provide end-to-end lineage of the data, including within the model and its operations.

Data management needs AI

AI and ML play a critical role in scaling the practices of data management. Due to the massive volumes of data needed for digital transformation, organizations must discover and catalog their critical data and metadata to certify the relevance, value, and security – and to ensure transparency. They must also cleanse and master this data. If data is not processed and made usable and trustworthy while adhering to governance policies, AI and ML models will deliver untrustworthy insights.

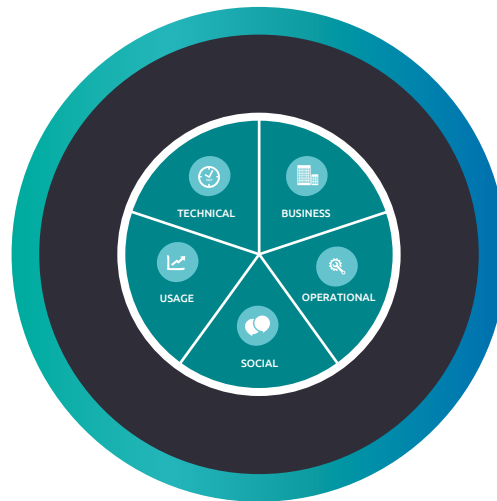


Don't take a linear approach to an exponential challenge

Traditional approaches to data management are inefficient. Projects are implemented with little end-to-end metadata visibility and limited automation. There is no learning, processing is expensive, and governance and privacy steps can't keep pace with business demands. So how can organizations move at the speed of business, increase operational efficiency, and rapidly innovate?

This is where AI shines. AI can automate and simplify tasks related to data management across discovery, integration, cleansing, governance, and mastering. AI improves data understanding and identifies privacy and quality anomalies.

AI is most effective when you think about how it can help you accelerate end-to-end processes across your entire data environment. That's why we consider AI essential to data management and why Informatica has focused its innovation investments so heavily on the CLAIRE engine, its metadata-driven AI capability. CLAIRE leverages all unified metadata to automate and scale routine data-management and stewardship tasks.



As a case in point, [Banco ABC Brasil](#) struggled to provide timely data for analysis due to slow manual processes. The bank turned to an AI-powered integration Platform-as-a-Service and automated data cataloging and quality to better understand its information using a full business glossary, and to run automated data quality checks to validate the inputs to the data lake. In addition, AI-powered cloud application integration automated Banco ABC Brasil's credit-analysis process.

Together, the automated processes reduced predictive model design and maintenance time by up to 70 percent and sharpened the accuracy of

predictive models and insights with trusted, validated data. They also enabled analysts to build predictive models 50 percent faster, accelerating credit application decisions by 30 percent.

With comprehensive data management, AI and ML models can lead to effective decision-making that drives positive business outcomes. To counter the exponential challenge of ever-growing volumes of data, organizations need automated, metadata-driven data management.

#DATA #DATAMANAGEMENT #AI #AUTOMATION #DIGITALTRANSFORMATION #DIGITALECONOMY

INNOVATION TAKEAWAYS

ACCELERATE ENGINEERING

Data engineers can rapidly deliver trusted data using a recommender system for data integration, which learns from existing mappings.

BOOST EFFICIENCY

AI can proactively flag outlier values and predict issues that may occur if not handled ahead of time.

DETECT RELATIONSHIPS AMONG DATA

AI can detect relationships among data and reconstitute the original entity quickly, as well as identify similar datasets and make recommendations.

AUTOMATE DATA GOVERNANCE

In many cases, AI can automatically link business terms to physical data, minimizing errors and enabling automated data-quality remediation.

THE RISE OF TRANSFORMATIONAL AI

GPT-3 ONE-CLICK CODE TRANSLATION | 23

Anitha Madhusudhan

AI DELIVERS SIMPLER SIMULATIONS FOR A COMPLEX WORLD | 26

Sam Mahalingam

AUGMENTED INSURANCE ADVISORY WITH AI | 29

Damien Philippon
Kiran Boosam

A CASE FOR CONTEXT AWARENESS IN AI | 32

Robert Engels

UNBOX BLACK-BOX MODELS WITH AI GLASS BOX | 35

Jatinder Kumar Kautish
Ashish Dhali





GPT-3 ONE-CLICK CODE TRANSLATION

ANITHA
MADHUSUDHAN
Consultant at AICOE, ID DSA India



THE REAL BEAUTY OF BUILDING A2B ON THE BACK OF OPENAI'S GPT-3 IS THAT IT CAN BE LEVERAGED VIA APIS, SAVING MASSIVE AMOUNTS OF COMPUTING POWER."

Translating legacy code into modern coding languages has been, until now, a time- and resource-intensive endeavor. Our solution leverages OpenAI's GPT-3 to automate code translation, a process that offers limitless market potential.

OpenAI's GPT-3 – one of the largest AI language models in the world – has been used to address some of the most challenging Natural Language Processing (NLP) use cases in business, from content summarization to text generation to code creation.

Founded in 2015 by Elon Musk, Sam Altman, and others, the mission of the OpenAI research lab is to develop AI for the benefit of humanity. Its GPT-3, or third-generation Generative Pre-trained Transformer, uses deep learning and training on 175B parameters and an extremely large open-source language corpus to produce natural language text. This has enabled it to perform some pretty impressive tasks: generating natural dialog for virtual-reality characters, providing lightning-fast semantic search results for help-desk queries, and translating customer feedback from a variety of sources into readable summaries.

GPT-3 has literally changed the game for solving NLP tasks. And we have now taken it a step further.

Using GPT-3's immense open-source language library, Capgemini India's AI Center of Excellence has created a code translation accelerator, dubbed the A2B Translator, that can quickly and efficiently convert code from one language to another with little more than a single click.

Code translation accelerated and automated

As digital transformation becomes ever more critical for business growth, relying on legacy code with limited functionality is simply becoming untenable. But updating code comes with multiple challenges, risks, and costs, leaving organizations in need of resource-efficient technology solutions to migrate their legacy code to a modern coding language.

Translating code has typically been accomplished either manually by a team of developers – a process that can take months or even years – or by older code converters that translate line by line, an equally laborious and often unreliable process.

Our A2B translator changes that by leveraging GPT-3’s vast capacity for natural language processing and layering on prompt engineering processes that enable its application-

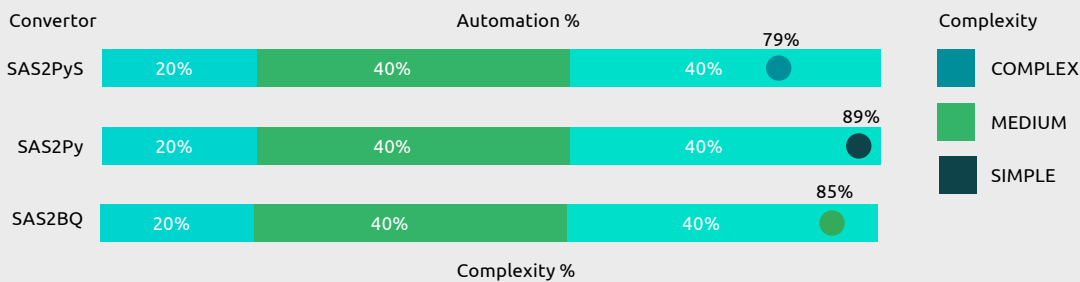
to-code translation. Rather than translating line by line, it extracts semantic meaning in blocks and recreates segments of code based on meaning rather than word-by-word translation.

How the A2B translator works

Our translator tool contains three unique code converters for changing SAS to Python, BigQuery, and PySpark. Using prompt engineering

principles, we give the translator a textual input (i.e.: a prompt) to get the output we want.

Using multiple frameworks and accelerators (the GPT-3 accelerator, as well as our in-house built accelerator), we can achieve near-complete automation of code translation.



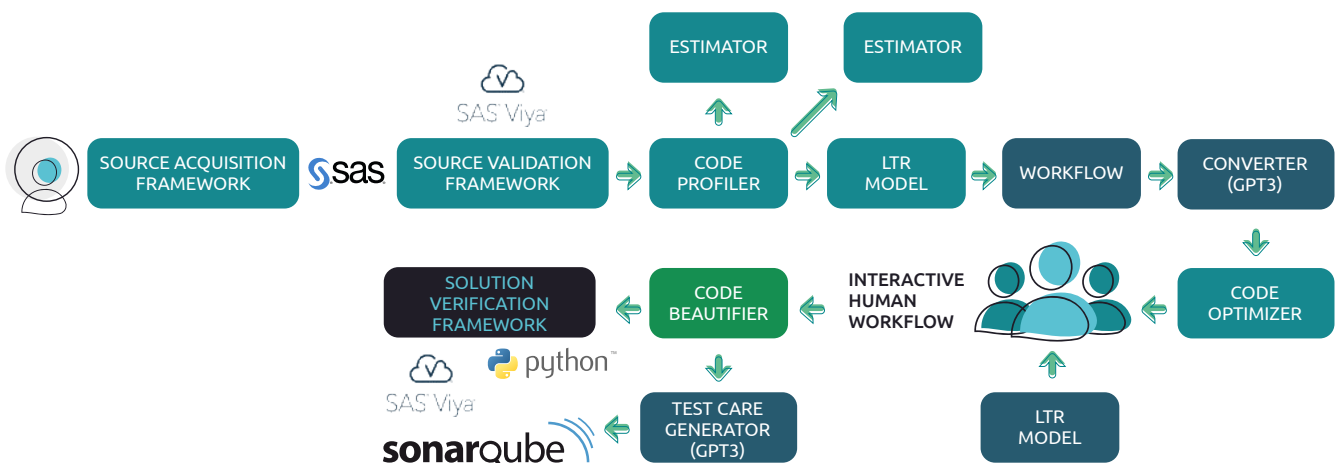
Representation of SAS code is complexity and the possible % automation

We then use human optimizers to beautify the generated code, accounting for about 10 to 20 percent of the translation process. For example, indentations are important for any code to execute successfully in Python, and the indentation usage is

very different in SAS, so indentations must be edited manually after the code conversion. The result is converted code that is more than 85 percent reliable.

But the real beauty of building A2B on

the back of OpenAI’s GPT-3 is that it can be leveraged via APIs, saving massive amounts of computing power. And, as a bonus, it’s a user-friendly tool, so literally anyone can use it to quickly and easily convert legacy SAS code.

