Deloitte.
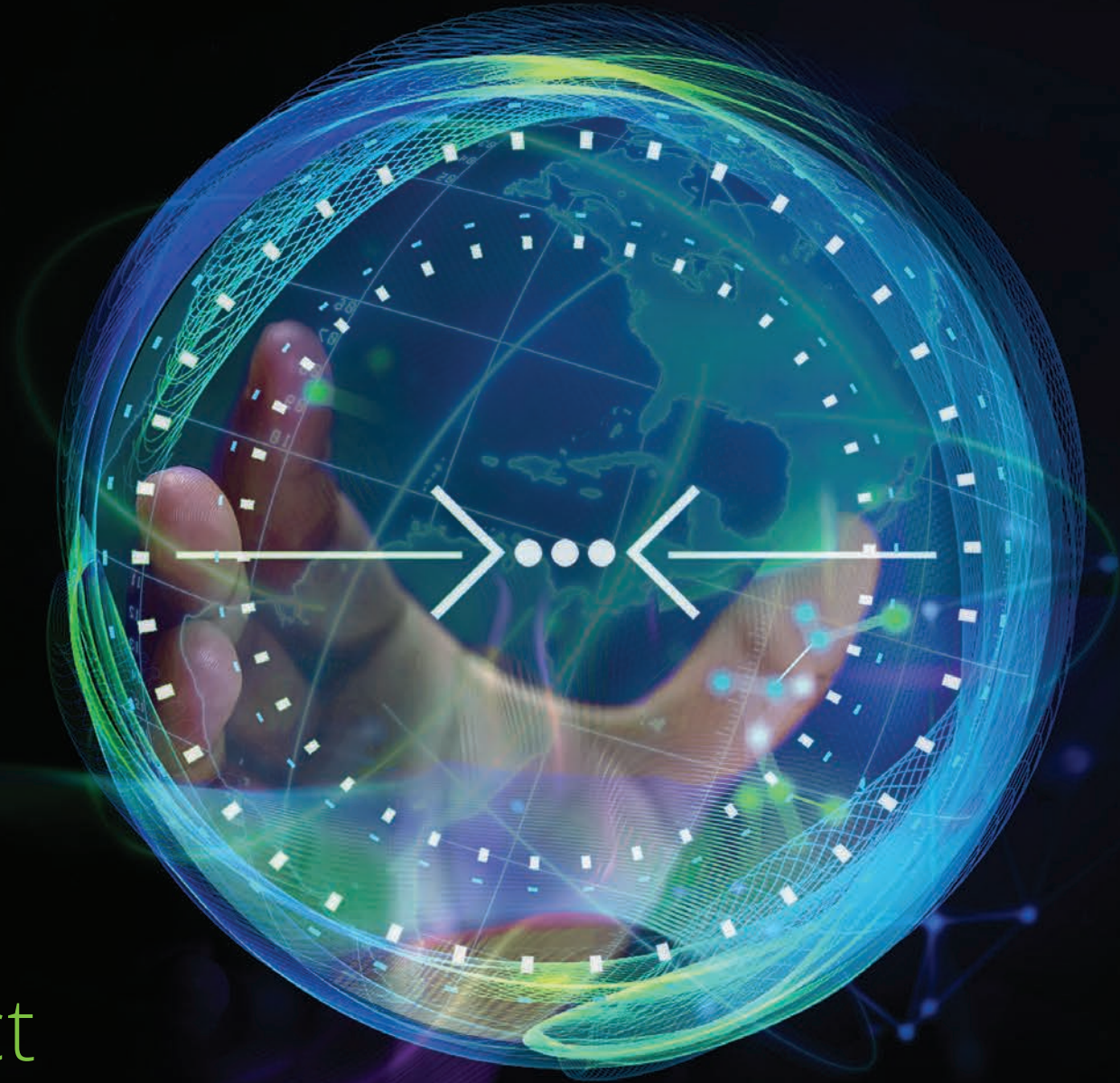
MLOps with impact

Drive business outcomes with
enterprise adoption of AI

A report by the Deloitte AI Institute

Fundamentally world-changing AI technology

As artificial intelligence (AI) is gaining importance, it is creating amazing results across many industries.
It can predict pricing. It can prevent maintenance failures. It can help doctors find early disease. It can detect and help resolve supply chain issues. It can even automate customer service around the clock.