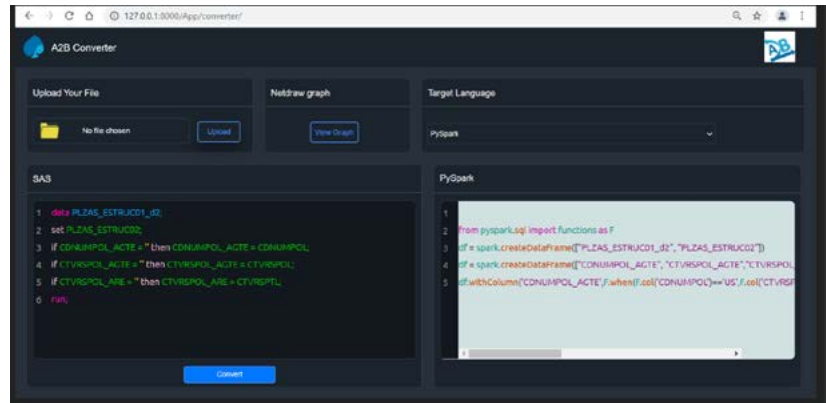


Limitless potential

While A2B can be deployed in any sector currently using SAS code, one industry rife with opportunities for converting SAS code to Python is the medical sector. And as code translation becomes more advanced with new GPT-3 models, the applications will expand to include nearly every industry.

While we are currently working in the realm of procedural language conversions, our plans are to expand A2B to include functional languages and, eventually, to machine learning code conversions as well.

Our ultimate goal? To create a comprehensive package of tools that can translate anything.



#DATAPOWERED #AI4GOOD #GPT3 #LANGUAGEMODELS #SAS
#PYTHON #PYSPARK #BIGDATA #AIACCELERATOR

INNOVATION TAKEAWAYS

RESOURCE SAVINGS

Using AI to automate code translation will save significant time and resources for companies looking to update legacy code with a user-friendly tool.

COMPUTE EFFICIENT

Leveraging OpenAI's GPT-3 for code translation offers the benefit of access via APIs, reducing the required computing power.

ADDITIONAL CONVERTERS

A and B technologies that can be used for code conversion offer enormous potential for language combinations, leaving vast untapped opportunities for building more such converters that can apply to any industry.



AI DELIVERS SIMPLER SIMULATIONS FOR A COMPLEX WORLD

SAM MAHALINGAM
Chief Technology Office, Altair



BY EMPOWERING WIDER DESIGN AND ENGINEERING TEAMS, REDUCED ORDER MODELING CHALLENGES INTERNAL SILOES AND SUPPORTS 'SIMULATION AT THE SPEED OF DESIGN.'

Just add AI: with the help of artificial intelligence (AI) – combined with the concepts of Reduced Order Modeling – a much more efficient approach to simulating system components is within reach. Bringing ease-of-use and democratization of digital twins to more people, it works for even the most complex components. And more opportunities for innovation breakthroughs are around the corner, way beyond the original realm of engineering.

Dynamism, creativity, and innovation characterize the engineering market. Whether it's the decarbonization of power generation, commercialization of space travel, or electrification of personal mobility, the pace of change is remarkable. Inevitably, it also means that systems are becoming more complex.

In response, manufacturers have increasingly embraced the concept of simulation-driven design. Modeling new product concepts in the digital

domain makes it possible to explore more design options earlier in the development process and reduce the need for physical prototyping. But simulation is getting tougher, too. A growing number of components are proving difficult to model. However, by combining AI with another proven concept, physics-based Reduced Order Modeling, a new approach to simulating difficult components is now within reach. As well as speeding product development, AI for Reduced Order Modeling enables a more

sophisticated approach to digital twins. As a result, data from the field will inform continual enhancement in product design. With AI learning new lessons and delivering new insights, design teams will be better equipped to make sense of complexity and push the boundaries of innovation.

Not too hot, not too cold

Reduced Order Modeling is already a popular means of streamlining simulation and sharpening its focus. Essentially, it provides designers with just the right level of model fidelity to inform decision making at a particular point on the development pathway. For example, at the outset, designers may only want to consider a relatively limited number of characteristics. Highly granular detail, and the data and human resources needed to create it, will simply limit their ability to run more of the “what ifs” that should guide them to the best outcome.

Now with added AI

As the name suggests, AI for Reduced Order Modeling adds the self-learning capabilities of AI. To illustrate the benefits, it's useful to start with a use case. In this respect, the battery in an electric vehicle (EV) represents a classic example of a component that is difficult to model. A significant number of tricky variables need to be considered, including the internal chemistry of the battery, how the vehicle is driven, and the effects of aging.

For the system designer, understanding the relationship between factors such as vehicle speed, range, and the battery's state of charge is critical to the accuracy and insight provided by overall system simulation. Crucially, compared to traditional approaches, AI for Reduced Order Modeling needs much less in the way of training data to create a robust model for this type of component. And that basic data can be drawn either from existing simulations or generated with test-bench measurements and by products in the field.

Democratizing simulation

With AI for Reduced Order Modelling, the user's responsibilities are limited to defining the input and output, and the relevant variables and characteristic parameters. Interpreting that information, and generating simulation results, is the job of a self-learning neural network – effectively, a functional building block or box that offers a straightforward modular fit within wider system simulations. The simplicity of this approach also democratizes it. AI for Reduced Order Modeling is not the preserve of simulation specialists. By empowering wider design and engineering teams, it challenges internal siloes and supports “simulation at the speed of design.”

Enhancing digital twins

AI for Reduced Order Modeling can embrace multiple physics. For example, the solution has also been tested for a Computational Fluid Dynamics simulation on a hairdryer, and simulation of the mechanized scooping of bulk material. In all cases, the results mirrored closely those of far more labor-intensive conventional simulations. In the bulk material example, it smoothed out detailed “spikes” in the graph, but this is a phenomenon typical of all Reduced Order Modeling. There is also exciting potential for use in digital twins, where training data can be generated and updated continually by the product in the field. AI for Reduced Order Modeling embedded within the design and development twin effectively locks in a virtuous circle of improvement, powered by the self-learning capabilities of a neural network.

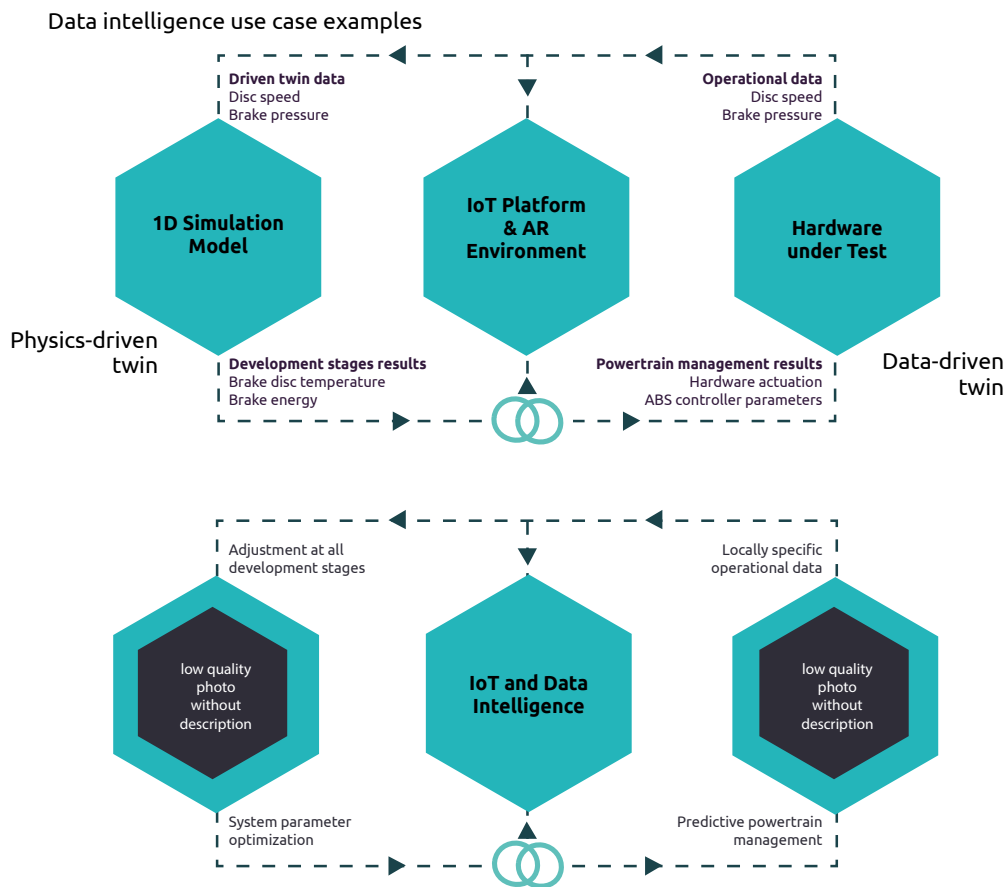
The age of convergence

At this point, market interest in AI for Reduced Order Modeling is strongest in sectors characterized by complex multi-physics simulations. The automotive sector is just one example. In addition to EV batteries, early use cases here are likely to include the optimization of chassis dynamics. But recognition of the benefits of combining data analytics and simulation is quickly gaining wider

traction. Notably, process industries such as steel- and cement-making are beginning to leverage the convergence of data, AI, and simulation to explore the potential and assess the risks of deploying new technologies and alternative fuels.

Evolution to revolution

Broader deployment of AI for Reduced Order Modeling will undoubtedly be assisted by the fact it represents a logical evolution of existing, simulation-driven design methodologies. However, there are genuinely revolutionary opportunities in terms of the value delivered. In the process sector, users are already identifying applications that can rapidly achieve threefold and fourfold returns on investment. As they surf the opportunities of relentless technological progress, enterprises such as these are not only looking to draw on the immense potential offered discretely by AI, data analytics and simulation. Increasingly, they are looking to combine them. In this respect, AI for Reduced Order Modeling represents another compelling example of what this new era in convergence will offer, both in the engineering domain and well beyond.



#DIGITALTWIN #AI #ML #AUTONOMOUSLEARNING #ALTAIRPARTNER

INNOVATION TAKEAWAYS

LESS IS MORE

By enabling robust simulation of challenging components using minimal training data, AI for Reduced Order Modeling accelerates complex system development.

JUST ADD AI

Combining the powerful, self-learning of artificial intelligence with Reduced Order Modeling further improves the effectiveness of modeling, converging data analytics, AI, and simulation.

POWER TO THE PEOPLE

AI-enabled Reduced Order Modeling empowers wider design and engineering teams, challenging internal siloes and supporting "simulation at the speed of design."



AUGMENTED INSURANCE ADVISORY WITH AI

DAMIEN PHILIPPON

Co-founder, COO, Zelros

KIRAN BOOSAM

Vice President, Global Insurance Strategy & Portfolios, Capgemini

Insurance is complex at its core, both in design and operations. It has always been product-centric and paper and workflow intensive, making customer advisory very complex and nuanced. In tandem, consumer buying behaviors have significantly shifted; they now expect intuitive experiences, personalized and in real-time. Complexity breeds opportunity for AI to ensure quality sales interactions and drive disruptive growth, with automation of customer advisory.



THINGS CHANGED WITH A NEW TYPE OF AI OFFERINGS DELIVERED BY PROVIDERS THAT ARE SPECIALIZED TO THE INSURANCE INDUSTRY AND THAT DELIVER OFF-THE-SHELF INSURANCE BUSINESS VALUE."

AI has been on the strategic agenda of insurance players for a couple of years now. Until recently, it has mainly resulted in proof of concepts without substantial ROI. Things changed with a new type of AI offering delivered by providers that are specialized to the insurance industry and that deliver off-the-shelf insurance business value. We see deployments at scale in many different areas of the value chain, moving up towards the consumer perspective. What is the use of automation if it does not give consumers the ability to be fully autonomous when needing to protect themselves, without compromising the quality and the relevance of the advisory involved? According to the

latest [Capgemini World Insurance Report](#), 69 percent of customers prefer buying car insurance online. On the other hand, the online conversion ratio is brutally low, around one percent. Why is that? One reason is that consumers don't receive adequate experiences with the level of advice they are looking for. As a result, consumers lose confidence and comfort in trusting the company to buy such a personal product. To bridge this gap, insurers must use technology to uncomplicate and personalize advisory, leveraging all types of data they have in their hands, and embrace an experience-first approach to digital distribution.

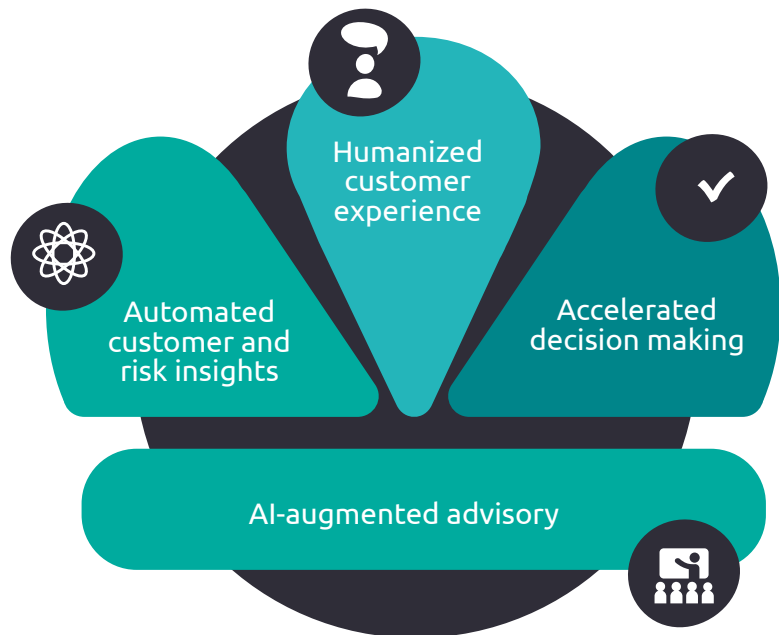
Multiple ways of AI

AI interventions automate data-driven personalized advisory around specific life events relevant to the customer, tailored products, and coverage options with impactful selling insights. It does this in multiple ways.

- Natural Language Processing (NLP) and computer vision aims at extracting insights from unstructured data sources like documents, emails, and voice discussions. Training modern NLP algorithms such as BERT (Bidirectional Encoder Representations from Transformers) on large insurance-specific corpus allows superior understanding of insurance named entities and insights generation.
- Machine learning algorithms learn from the history of data points what is suitable to recommend at the right time with the right message. Once in production, a crucial aspect is to monitor algorithms' potential drifts, to manage their lifecycle.
- Reinforcement Learning algorithms leverage every implicit or explicit feedback from the customer, like clicks, quotes issued, or contracts subscribed, to update in real time the weights of the model and capture new behaviors unseen in the past training data. Several mechanisms reinforce learning algorithms, leveraging already seen consumer behaviors (exploitation), but also investigating new potential combinations (exploration).

Responsible AI

Further, as insurance is a heavily regulated industry, such AI must be developed within a [Responsible AI framework](#) (ideally validated by financial regulators). This Responsible AI framework must bring state-of-the-art capabilities to work around the black-box issue, leveraging libraries like SHAP (features importance), LIME (local explainability), and DiCE



(counterfactual explainability). It also must pay attention to inherent biases and non-discrimination issues by defining “protected” features (like age or gender) and check a couple of critical points (equality of prediction distribution, equality of performance, robustness to changes) for each of them. And it must raise warnings when it sees outliers at the prediction time when the prediction context is too different from the training context to avoid irrelevant recommendations.

Leading bancassurance and insurance companies like, BPCE, Crédit Agricole, and Matmut have relied on AI to improve the experience they deliver to their policyholders and delight customers. They leverage Zelros technology in various ways:

- to drive more traffic to their self-care portal and their insurance funnel
- to increase lead conversion from quote to contract
- to increase upsell and cross-sell
- to automate underwriting decisions.

For any insurance player, it is now imperative to provide rich engagement to customers by empowering integrated physical and digital channels. Setting up a powerful and competitive digital sales channel is critical in the new normal, unless they want to see higher churn in their policyholders. Data and AI must be employed to turn customer portals, mobile apps, and agency channels into a relevant experience that meets consumer expectations. And this is only the beginning, as we see third-party data enrichment as another opportunity to elevate mass hyper-personalization. This kind of AI, coupled with new types of data like weather, points of interest, statistics, and telematics, mean consumers will see their insurance partner being one step ahead of their needs, offering on-demand advance digital/robo advisory across channels. With self-discovery-type experience design, consumers won't even sometimes see it anymore; insurance policies will be based on a tailored product, because embedded AI understands the customer need.



Underwriting and actuarial	Claims process	Policy servicing	Sales and marketing	Operations: HR, finance, IT, call centre and others
Renewal reviews, pricing and projections	Fraud analytics	Effect of premium rate change on policy holders	Campaign analytics	Call center analytics
Price optimization	Claims processing analytics	Premium leakage analytics	Cross-sell/upsell	Reputation intelligence
Predictive underwriting	Geospatial analytics	Churn prediction	Agent performance / advisory analytics	Knowledge management
Risk segmentation for renewal underwriting	Recovery prediction models	Lifetime value analysis	Churn / persistence propensity	Sentiment analytics
Portfolio analysis	Continuous monitoring for claim assessment	Exploratory analytics on cancellations and refunds	Quote conversion Analytics	Predict impact of external factors on business
				Cognitive document content processing
				PII redaction
				Cash flow forecasting

The application of AI in insurance, particularly in the areas of process or workflow automation, has already realized several possibilities – from simple cognitive actions like reading text to making an autonomous decision on a quote or a claim.

Rapid advances in the technology will only lead to disruption of traditional insurance models, so insurers need to future proof their business. Insurers need to see AI as an opportunity to innovate and offer newer products in a responsible way. With expanding application in areas not limited to claims assessment, preventive maintenance, agent augmentation, risk advisory, etc., AI is poised to become an integral part of the insurance ecosystem.

#DATAPOWEREDADVISORY #AIDRIVENDISTRIBUTION #RESPONSIBLEAI

INNOVATION TAKEAWAYS

SIMPLIFY INSURANCE WITH AI

Inspire confidence and comfort within customers with uncomplicated and personalized advisory at the right time.

DISTRIBUTION INNOVATION IS THE WAY FORWARD

Set up a powerful digital sales channel with rich customer engagement to complement human-assisted channels.

OPPORTUNITY FOR DISRUPTIVE GROWTH

AI-driven automation leveraging extensive data will lead to more products being sold online and for more segments than ever.

DEVELOP RESPONSIBLE AI INTERVENTIONS

Mitigate inherent biases and black-box issues to comply with heavily regulated insurance industry standards.



A CASE FOR CONTEXT AWARENESS IN AI

ROBERT ENGELS

CTO I&D Northern Central Europe,
Capgemini



THERE IS A NEED FOR CONTEXTUAL KNOWLEDGE IN ORDER TO MAKE APPLIED AI MODELS TRUSTABLE AND ROBUST IN CHANGING ENVIRONMENTS."

There have been catastrophic effects of AI use in self-driving cars, including road crashes, social media, and failures in critical infrastructures, making some ask: can we trust AI in production? Also, what can we do to make AI more robust while operating in dynamic surroundings and, most importantly, how can we make AI understand the real world?

Does applied AI have the necessary insights to tackle even the slightest (unlearned or unseen) change in context of the world surrounding it? In discussions, AI often equals deep-learning models. Current deep-learning methods heavily depend on the presumption of "independent and identically distributed" data to learn from, something which has serious implications for the robustness and transferability of models. Despite very good results on classification tasks, regression, and pattern encoding, current deep-learning methods are failing to tackle the difficult and open problem of generalization and abstraction across problems. Both are prerequisites for general learning and explanation capabilities.

There is great optimism that deep-learning algorithms, as a specific type of neural network, will be able to close in on "real AI" if only it is further developed and scaled up enough (Yoshua Bengio, 2018). Others feel that current AI approaches are merely a smart encoding of a general distribution into a deep-learning networks' parameters, and regard it as insufficient to act independently within the real world. So, where are the real intelligent behaviors, as in the ability to recognize problems and plan for solving them and understanding of the physics, logics, causality, and analogy?

Understanding the real world

What is needed is a better understanding by machines of their context, as in the surrounding world and its inner workings. Only then can machines capture, interpret, and act upon previously unseen situations. This will require the following.