Every department in every company wants some aspect of AI to drive business value. The technology is fundamentally world changing. Its invention can be equated to that of the lightbulb.

Moving AI/ML models from development to production

The machine learning (ML) models that drive AI are no easy feats. **Deploying AI solutions in production is challenging.** If business stakeholders and technologists struggle to collaborate effectively, resulting investments in AI can fail to address the business need. Too often, the focus of data science teams can lie more on designing and deploying highly accurate AI/ML models than working with business and product teams to ensure end-to-end orchestration with business workflow solutions. ML operations (MLOps) can become a costly endeavor. By redefining the framing of MLOps, organizations can better meet the needs of the business and drive value.

Value arrives in a calculated, ongoing process

MLOps can tie models to business value. However, AI is not a typical technology deployment from the past. ML models need to be observed with feedback loops to ensure optimal capabilities. It's not a "once and done" scenario. It's a calculated, ongoing process—and a mindset—that gives data science teams a structured way to rapidly develop, deploy, monitor, and maintain AI/ML solutions that make a real impact on the business. It is not a single tool or technology; there are many in MLOps. A single tool is rarely the answer, as it would be unlikely to drive outcomes that represent an enterprise in its entirety. And none are substitutes for a good process. MLOps is an end-to-end AI/ML life cycle management approach necessary for governance and agility. And it needs to have guardrails in place so the model doesn't get off track and create a public relations nightmare.

Here are some real-world challenges and results from real companies using MLOps to drive business value today.
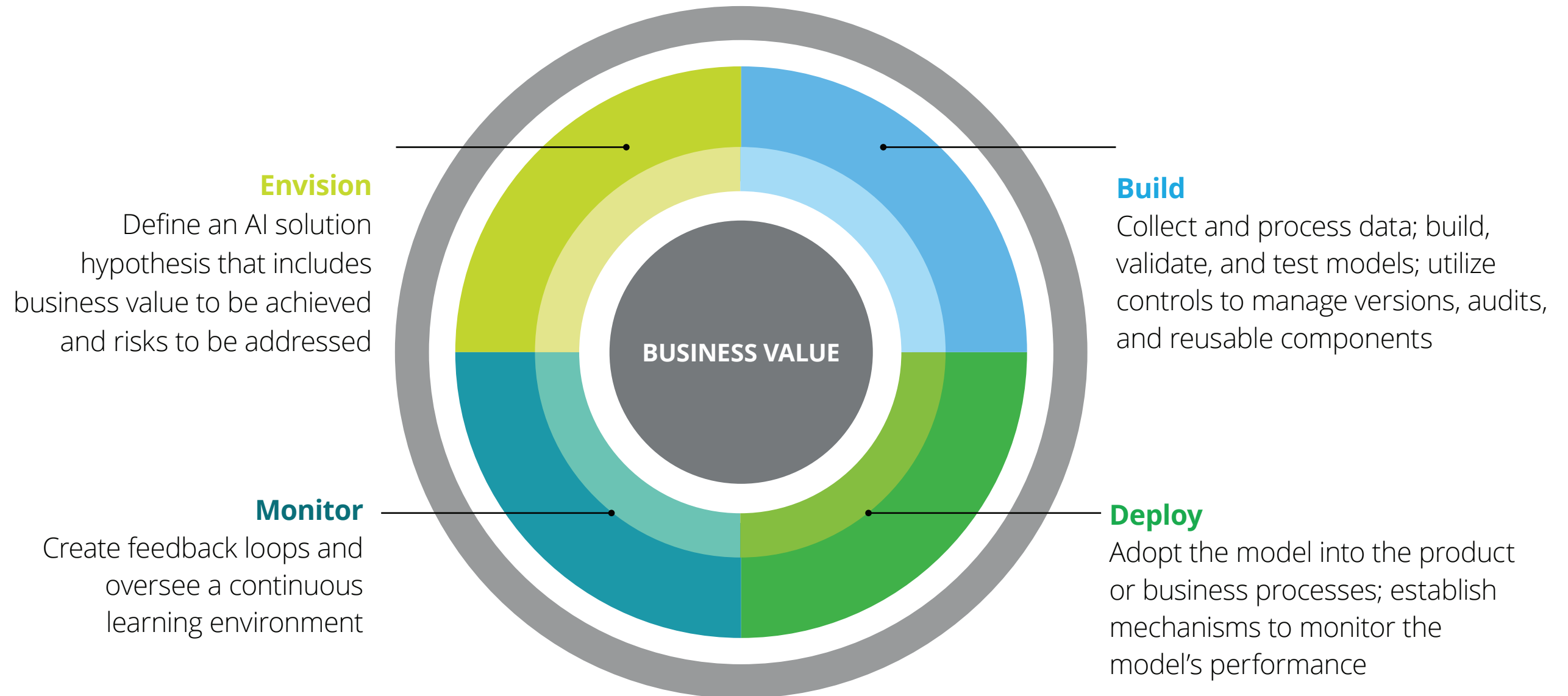
CASE #1
A life sciences company challenge