- Understanding of logical constructs as causality (as opposed to correlation). If it rains, you put on a raincoat, but putting on a raincoat does not stop the rain. Current ML struggles to learn causality. Being able to represent and model causality will to a large extent facilitate better explanations and understanding of decisions made by ML models.
- The ability to tackle counterfactuals, such as “if a crane has no counterweight, it will topple over.”
- Transferability of learned “knowledge” across/between domains; current transfer learning only works on small tasks with large domain overlap between them, which means similar tasks in similar domains.
- Withstand adverse attacks. Only small random changes made in the input data (deliberately or not) can make the results of connectionist models highly unreliable. Abstraction mechanisms might be a solution to this issue.
- Reasoning on possible outcomes, finding problematic outcomes and a) plan for avoiding them while reaching the goal or b) if that is not possible, find alternative goals and try to reach those.

In the first edition of this review, we already made the case for extending the context in which AI models are operating, using a specific type of model which can benefit from domain knowledge in the form of knowledge graphs. From the above it follows that knowledge alone probably will not be enough. Higher-level abstraction and reasoning capabilities are also needed. Current approaches aim at combining “connectionist” approaches with logical theory.

1. Some connectionists feel that abstraction capability will follow automatically from scaling up networks, adding computing power, and using more data. But it seems that deep-learning models cannot abstract or generalize more than learning general distributions. The output will at the most be a better encoding but still not deliver symbolic abstraction, causality, or showing reasoning capabilities.

2. Symbolic AI advocates concepts as abstracted symbols, logic, and reasoning. Symbolic methods allow for learning and understanding human-made social constructs like law, jurisprudence, country, state, religion, and culture. Could connectionist methods be “symbolized” as to provide capabilities as mentioned above?

3. Several innovative directions can be found in trying to merge methods into hybrid approaches consisting of multiple layers or capabilities.

- Intuition layer: Let deep-learning algorithms take care of the low-level modeling of intuition or tacit skills shown by people having performed tasks over a long time, like a good welder who can hardly explain how she makes the perfect weld after years of experience.
- Rationality layer: The skill-based learning where explicit learning by conveying rules and symbols to a “learner” plays a role, as in a child told by her mother not to get too close to the edge. A single example, not even experienced, might be enough to learn for life. Assimilating such explicit knowledge can steer and guide execution cycles which, “through acting,” can create “tacit skills” within a different execution domain as the original layer.
- Logical layer: Logics to represent causality, analogy, and providing explanations
- Planning and problem-solving layer. A problem is understood, a final goal defined, and the problem divided in sub-domains/problems which lead to a chain of ordered tasks to be executed, monitored (with intuition and rationality), and adapted.



In general, ML models that incorporate or learn structural knowledge of an environment have been shown to be more efficient and generalize better. Some great examples of applications are not difficult to find, with the Neuro-Symbolic AI by [MIT-IBM Watson lab](#) as a good demonstration of how hybrid approaches (like NSQA in this case) can be utilized for learning in the connectionist way while preserving and utilizing the benefits of full-order logics in enhanced query answering in knowledge-intensive domains like medicine. The NSQA system allows for complex query-answering, learns along, and understands relations and causality while being able to explain results.

The latest developments in applied AI show that we get far by learning from observations and empirical data, but there is a need for contextual knowledge in order to make applied AI models trustable and robust in changing environments.

#DATAPOWERED #AI4GOOD #DATAMASTERS #KNOWLEDGEGRAPHS
#CONTEXTUALAI #CAUSALITYINAI #PLANNINGINAI #PROBLEMSOLVINGINAI
#HYBRIDAI #NEURALSMBOLICAI

INNOVATION TAKEAWAYS

HYBRID APPROACHES

are needed to model and use causality, counterfactual thinking, problem solving, and structural knowledge of context.

NEURAL-SYMBOLIC PROCESSING

combines the benefits of connectionist and symbolic approaches to solve issues of trust, proof, and explainability.

CONTEXTUAL KNOWLEDGE

AI needs modeling more of the world to be able to understand the physics and logics, causality, and analogy in the surrounding world.



UNBOX BLACK-BOX MODELS WITH AI GLASS BOX

**JATINDER KUMAR
KAUTISH**

Chief Architect & Principal Data
Scientist, AI & Analytics, Capgemini

ASHISH DHALI

Senior Manager/Lead Data Scientist,
AI & Analytics, Capgemini



OVERALL, THE AI PROCESS SHOULD GENERATE STANDARDIZED AND BETTER RESULTS, IRRESPECTIVE OF DATA AND CONCEPT CHANGES."

The share of customers trusting AI interactions has increased in the past few years, but it is still less than half. Most customers expect AI black-box models to explain their results with transparency and traceability. This would support wider acceptance of AI solutions for decision-making. The AI Glass Box solution is developed from the ground up to “unbox” black-box AI models. It is an open-source and platform-agnostics asset, developed and piloted by Capgemini across multiple customers and geographies.

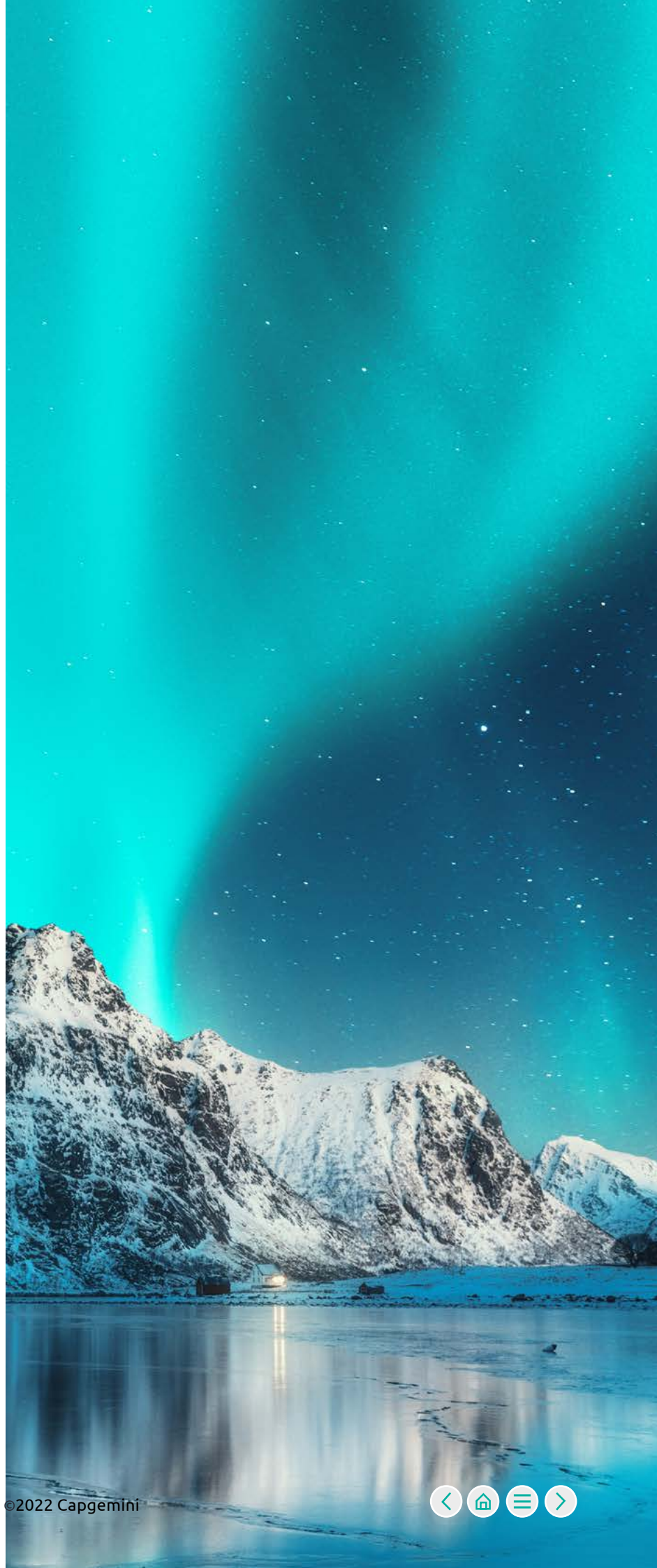
AI is being rapidly adopted at the core of many business processes across different departments within every organization. But multiple challenges arise to track all activities across different teams, due to team-specific standards, language frameworks, separate repositories, and validation processes – with the business

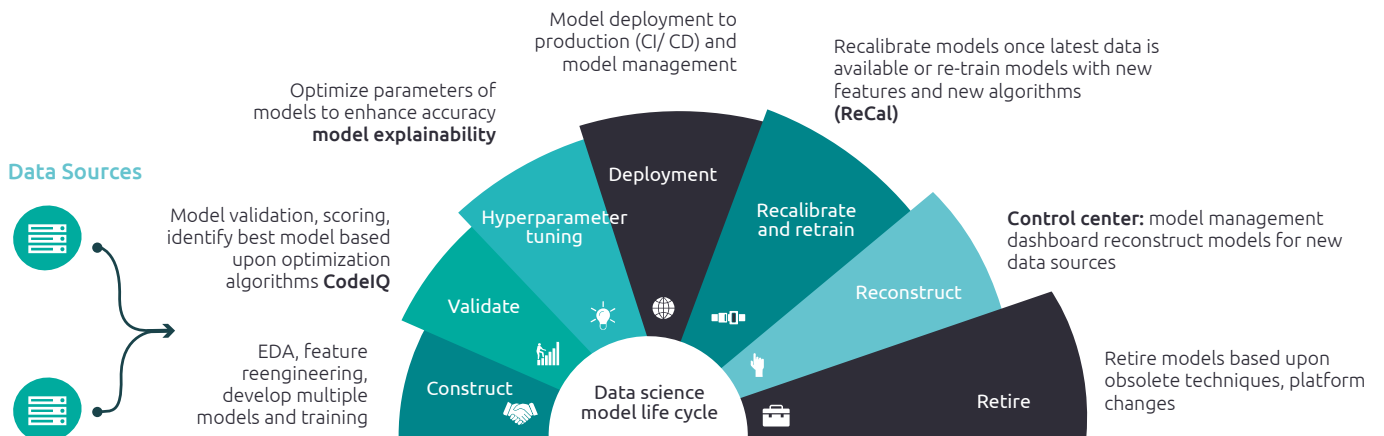
problem, datasets, and final objectives remaining the same. Customers, on the other hand, seek AI outcomes that are explainable, transparent, traceable, and free from bias or prejudice. Overall, the AI process should generate standardized and better results, irrespective of data and concept changes.

Glass Box

AI Glass Box automates end-to-end AI-model-management processes. It is studded with modules which amplify each step and can be implemented across various industries and domain. Its strong open-source foundation provides the base for any ML model validation and “MLOps” implementation of continuous AI delivery.

Once the AI model is baselined in GitHub by data scientists, AI Glass Box gets triggered by pulling the latest model and model artifacts. It traverses the data and model to generate a process document with actionable insights. AI Glass Box utilizes a best-practices store to leverage industry standards and best practices. It is configurable and scalable according to use case and regional and sector specifics.





AI Glass Box tracks all development experiments across frameworks and highlights key improvements in features, engineering, transformations, leveraged models, and scope for hyperparameter tuning with accuracy gain. CodeIQ is a key differentiator module that identifies unnecessary code, passwords, hard coding, and SQL injections. It also performs security vulnerability analysis and highlights the risky and whitelisted packages, library, versions, and techniques according to industry and company standards.

Another key differentiator module is XplainME. It explains the contribution of each input feature towards the model prediction, and also identifies bias and fairness. Along with model approval workflow, the OneView MLOps feature provides a single interface to enable all required native services of cloud for MLOps. The Reboot module uses a proactive approach to detect data drift and model health and empowers the business with retraining decisions. AI Glass Box also aims to manage all responsible AI concepts under a single umbrella. It follows the Switch Board framework, so only required modules for projects need to be configured, hence it is not mandatory to install a whole suite to get a productivity gain.

Examples of successes

Success stories from two assignments show what AI Glass Box can do. The first one covers the model build phase and the second highlights the post-model deployment phase, which is a part of the MLOps approach.

A leading consumer goods and healthcare organization faced challenges implementing standards and best practices and to track all AI development experiments across teams. AI Glass Box's centralized model-validation framework was implemented to track all open-source-based experiments and publish top-down insights in an interactive interface, such as narrations of what is there, should be there, and must be there. Comparative analysis of all experiments was performed and ranked based on configurable metrics. It maintained a repository of all models' versions, artifacts, scripts, and feedback reports.

The second example is a leading communication provider at which the business team was witnessing a lower accuracy of AI models. After deploying the models into production through CI/ CD pipelines, AI Glass Box was implemented to monitor the health of all models, performing drift detection using advanced statistical techniques and raising alerts, if recalibration is recommended. It published actionable insights through a control center for leadership view. AI Glass Box's innovative and unique architecture combines advanced statistical techniques with attribute characteristics to translate the scores to business decisions. It then notifies respective teams whenever any model needs to be re-calibrated with a new dataset or re-trained with new features.

To summarize, AI Glass Box is:

- Based on a platform-agnostic and open-source framework – it is an open-source model validation and model management framework which integrates with cloud as well as on-premises platforms.
- Configurable and scalable – it has a configurable and scalable best-practices store that complies with organizational and regulatory standards and best practices.
- Providing Switch Board flexibility – it boasts multiple modules to establish responsible AI features, such as Xplore, CodeIQ, DeepParam, XplainME, ReBoot, ReCal, OneViewMLOps, and Control Tower. Modules can be switched on or off at will.
- Rebooting models – an early data-drift detection and model-monitoring module constantly checks the data changes and model health and recommends when to retrain models.
- Model agnostic – data scientists prefer their own frameworks to experiment but use cases and data sets remain the same. AI Glass Box tracks all experiments and ranks across different frameworks.



#AIEVALUATOR #MLMASTERS #ETHICALAI #AI@SCALE
#RESPONSIBLEAI, #EXPLAINABLEAI #MLOPS

INNOVATION TAKEAWAYS

CUSTOMERS WANT TO TRUST AI

As they are more and more confronted with AI-driven decisions and communication, customers increasingly ask for AI that is explainable, transparent, and fair.

AI GLASS BOX UNBOXES THE BLACK BOX

AI Glass Box is a toolbox of different modules that automates the process of making AI algorithms comply to different (ethical) needs.

NO MATTER HOW YOU DO IT

Enforcing ethical AI qualities should be consistent and repeatable, even if different approaches, frameworks, and technologies are applied during the AI development process.

THE RISE OF DATA FOR GOOD

LEVERAGING THE VALUE OF DATA FOR SOCIETY 5.0 | 40

Erik Hoorweg
Judith Kennes

DATA AND THE SUSTAINABILITY ECOSYSTEM | 43

Philip Harker
Courtney Holm

BATTLING RIVER BLINDNESS DISEASE WITH THE HELP OF AI | 46

Dr. Daniel Kühlwein
Dr. Janina Kühlwein

IMPROVING LIVES OF FARMERS WITH AI AND AUTOMATION | 49

Marijn Markus
Luc Baardman

AI FOR DYSLEXIA PATIENTS | 52

Dr. Tanu Solanki
AI by Wordtune





LEVERAGING THE VALUE OF DATA FOR SOCIETY 5.0

ERIK HOORWEG

Vice President, Lead Public Sector
Capgemini Invent

JUDITH KENNES

Enterprise Architect, Capgemini Invent



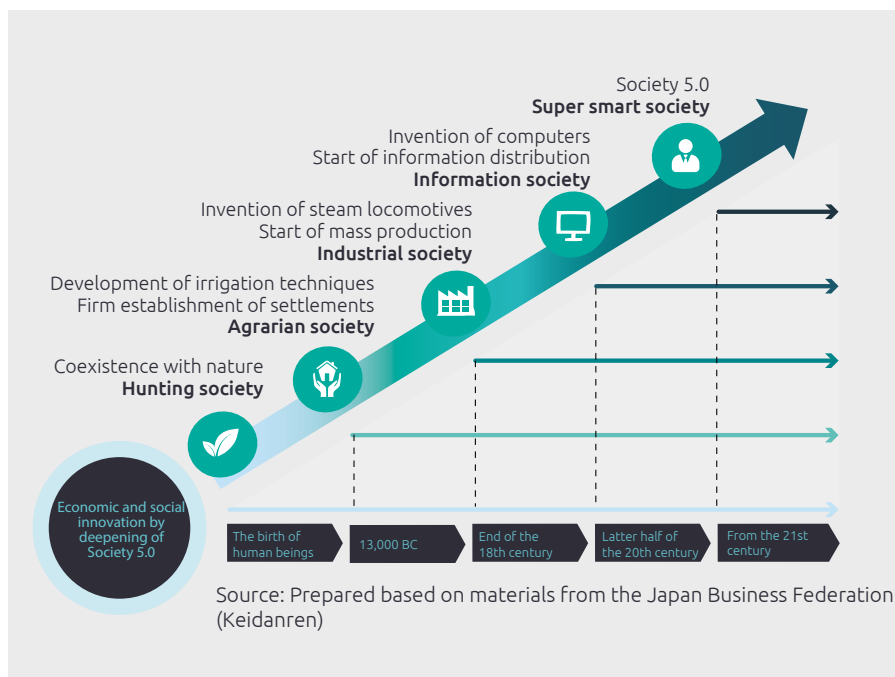
BY LEVERAGING THE VALUE OF DATA TOGETHER, WE'LL STRENGTHEN TRUST AND INCLUSION AND, IN THE END, CREATE A TRULY SUPER-SMART SOCIETY."

Society 5.0: a super-smart society in which borders between humans and technology blur and the virtual and physical world become ever more entwined, and where “things” and citizens communicate, AI is used for analyzing the immense flow of data and virtual reality gives you a peak into a new urban landscape or building. Find out what conditions should be met to reap the benefits from Society 5.0.

The past decades have shown unprecedented change in the way we live, produce, consume, travel, and enjoy our free time. As has been said before, "we do not live in an era of change but in a change of era." From a societal point of view, this new era

is increasingly labelled Society 5.0: a super-smart society in which borders between humans and technology blur and the virtual and physical world become ever more entwined. Society 5.0 follows upon the fourth

industrial revolution. In Industry 4.0, new technologies enable companies to create competitive advantage and more efficient production by connecting and communicating between all kinds of industrial processes. The idea behind Society 5.0 is to use technology and leverage the value of data for tackling societal issues like sustainability, social polarization, aging population, citizen centricity, and mobility.



Important technologies for Society 5.0 are Internet of Things (IoT), big data, cloud, artificial intelligence (AI), and augmented virtual reality (AR/VR). Through IoT, “things” are connected and continuously exchange data. Think of a streetlamp that communicates its out-of-order status. When many things communicate, they produce a lot of data. Combine this with all the data people produce on social media and organizations generate in their day-to-day operations and there is a lot of data – big data – that can be used for all sorts of analyses. To store, process, and analyze the data, cloud technologies make it easy to scale solutions accessible for everyone. AI algorithms can be leveraged to intelligently analyze big amounts of data. With AR/VR, it is possible to “draw” extra information on the real world as seen through VR glasses or a mobile screen. Bringing an urban design virtually to life helps in discovering practicalities and impracticalities by, for example, projecting cables and sewage systems virtually on the ground to speed problem solving.

If we can already do so much with new technologies, how do we use them successfully? And how do we transition to a Society 5.0? We have indicated five pivotal conditions for success.

1. **Skills.** [According to the World Economic Forum](#), the increase of AI and automation will impact more than 1.2 billion employees in the next 10 years. Not only will jobs change, but new jobs will also. Two-third of today’s toddlers will have a job that [does not exist yet](#). The ability to learn new skills is pivotal for individuals to remain relevant in the job market and it is essential for companies to survive and grow. That notion seems not yet fully embraced by industry leaders: 45 percent of them communicate on automation initiatives but only 15 percent initiate reskilling or upskilling initiatives. It is time for some strategic workforce planning.

2. **Data spaces.** Leveraging the value of data often means sharing of data. The planned investment in supporting the development of data sharing and associated technology solutions in the European Union alone in 2021 to 2027 adds up to approximately 10 billion euros. The [Data Governance Act](#) intends to boost innovation and growth in Europe by providing a foundation for data sharing in controlled ecosystems (data spaces) to uphold values like data sovereignty. The real challenge in building successful data spaces lays in the commercial value of sharing. While legal and technical risks are being minimized, willingness and business demand to share data must be understood and answered.