What intelligence do sales professionals need at their fingertips?

The better question might be: How would you like to influence >$500 million in sales? That was the goal for one global life sciences company. The company started with five countries in which they operate. Applying a distinct MLOps process with the right tools and people to make it happen, an omni-channel excellence program was built, including AI models, data pipelines, data exploratory tools, and alerting capabilities to transform their sales efforts with health care providers. And they did. **They kept business value at the forefront at every step and evenly balanced the process, people, technology, and guardrails to be successful. How?**

The four key phases of MLOps

The Envision phase helps the entire team understand what they are building and why.[1] The vision for this life sciences company was to drive more targeted, personalized engagement with customers. The sales team needed better intelligence to drive their efforts and in less time. At the same time, they wanted to ensure that the models were built thoughtfully, fairly, and accurately.

Envision
Define an AI solution hypothesis that includes business value to be achieved and risks to be addressed

Build
Collect and process data; build, validate, and test models; utilize controls to manage versions, audits, and reusable components

BUSINESS VALUE

Monitor
Create feedback loops and oversee a continuous learning environment

Deploy
Adopt the model into the product or business processes; establish mechanisms to monitor the model's performance

CASE #1

Every company initiating a project like this needs a road map

For this company, Deloitte helped them develop detailed outputs and customer journeys that would produce actual results for sales. **As drug treatments become more personalized, easier ways to reach individual customers will become critical to the company's mission for better health outcomes as well.** The ML models built would help answer a variety of questions.

Granular local performance and insights
What have the customers been buying? How frequently?

Micro-segmentation
How is one primary care doctor different from the next? Or from a fellow physician in the same practice?

Channel affinity
Which channels of communication does each customer prefer?

Channel effectiveness
Which channels of communication do the customers predictably respond to?

Channel sequencing
What are the most effective paths to reach customers and in what order?

Content optimization
Do they need simpler or more robust communications, and how much is right for them?

Envision

CASE #1
The program included Trustworthy AI guardrails

The Deloitte MLOps team incorporated dimensions of a rigorous Trustworthy AI™ framework to ensure fairness. The team clearly identified the accountable people who would help ensure the models didn't waver in that effort over time and produced a RACI to communicate and drive change. The team developed a universal common data model to establish guardrails for use of internal and external data. **A careful governance mechanism was designed with multiple stakeholders, creating a clear, detailed process flow so the sales force could provide regular updates on the value the models were providing and how they might be improved to the data scientists.** This last step was particularly important and included specific differences to accommodate each global location, depending on how often the sales team wanted to both receive and provide updates.
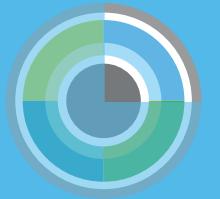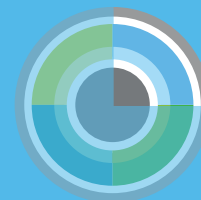
Build

CASE #1
It takes a village

The Build step is where models are brought to life. Through collaboration among all the teams, data scientists will learn more about the business value the company wants to achieve, and business teams will understand more about how data science can give them meaningful information to use. And it's not just data scientists. **Creating and sustaining models at scale requires people with capabilities across data science, ITOps, and UX who work seamlessly toward a common goal.**[2] In this particular case, the collaboration with SalesOps, BrandOps, and digital leaders was key in defining agile workloads to iteratively build.

Build

CASE #1
With sales and scientists in sync, models come to life

Each model that was built ran through this iterative process:

Review requirements
The team aligned on expectations of brand teams and data availability to build, deploy, and sustain models for a particular brand and region.

Collect data
We paired internal data with third-party data, which led to better models and better insights for the sales teams.

Build models
The models were built, tested, and validated.

Set controls
Then the team recorded versions so they could be easily audited later, and reusable components were identified to make future models easier to build.

Show insights
Here is where it all comes to useable life—giving sales professionals the information they need to better engage with their physician customers in an easy-to-understand format.
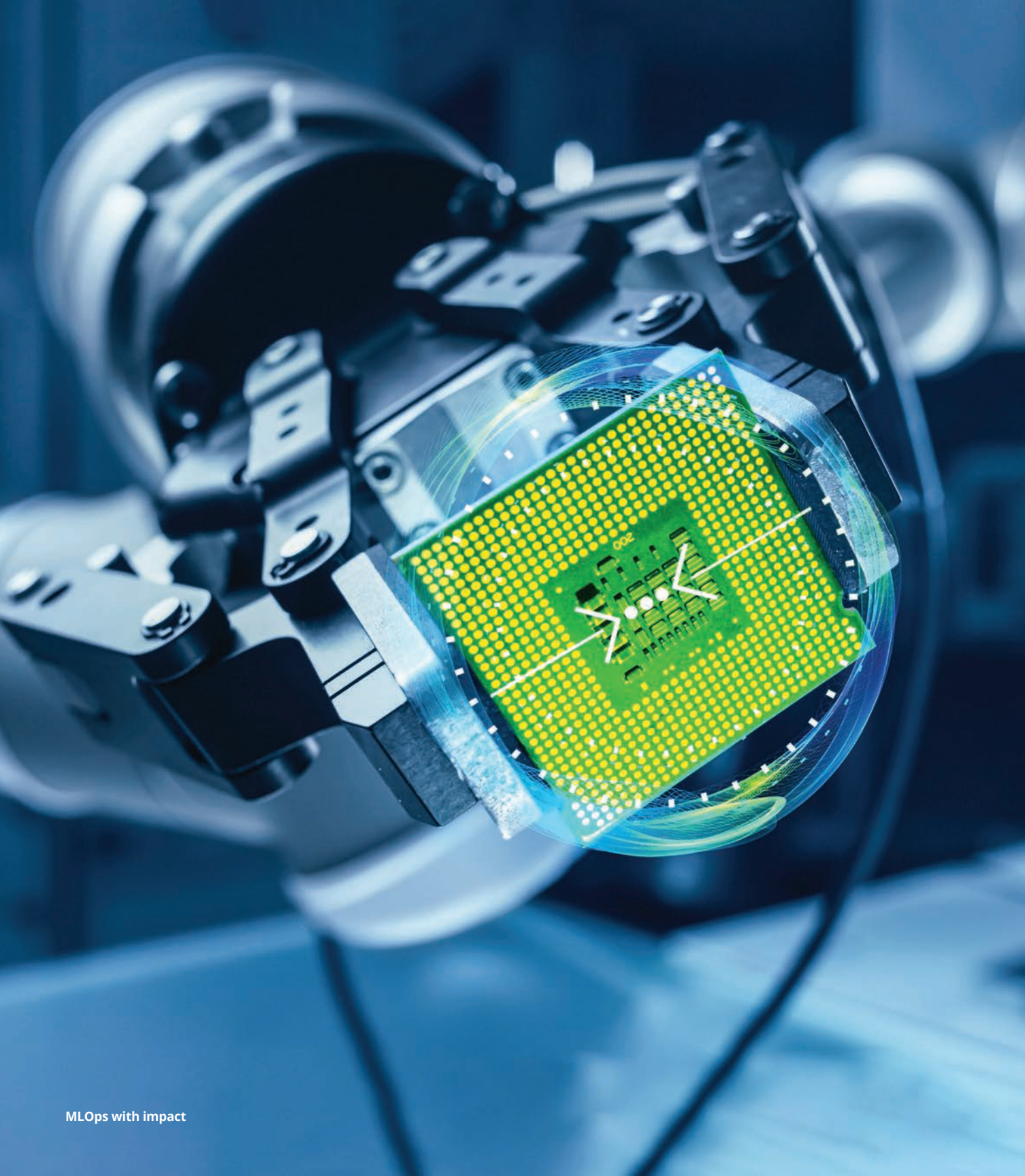
Deploy