3. **Platforms.** Easy communication between citizens and government is characteristic of Society 5.0: citizens want to be understood by the government and quickly served to full satisfaction. This is possible when governments and departments work together in breaking silos and deliver services through an organization-wide digital platform, composed of smaller specific-purpose platforms such as a platform for omnichannel interaction, the engagement platform for a 360-degree view of citizens, and the workflow platform for connecting required actions of different systems and people to deliver a requested service to the citizen. This can also be a data platform for storing, managing, and analyzing data used by the other platforms, the integration platform for connecting all platforms, and a cloud platform to run all platforms on. In this way, governments can act as one and deliver the best possible services.



4. Privacy and security. Integrating our physical and virtual environment to a great extent as the basis for Society 5.0 is not without risks. Gathering information and combining sources of data can lead to unwanted insights and exposure of personal information and even lead to great personal disadvantages, through expanded surveillance, profiling, and false matches, for example. Other risks are around theft and misuse of data and systems. Society 5.0 will only come to full bloom when we can trust the society. A continuous focus on protecting privacy and delivering security holds true for any society, and this includes the virtual world for Society 5.0.

5. Public values and ethics. Currently, trust in public institutions is under pressure in many countries. The Capital Hill attack in the USA, the “gilets jaunes” in France, and the allowance affair by the Dutch Tax Office are all examples of growing distrust. In order to create a Society 5.0 based on inclusion and trust, it’s paramount to safeguard European public values in the use of new technologies. To do so, ethics by design ought to be adopted in the development of new algorithms.

We are already on the road towards Society 5.0 and, to create the future we want, the five pivotal conditions as described above should be addressed. Creating Society 5.0 requires us to do it together. By leveraging the value of data together, we’ll strengthen trust and inclusion and, in the end, create a truly super-smart society.

#SOCIETY5.0 #DATAPOWERED #AI4GOOD #ESKILLS #DATASPACE

INNOVATION TAKEAWAYS

SOCIETY 5.0 NOT JUST A VERSION NUMBER

It builds on the innovative concepts of Industry 4.0, in which the physical and virtual worlds are becoming more intertwined

LEVERAGE DATA FOR SOCIETY

Within Society 5.0, AI, cloud, and immersive technologies leverage data to create solutions to address societal challenges and opportunities.

CRITICAL CONDITIONS

In order to successfully reap the benefits from Society 5.0, prerequisites must be met regarding e-skills, data spaces, platforms, privacy and security, and public values.



DATA AND THE SUSTAINABILITY ECOSYSTEM

PHILIP HARKER

VP, Sustainability Ninja
Insights & Data, Capgemini

COURTNEY HOLM

VP, Sustainability Solutions,
Capgemini Invent

As the saying goes in the UK, “make hay whilst the sun shines.” And the sun is shining very brightly these days indeed, especially when it comes to data. While sustainability and environmental data may not yet be abundant, sustainability data ecosystems are rapidly evolving to serve the demand for increasing compliance at governmental, intergovernmental, and industry levels.



WITH THE SAME INTENT AND DISCIPLINE AS MONEY, CARBON SHOULD BE TREATED LIKE A CURRENCY, FLOWING THROUGH THE ORGANIZATION, AND WITH IT THE SCRUTINY, CONTROLS, AND SYSTEMS.”

“Data isn’t oil,” says Zhiwei Jiang, CEO of Capgemini’s Insights & Data global business line. “At least not anymore. Data is more like sunshine, abundant and unlimited in its potential. It has a positive impact on the environment. And critically, sunshine wants to be shared – not hoarded.”

Carbon and its equivalents (CO2E) have a currency, and with carbon reduction ambitions and even net-zero initiatives it is essential that organizations commit to adopt change, act in ways that reduce carbon across the business, and further monitor and report carbon, as in Scope 1, 2, and 3.



Capgemini’s net-zero value proposition: we help organizations accelerate their net-zero transformation from commitment to sustainable achievements.



A new currency is born

With the same intent and discipline as money, carbon should be treated like a currency, flowing through the organization, and with it the scrutiny, controls, and systems to avoid diversion and waste – just as with money. Hence the need for systemized and structured data to give transparency and insights to not only support decisions for reporting on Scope 1, 2, and 3 but moreover to empower planning and modelling for business planning and execution. Equally, data residing in siloes isn't going to answer the enterprise requirements for sustainability reporting; data needs to be accessible and available in a form fit for consumption. The potential is with the executives of the software industry and their ability to make data flow and be shared through adoption by the enterprise. Here are some examples.

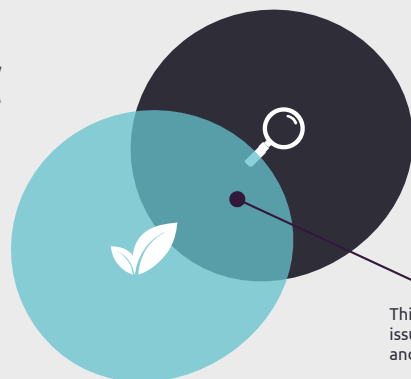
Salesforce has an established service, the [Sustainability Cloud](#), which leverages its proven approach for rapid templating of action of the interpreted adoption of greenhouse gas (GHG) data. This, coupled with [MuleSoft](#) for data integration and APIs, will not only ease the market burden on the shortage of sustainability data skills but also quicken the rolling deployment of GHG reporting of Scope 1, 2, and potentially 3 emissions.

SAP is also building on its core capability and has a unique position in this space, with a perhaps more strategic stance: that of being the guardians of transactional data. Will it become the “ERP for carbon” as is expected? As organizations mature their carbon reporting to be more accountable, their organizations will evolve more into a [sustainable P&L](#), where the business understands, measures, and reduces the environmental, social, and financial impact of its operations. This in itself will play to the strengths of SAP's system-of-record heritage. It has offerings with [Product Footprint](#), circular economy, and enterprise reporting which will surely be harmonized into a suite of holistic products.

Then there is **Microsoft**, with a different advantage. Many organizations today are reporting on their efforts in natural reporting fashion, building custom reports to meet their obligations or regulations. With Microsoft having the de-facto toolset for extracting, cleansing, sorting, and reporting data with 365, Azure, and PowerBI tools, its ubiquitous use will be a significant liberator, enabling transparency within the confines of legacy silos to enable reporting ease. Certainly, in the short-term while enterprise-class solutions are being further developed, Microsoft may well have an advantage and, indeed, the imminent release of [Microsoft Cloud for Sustainability](#) will be a testament to its tooling.

Climate and energy expertise

The cornerstone of any net-zero transformation



Data and full-cycle approach

The enabler of a successful net-zero transformation

This dual expertise allows us to have a solid grasp of environmental issues, define a coherent sustainability strategy across the value chain, and integrate it into the design process.

Capgemini's capability to meet the demand for expertise in data and sustainability.

A new dawn is approaching for sustainability solutions, as is evidenced by these three examples (and there are many others). With all of these come the need for contextual data and sustainability expertise to implement the technology, to instill confidence in the solution and drive business outcomes. This is an evolution, not revolution.

Darwinism: adapt or else...

Consider how many businesses rarely operate in isolation. A consumer-packaged-goods company making shampoo, for example, will understand the ingredients to make the product and the manufacturing and packaging processes, and this data is most likely to reside in its core ERP and PLM systems. The products are shipped from the factory to a warehouse for distribution, to retailers and, naturally, onto consumers. Data will need to coalesce and harmonize along the journey to measure actuals or by defined proxy to report the impact. These data points will reside outside the CPG organization, with the logistics provider and retailers. Data can also come from open data sources standardizing on routes and traffic that are set with revisions. These supply chains are built over a long period,

typically through the lens of a financial business case. Will the currency of carbon and the need to report on Scope 3 emissions trigger the need for these supply-chains to be recalibrated? Indeed, on logistics, the current routes are typically optimized by distance, traffic, and cost, so what about optimizing routes based on impact on the environment, like avoiding a school or a designated low-emissions zone? Here's where open data sources from Google, governments, and academic institutions and not-for-profits can collaborate to form innovative solutions based on a #DataForGood basis.

Tooling for the (new) trade

This is the reality that will evolve, and suddenly solutions get complex. If this is coupled with supply-chain transparency initiatives like digital twins and blockchain, there is indeed another level of complexity. So, simplification is needed, and will come with standards: data systemized and shared openly in a governed and secure ecosystem.

Expectations will evolve and with it the need to trust the inputs and outputs of these data ecosystems along with the need for organizations to share data with many other parties, and to do so for mutual benefit. Core to this is the need for data mastery, and [in our research paper](#) we discuss the various models for collaboration and opportunities to serve and monetize data.

One thing is for sure, expertise in climate, energy, and data lifecycles will be paramount for these to succeed and evolve at pace.

#SUSTAINABLEECOSYSTEMS #NETZERO #CARBONFOOTPRINT #DATAFORGOOD

INNOVATION TAKEAWAYS

CARBON DATA IS A CURRENCY

And to manage it effectively, just as any other currency, its data needs to freely flow within and between organizations.

TECHNOLOGY PROVIDERS ARE MOVING

Providing platforms, standards, and tooling to collect, exchange, and process carbon-related data.

COLLABORATION CHANGES BUSINESS

On their journey to reducing carbon emissions, and organizations find new ways to partner and to collaborate on data, innovating their business models while doing so.



BATTLING RIVER BLINDNESS DISEASE WITH THE HELP OF AI

Using deep learning to improve health outcomes for millions

**DR. DANIEL
KÜHLWEIN**

Managing Data Scientist,
AI Center of Excellence

**DR. JANINA
KÜHLWEIN**

Project Manager AI Oncho, Institute
for Medical Microbiology, Immunology
and Parasitology



TO FULLY REALIZE THE BENEFITS OF AI IN THE MEDICAL SECTOR, DATA SHARING IS ESSENTIAL."

AI and deep learning are ready to improve visual diagnosis of medical conditions. What's missing is high-quality data to train the systems.

The potential for AI and deep learning to accelerate analysis and deliver superior results with fewer errors has excited medical researchers for several years. For example, one of the first successful applications of these technologies to the medical research field occurred in 2017, when researchers at Stanford University in California used a deep convolutional neural network and image analysis to successfully [classify skin lesions](#). The system performed as well as a panel of 21 dermatologists and results such as this made deep

learning a promising candidate for medical applications to automate visual inspections.

But many early AI projects failed to develop into full solutions. In fact, it wasn't until September of 2021 that the US Food and Drug Administration approved the first use of an AI-powered system for diagnostic use. It is a [clinical-grade AI-based solution](#) from Paige AI in New York that assists pathologists in detecting prostate cancers. The company invested more than 10 years in its development.