CASE #1
Don't underestimate the power of early buy-in

In order to help drive adoption globally, this life sciences company wanted the new system built to ensure adoption across different countries. The team deployed the modeling outputs for this program and carefully integrated this new tool with existing tools the sales teams were already familiar with to drive adoption. This is the essence of "change management" for an innovative program such as this. **If business users don't think it's easy to use, they won't. It's that simple.**

It's important to note that the team worked directly with the business users for this project. Sales professionals need to have input on what data they need, how finely it can be sliced and diced, when it will be updated, and how to use the data in the dashboard. If buy-in is created early on, deployment goes all the easier. The modeling outputs were factored into the programs and processes that sales needed to create more value.

Deploy

CASE #1
Change management is an essential function of a successful project

The effort should be measured. This company took the deployment seriously by looking at sales professionals' adoption rates by brand and region. A/B testing was used to measure receptiveness and adoption by the sales teams—and it was done by teams in local markets to reflect their unique needs. Finally, the models were fine-tuned and rolled out nationally and then globally. **This new program is a combination of business rules and probabilistic theory from the AI algorithm outputs.** So, it did create a different process. Our approach helped reduce the age-old question: "Why are we doing this a new way?" The business intelligence was also designed to give the company a better way to forecast staffing needs for sales teams.

Monitor

CASE #1
Model retraining/feedback

Once deployed, the models were measured for criteria such as drift and shop values. A key component was creating a feedback loop with the model developers and the sales professionals so development teams knew what was meaningful to them or what additional training they needed on it. Each product owner could determine the appropriate "refresh" cycle for the models—from quarterly to monthly to weekly—to make sure the models stayed on track. **Monitoring, for any company, includes overseeing a continuous learning environment with business users. It can't be done in isolation.**

Throughout the process, effective governance and oversight are critical. A front line goal is to make sure that the team is aligned on the criteria of regulatory, privacy, security, and bias policies and determining the business metrics to measure and evaluate the success of the AI effort.
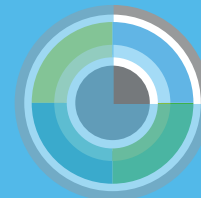
CASE #2
A smart fast-food restaurant with an edge

Distributed AI empowers decision making at the point of action

Imagine the design of a state-of-the-art digital transformation so that it can dramatically improve both employee and customer experience. Deloitte's ability to combine technology with business value helped this quick-service restaurant chain define a breakthrough restaurant experience by prioritizing more than 100 use cases aimed at increasing revenue or reducing costs. The strategy involves:

- Automation at scale throughout the network.

- Futuristic digital experience with a "never down" experience.

Deloitte helped bring together AI, cloud, and edge computing to implement an AI-enabled voice assistant to modernize the experience. This system can scale across many stores with centralized management and still be self-contained at the store level in case of outages. Customers get served faster, menus are updated centrally with ease, and restaurant operations achieve greater efficiency.

Build

CASE #2
A drive-through dream come true

With a clear road map on how to create value for stores, technology teams set out to build a robust solution that included:

A robust cloud platform. Software and ML models securely scale across a multitude of restaurants from a centralized cloud without sacrificing speed to market. The platform enables increasingly complex menus, standardized offerings, governance, and security.

AI on the edge. The edge AI platform with containerized applications transforms restaurant operations using IoT sensors to detect customer vehicles and AI voice agents to send customer order details to the kitchen faster and more efficiently.