A condition affecting more than 20 million people

Still, the effort and time invested in such projects has tremendous potential to improve health outcomes around the world. As an example, it could help with the elimination of onchocerciasis – more commonly known as “river blindness.” It belongs to a group of conditions known as Neglected Tropical Diseases and has currently infected more than 20 million people and caused permanent blindness in more than one million.

Common to sub-Saharan Africa, Central America, and South America, river blindness is an infection spread by black flies that typically live near fast-flowing rivers. When the flies bite humans, worm larvae invade the body. These grow into worms which reside in nodules under the skin where they can produce millions of baby worms (microfilariae), which then travel through the body. If these microfilariae reach the eyes they can cause irreversible blindness. Typical symptoms include extreme irritation, inflammation, and itching.

To combat this disease, the World Health Organization’s Onchocerciasis Technical Advisory Subgroup is spearheading a [global effort to eliminate transmission](#) in 10 countries by 2030.

The road to elimination

While drugs currently exist to kill the microfilariae, no medication has yet been developed to target the adult worms. Since adult worms can live and reproduce for up to 15 years, it is difficult to stop transmission through current drug regimens and patients often must receive treatment for decades. Hence the need for new drugs that also target the adult worms.

Evaluating the efficacy of different treatment regimens involves a histological examination of tissue samples. This is a time-consuming process for several reasons. There are only a handful of experts worldwide trained to do the work. Proper examinations can take up to three

months for 150 nodules and more than a year for 2,000 nodules. As a consequence, evaluating and registering new treatments can take years.

Using deep learning to speed up clinical trials

At Capgemini’s AI Center of Excellence, we believe AI can be used to address these challenges and help bring new treatments to market. To that end, Capgemini and the Institute for Medical Microbiology, Immunology and Parasitology (IMMIP) at the University Hospital Bonn in Germany have joined forces to apply the power of deep learning to automate the evaluation of tissue samples.

Capgemini’s role in the project is multifaceted. In addition to developing the algorithms used to evaluate tissue samples, we are also creating suitable training data and designing and implementing the solution. Other project partners include the German Center for Infection Research (DZIF), the Drugs for Neglected Diseases Initiative (DNDI), the Kumasi Centre for Collaborative Research in Tropical Medicine (KCCR) in Ghana, and Washington University in St. Louis.

Our project – AI for Onchocerciasis – plans to use AI to standardize testing and accelerate the examination of tissue samples. We expect these two improvements to shave several months off the time required to evaluate clinical studies.



How it works

Our approach starts with digitizing and labeling microscope slides of tissue samples subject to four different stainings. We use these to train and test deep-learning models to detect and classify worm sections in a histological image.

To handle the massive images involved (up to 35GB) we have developed a two-phase approach. First, we apply object detection on low-res images to identify regions of interest. The underlying models are based on Faster Region-Based Convolutional Neural Networks (R-CNN) and fine-tuned via transfer learning on labeled images.

In the second phase, we crop high-res section images that contain the necessary level of detail to make a judgment and classify them using Convolutional Neural Networks (CNN). This results in a separate model for each attribute we seek to identify when inspecting a new tissue sample. The initial results are very promising, and we will do a first test on ongoing clinical trials in 2022.

Crowd-sourcing AI development with the Global Data Science Challenge

While our results are already positive, the field continues to develop rapidly and there is room for improvement. To help with this, we are engaging with the fifth edition of the Capgemini Global Data Science Challenge (GDSC). We will launch this challenge, called [Code for a Cure](#), in early 2022, and task participants with creating an AI-based solution to automate the current manual evaluation process.

The GDSC is an opportunity to accelerate our progress, and we're looking forward to the outcome of this competition. At the same time, using the GDSC to tackle river blindness represents an accessible, hands-on introduction to this exciting field. It will help teach AI to the next generation of data scientists, and help non-experts identify opportunities to add AI to their projects.

Looking ahead

The partners in this project are already working on next steps. We want to verify our results in ongoing clinical trials and develop a user interface to allow researchers who lack

programming knowledge to use our solution. And because this project represents a significant opportunity to improve health outcomes for millions of people, we will share our results through open-source venues with the goal of enabling other researchers, in keeping with the vision of AI for good. This is fitting, as our solution builds on the work of others and would not have been possible without sharing via the open-source movement.

AI modeling techniques are ready to automate visual tasks and improve their results. They can shave precious time off drug discovery and approval processes and contribute to a massive transformation of the health industry.

The biggest obstacle to this is access to high-quality data for training the AI systems. To fully realize the benefits of AI in the medical sector, data sharing is essential. The good news here comes from an unlikely source: the global pandemic.

COVID-19 has demonstrated the value of working together to create medical solutions to pressing problems. Already, we're seeing a big push towards data sharing from both the research community and the health sector.

#AI4GOOD #ENDNTDS #GDSC

INNOVATION TAKEAWAYS

LET'S HAVE A LOOK AT YOU

In medical practice, diagnosis often relies on visual examination. AI-powered systems can support this, and even automate the process.

GARBAGE IN, GARBAGE OUT

The algorithms are ready to enable AI-powered visual examinations. Getting the right data set is the biggest issue. Data sharing is essential.

CROWD-SOURCING AI WITH GDSC

Capgemini's Global Data Science Challenge combines AI for good with hands-on opportunities to teach AI to data scientists and introduce it to those in other fields of study.



IMPROVING LIVES OF FARMERS WITH AI AND AUTOMATION

Project FARM: how automated data insights help farmers regain control over their local food supply chain

MARIJN MARKUS

AI Lead & Senior Data Scientist,
Capgemini

LUC BAARDMAN

Ecosystem Facilitator,
Capgemini

Project FARM uses AI and automation to give smallholder farmers access to innovative technology. With FARM, agricultural value-chain players access an online platform that improves relationships between farmers, traders and suppliers, chain operators, and governments. This strengthens the information position of individual players and stabilizes their way of living. FARM supports data-driven decision-making at any moment in the growing season. The project's aim is to improve the lives of one million smallholder farmers worldwide.

The world population is expanding. The good news is that the number of people living in extreme poverty has been steadily decreasing globally. The bad news is that there is a steep increase in the demand for food. In

Sub-Saharan Africa alone, demand will increase by 55 percent over the next 10 years. Worldwide, there are 500 million smallholder farmers. They farm small volumes of produce on small plots (under two hectares). The majority of them have low access to input material, knowledge, and financial resources, and experience a weak link to markets. Because of this, their productivity will likely remain low and many farmers remain trapped in poverty cycles. And yet, not only local but also global food supply-chain players (and thereby a great part of the world's population) rely on these smallholder farmers, for they produce 70 percent of the world's food, [according to the UN](#).

To produce crops, smallholder farmers use relatively simple, rudimentary methods. Their livelihood takes place on a human scale, with a few hectares of land, no tractors, and manual work



FARM IS AN ONLINE PLATFORM THAT SUPPORTS BUILDING TRUSTED RELATIONSHIPS BETWEEN THE THREE ESSENTIAL PARTIES FOR SUCCESSFUL PRODUCE TRADE: THE SUPPLIER, THE TRADERS, AND THE FARMERS."

done by themselves, family members, and/or hired workers. No drones, but a donkey – in the best case. Information flows from field advisors and through word-of-mouth. To these farmers, small improvements can make a huge difference. Many NGOs worldwide aim to aid these smallholder farmers through providing tools, knowledge, advice, or credit. Project FARM teams up with these NGOs.

Core: Automated insights from data

Of course, we are not farmers. We are analysts, architects, and experts in IT. We are system integrators who are in the position to aid these smallholder farmers. We do that by giving NGOs access to the insights that are now still hidden – in raw data.

Project FARM uses artificial intelligence to automatically generate insights from big data through geo-visualization and pattern recognition. FARM is therefore also an online platform that supports building trusted relationships between the three essential parties for successful produce trade: the supplier, the traders, and the farmers. Through the platform, they can find out about something as rudimentary as each other's names and location but also share data safely and make predictions based on previous years of yield and agricultural input. The pattern-recognition models provide the technical basis for the latter.

So, where does the big data, AI, and machine learning come in? FARM uses various machine-learning techniques to generate recommendations for farmers, made actionable through digital portals and cell phone SMS-services. After all, not every farmer in the developing world has access to a smartphone – but almost everyone has feature phones like a good old Nokia.

Examples of insights

Project FARM has its roots in all kinds of data sources – from weather data such as humidity or temperature, rainfall, or air pressure, and from historic agriculture data to flood and fire risks. Additionally, FARM uses



inputs like pictures of farms to indicate soil fertility. It also leverages Sentinel 5 satellite aerial imagery and an analysis of wider-than-RGB-spectrum aerial images to map Soil Organic Carbon and measure CO2 content. Many of these features serve as inputs for the biggest and perhaps most complex part of Project FARM: the prediction of a farmer's yield end of season, and what the farmer can do to optimize this yield.

By automating both the data gathering and insight-generation steps, Project FARM strengthens the information position of farmers and aids in their decision making. It also enables farmers to connect and grow in engagement with digital and financial services like NGOs and – in the future – creditors. FARM does not exist out of loose components; it is the interoperability of modules combined that create the value for the entire value chain.

To that agricultural value chain, the benefits of FARM are manifold: input providers can effectively supply smallholder farmers with agricultural knowledge and advice, and traders and processors can efficiently source products. Financial-service providers can close financing gaps against lower risk and costs, and farmers can make a fair and sustainable income.

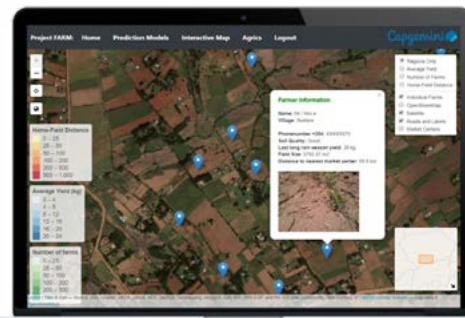
Feeding the world

The FARM solution has been built in collaboration with NGOs operative in Kenya (2019) and now India (in 2021) in collaboration with our partner IBM. IBM has provided the software and hardware necessary for FARM to run but all components have been built to be cloud-agnostic. In total, 25,000 data points have been registered on the FARM platform and we are looking for ways to expand the information position of the platform, enriching with global weather data that can indicate and give accuracy to severe climate disasters predictions.