Intelligent restaurant operations. Kitchen automation accelerates order-to-delivery times. Linux-based servers designed for process-intensive edge computing allow for failover and resiliency in infrastructure capabilities and monitor machines to help predict maintenance issues before they occur.
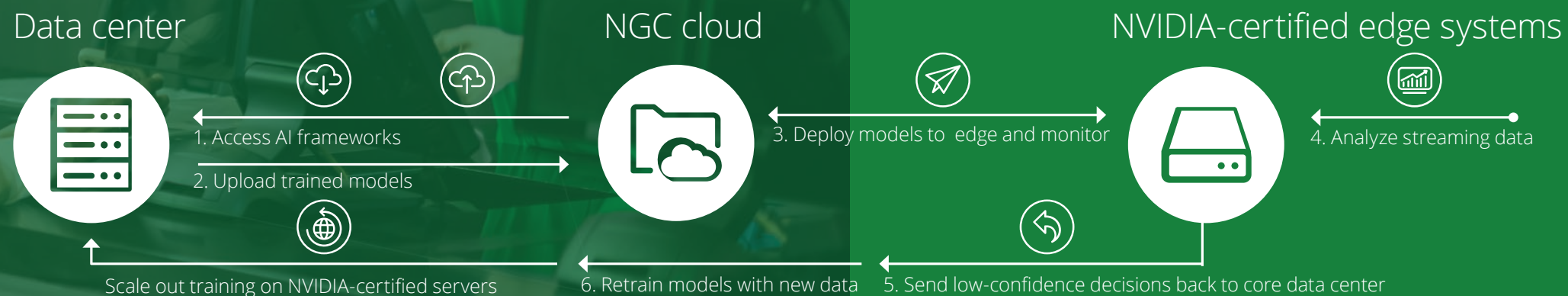
Deploy

CASE #2
A few to many

Once built, the team deployed the solution to more than 350 restaurants in a year with plans to extend to almost 40,000. The solution enables model deployment and inferencing on the edge, so stores aren't dependent on centralized systems, just supported by one. It provides preconfigured deep learning models for object detection using LIDAR datasets, reusable code for seamless deployment of models on edge devices.

Employees were able to reach dedicated agents as systems went into place to answer any questions as the models were deployed and received extensive training on new processes before go-live dates.

Data center

NGC cloud

NVIDIA-certified edge systems

1. Access AI frameworks

2. Upload trained models

3. Deploy models to edge and monitor

4. Analyze streaming data

Scale out training on NVIDIA-certified servers

6. Retrain models with new data

5. Send low-confidence decisions back to core data center

Monitor

CASE #2

Continuous improvement requires careful monitoring

Employee feedbackis critical to understanding what is working well and what seems to be going in a less beneficial direction. Couple that feedback with the technological guardrails to assess if models have drifted away from their original intent, and the company would have a route to not only sustain the program but also improve on it.

The company inferenced deep learning models at the point of decision-making. Such close model monitoring enables the company to detect anomalies, provide preventative maintenance, and empower employees at the store/restaurant.

Monitor

CASE #2

Talk about upsizing your value meal

Deloitte helped this fast-food company build a first-of-its-kind cloud and edge computing implementation to deploy an AI-enabled voice assistant that helped modernize operations and transform the way customers order food, all while making the customer experience better, easier, and faster.

The company could create a personalized, consistent experience, processing orders faster and increasing revenue by automating suggestions for additional purchases. Kitchen operations shifted from a reactive mindset to a proactive one with models that could detect anomalies, saving money while keeping machines running and giving employees the power to act.

The system provides seamless scalability, centralized management, personalized automation, self-healing resiliency, and lower overhead.

MLOps carries real impact when executed right

There is no doubt that ML models are all the rage. The proliferation of tools, the MLOps modules offered by technology behemoths and startups alike are but a testimony to its relevance and importance. Yet, this technology is different in a lot of ways from what has been deployed traditionally. Artificial intelligence carries with it not only the standard deployment tactics but also a whole new set of challenges. It requires an end-to-end management life cycle that continuously needs monitoring with the guardrails and business value checks to keep it on track along the entire way. It also requires focused change management in good measure.

And while all this may not be the easiest to accomplish, real-world world companies are making headway and realizing value. You can, too.

Reach out for a conversation.

Ayan Bhattacharya
Managing Director, AI SGO
Strategy and Analytics
Deloitte Consulting LLP
aybhattacharya@deloitte.com

Sanghamitra Pati
Managing Director, ReadyAI™
Strategy and Analytics
Deloitte Consulting LLP
spati@deloitte.com

ENDNOTES

1 Deloitte AI Institute, *Take a new view on MLOps*, 2022.
2 Deloitte, *MLOops to MLOps*, 2021.

Deloitte.