Best practices from Kenya show that scale is everything. Without enough valid data points, predictions will not be accurate. FARM is now being prepared to go-live in the West Bengal region, where an NGO provides 10,000-plus local farmers with agricultural training. Project FARM is set to grow and expand on its features in the years to come. After all, the fertile soil for Project FARM is data – and the data is already there.

Who knew that simple automated data insights could have the power to feed the world and eventually might give that little extra push to create a poverty-free world with zero hunger and reliable access to information?

Doing good with AI, providing IT to those in need, is how we envision our place in a world that is changed by technology.



#AI4GOOD #AGRICULTURE #AI #DATA

INNOVATION TAKEAWAYS

AI FOR FARMING

Project FARM uses artificial intelligence (AI) to automatically generate insights from big data through geo-visualization and pattern recognition.

AUTOMATING ALL THE WAY

By automating both data gathering and insight generation, Project FARM strengthens the information position of farmers and aids in their decision making.

SIMPLE IS ENOUGH

The most advanced AI-driven insights can be accessed through simple online portals or even phones that only support text messages.



AI FOR DYSLEXIA PATIENTS

DR. TANU SOLANKI

Senior Data Science
Consultant, Capgemini

AI FROM WORDTUNE



PEOPLE WITH DYSLEXIA
CAN IMPROVE THEIR
TEXT USING AI TOOLS
LIKE WORDTUNE."

Artificial intelligence has evolved into a powerful tool that simplifies our lives in a variety of ways.

AI-powered tools such as Wordtune help people with dyslexia with reading and writing. Word choice, spelling correction, sentence selection, tone correction, and expanding and contracting sentences can all be improved using the tool.

Communicating in a foreign language can be challenging, especially if you are not a native speaker or don't understand the rules of the language. One in five students or about 20 percent of the population suffers from dyslexia, the most common learning disability. Dyslexia affects more than 40 million Americans, but only two million have been diagnosed. AI tools can detect early signs, correct and auto-complete sentence formats, check for spelling errors, offer multiple replacements for a single sentence, and summarize articles.

Social-media platforms allow us to

present ourselves to the world via content. We can also showcase our capabilities to the world using blogs, articles, and websites. AI-powered tools can create an entire article based on keywords. Paraphrasing and summarizing articles can also be achieved by AI. It is possible to improve grammar, avoid plagiarism, and proofread. Content personalization creates relevant content based on visitor data (such as location, demographics, and shopping habits). By using these tools, you can scan your text for high-performing keywords that will drive traffic to your website. For those with dyslexia, many AI-based

tools can be used for screening, grammar and spelling checks, paraphrasing, and reading assistance. AI technology is significantly improving the lives of those with dyslexia.

Dyslexia is derived from the Greek words "dys" (poor) and "lexis" ("language"). Identifying speech sounds and learning how they relate to letters and words (decoding) are two of the difficulties associated with dyslexia. Dyslexia, also called reading disability, affects certain areas of the brain, leading to reversal of sounds in words or confusion of words that sound similar. Some people with dyslexia are unable to perform many of the tasks that come naturally to others.

How AI-enabled tools help those with dyslexia

Some AI tools use computer games and machine learning to detect dyslexia in Spanish-speaking individuals. These also include 40,000 games for children with dyslexia that track and analyze observable symptoms. There are some tools that can help people find the right word, spell check, and paraphrase the sentence for better understanding. Voice commands can be used to take notes using AI listening tools. Reading comprehension can be improved by speaking the text out loud and then breaking it down into syllables. People with dyslexia can learn to read independently in this way. In addition to taking dictation, the machine can flag spelling errors during the writing process.

Wordtune

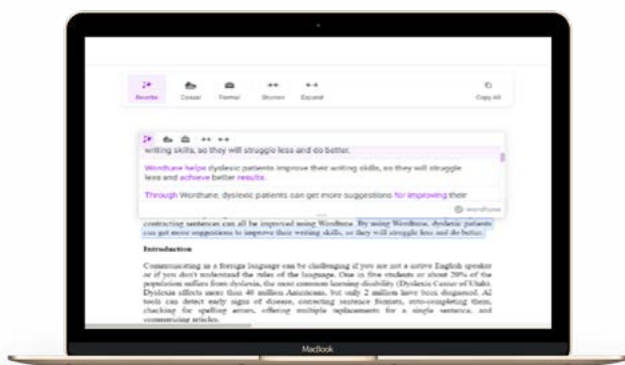
AI21 Labs, founded in 2018, developed Wordtune. The app hopes to revolutionize how we write and read. Advanced AI is used to recognize context and semantics in written texts. It is the first AI-based writing companion that goes beyond grammar and spelling corrections. The software includes an editor, a proof-reader, a tone checker, and a thesaurus.

Wordtune used by those with dyslexia

Dyslexia makes it extremely difficult for patients to locate correct words and sentences. People with dyslexia can improve their text using AI tools like Wordtune. It allows the user to rearrange sentences into multiple forms by using options like Rewrite, Casual, Expand, Shrink, Formal, and word selection. It is enough for the user to choose the form in which to represent the sentence and then select the desired option. If a user does not know how to expand a sentence, the Expand option provides several choices. If a user wants to shorten a sentence, this could be accomplished using the Shorten option.

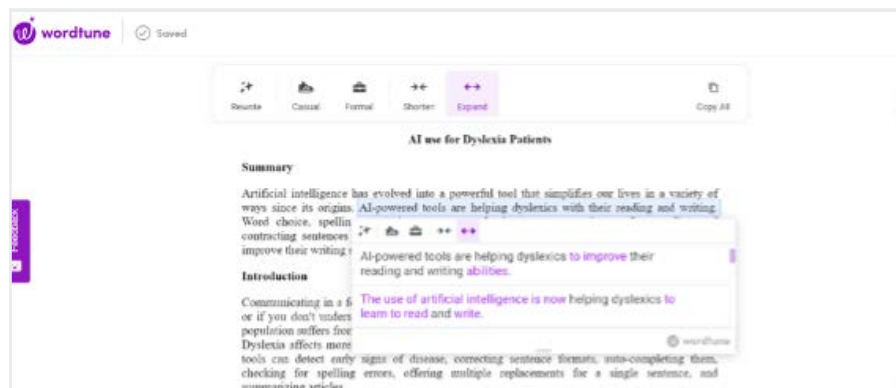
Wordtune used by educators

It is estimated that 10 percent of Indian children suffer from dyslexia, which means nearly 35 million people of student age. Teachers can use Wordtune to offer students word options as they struggle to understand the sound of the word. If students cannot understand a word, they can form synonyms. Students with dyslexia benefit from multiple sentences and printing assignment submission is another option.



Future of AI

There were and still are many suspicions about artificial intelligence. A machine cannot be trusted to be accurate and trustworthy, some believe. Artificial intelligence is always being developed and experts predict that instead of searching all over Google, you will one day be provided with virtual assistants that use your likes, emotions, habits, and reactions to recommend specific content to you. Most AI-generated content today is of short and medium lengths. AI tools will be able to write entire books in the future.



#DATAPOWERED #AI4GOOD #AICONTENTCREATION #AI4DYSLEXIA

INNOVATION TAKEAWAYS

LANGUAGE POWER

AI systems capabilities are maturing to not only interpret and analyze natural language but also to generate new content, both entirely new or as improvements of existing text.

TUNED BY WORDTUNE

Created in 2018 by AI21 Labs, Wordtune is the first AI-based writing companion that goes beyond grammar and spelling corrections. The software includes an editor, a proof-reader, a tone checker, and a thesaurus.

INCLUSIVE AI

By making the full power of AI-enabled language available, people with dyslexia are better able to express themselves while improving their language skills. AI systems such as Wordtune are therefore inclusive technologies.



ANITHA MADHUSUDHAN
Consultant at AICOE,
ID DSA India



ARVIND RAO
Partner Architect Advisor,
Google



ASHISH DHALI
Senior Manager/Lead Data
Scientist, AI & Analytics



COURTNEY HOLM
VP, Sustainability Solutions,
Cappgemini Invent



DAMIEN PHILIPPON
Co-founder, COO,
Zelros



DR. DANIEL KÜHLWEIN
Managing Data Scientist,
AI Center of Excellence,
Cappgemini



DINAND TINHOLT
VP, CPRD Lead,
Cappgemini North America



ERIK HOORWEG
VP, Lead Public Sector,
Cappgemini Invent



EVE BESANT
SVP, Worldwide Sales
Engineering, Snowflake



DR. JANINA KÜHLWEIN
Project Manager AI Oncho,
Institute for Medical Microbiology,
Immunology and Parasitology, Bonn University



AUTHORS



JATINDER KUMAR KAUTISH
Chief Architect & Principal Data
Scientist, AI & Analytics,
Cappgemini



JITESH S. GHAI
Chief Product Officer,
Informatica



JUDITH KENNES
Enterprise Architect at
Cappgemini Invent



KIRAN BOOSAM
Vice President, Global Insurance
Strategy & Portfolios, Cappgemini



LUC BAARDMAN
Ecosystem Facilitator,
Cappgemini Invent



MARIJN MARKUS
AI Lead & Senior Data
Scientist, Cappgemini



MARIUS NILSEN
Architect,
Cappgemini Norway



PHILIP HARKER
VP, Sustainability Ninja,
Cappgemini



ROBERT ENGELS
CTO I&D Northern Central
Europe, Cappgemini



SAM MAHALINGAM
Chief Technology Officer,
Altair



DR. TANU SOLANKI
Sr. Data Science
Consultant, Cappgemini



YASHOWARDHAN SOWALE
VP, Enterprise Architect,
Insights & Data, Cappgemini





For more details contact us:

Ron Tolido

EVP CTO and Chief Innovation Officer,
Insights & Data, Capgemini

ron.tolido@capgemini.com

Monish Suri

Director, Global Innovation Lead & Google Alliance Lead,
Insights & Data, Capgemini

monish.suri@capgemini.com

Johanna Sundh

Global Marketing Communications
Director, Data, AI & Analytics, Capgemini

johanna.sundh@capgemini.com

Laura Immonen-Beatty

Global Content Marketing Lead,
Insights & Data, Capgemini

laura.immonen-beatty@capgemini.com

Interested in learning more?

[Download Wave 2](#)

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 everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of over 300,000 team members in nearly 50 countries. With its strong 50-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 2020 global revenues of €16 billion.

Get the Future You Want | www.capgemini.com