AWS Announces Eight New Amazon SageMaker Capabilities

Amazon SageMaker Role Manager makes it easier for administrators to control access and define permissions for improved machine learning governance

Amazon SageMaker Model Cards make it easier to document and review model information throughout the machine learning lifecycle

Amazon SageMaker Model Dashboard provides a central interface to track models, monitor performance, and review historical behavior

New data preparation capability in Amazon SageMaker Studio Notebooks helps customers visually inspect and address data-quality issues in a few clicks

Data science teams can now collaborate in real time within Amazon SageMaker Studio Notebook

Customers can now automatically convert notebook code into production-ready jobs

Automated model validation enables customers to test new models using real-time inference requests

Support for geospatial data enables customers to more easily develop machine learning models for climate science, urban planning, disaster response, retail planning, precision agriculture, and more

LAS VEGAS—Nov. 30, 2022—At AWS re:Invent, Amazon Web Services, Inc. (AWS), an Amazon.com, Inc. company (NASDAQ: AMZN), today announced eight new capabilities for Amazon SageMaker, its end-to-end machine learning (ML) service. Developers, data scientists, and business analysts use Amazon SageMaker to build, train, and deploy ML models quickly and easily using its fully managed infrastructure, tools, and workflows. As customers continue to innovate using ML, they are creating more models than ever before and need advanced capabilities to efficiently manage model development, usage, and performance. Today's announcement includes new Amazon SageMaker governance capabilities that provide visibility into model performance throughout the ML lifecycle. New Amazon SageMaker Studio Notebook capabilities provide an enhanced notebook experience that enables customers to inspect and address data-quality issues in just a few clicks, facilitate real-time collaboration across data science teams, and accelerate the process of going from experimentation to production by converting notebook code into automated jobs. Finally, new capabilities within Amazon SageMaker automate model validation and make it easier to work with geospatial data. To get started with Amazon SageMaker, visit aws.amazon.com/sagemaker.

"Today, tens of thousands of customers of all sizes and across industries rely on Amazon SageMaker. AWS customers are building millions of models, training models with billions of parameters, and generating trillions of predictions every month. Many customers are using ML at a scale that was unheard of just a few years ago," said Bratin Saha, vice president of Artificial Intelligence and Machine Learning at AWS. "The new Amazon SageMaker capabilities announced today make it even easier for teams to expedite the end-to-end development and deployment of ML models. From purpose-built governance tools to a next-generation notebook experience and streamlined model testing to enhanced support for geospatial data, we are building on Amazon SageMaker's success to help customers take advantage of ML at scale."

The cloud enabled access to ML for more users, but until a few years ago, the process of building, training, and deploying models remained painstaking and tedious, requiring continuous iteration by small teams of data scientists for weeks or months before a model was production-ready. Amazon SageMaker launched five years ago to address these challenges, and since then AWS has added more than 250 new features and capabilities to make it easier for customers to use ML across their businesses. Today, some customers employ hundreds of practitioners who use Amazon SageMaker to make predictions that help solve the toughest challenges around improving customer experience, optimizing business processes, and accelerating the development of new products and services. As ML adoption has increased, so have the types of data that customers want to use, as well as the levels of governance, automation, and quality assurance that customers need to support the responsible use of ML. Today's announcement builds on Amazon SageMaker's history of innovation in supporting practitioners of all skill levels, worldwide.

New ML governance capabilities in Amazon SageMaker
Amazon SageMaker offers new capabilities that help customers more easily scale governance across the ML model lifecycle. As the number of models and users within an organization increases, it becomes harder to set least-privilege access controls and establish governance processes to document model information (e.g., input data sets, training environment information, model-use description, and risk rating). Once models are deployed, customers also need to monitor for bias and feature drift to ensure they perform as expected.

- **Amazon SageMaker Role Manager makes it easier to control access and permissions:** Appropriate user-access controls are a cornerstone of governance and support data privacy, prevent information leaks, and ensure practitioners can access the tools they need to do their jobs. Implementing these controls becomes increasingly complex as data science teams swell to dozens or even hundreds of people. ML administrators—individuals who create and monitor an organization's ML systems—must balance the push to streamline development while controlling access to tasks, resources, and data within ML workflows. Today, administrators create spreadsheets or use ad hoc lists to navigate access policies needed for dozens of different activities (e.g., data prep and training) and roles (e.g., ML engineer and data scientist). Maintaining these tools is manual, and it can take weeks to determine the specific tasks new users will need to do their jobs effectively. Amazon SageMaker Role Manager makes it easier for administrators to control access and define permissions for users. Administrators can select and edit prebuilt templates based on various user roles and responsibilities. The tool then automatically creates the access policies with necessary permissions within minutes, reducing the time and effort to onboard and manage users over time.
- **Amazon SageMaker Model Cards simplify model information gathering:** Today, most practitioners rely on disparate tools (e.g., email, spreadsheets, and text files) to document the business requirements, key decisions, and observations during model development and evaluation. Practitioners need this information to support approval workflows, registration, audits, customer inquiries, and monitoring, but it can take months to gather these details for each model. Some practitioners try to solve this by building complex recordkeeping systems, which is manual, time consuming, and error-prone. Amazon SageMaker Model Cards provide a single location to store model information in the AWS console, streamlining documentation throughout a model's lifecycle. The new capability auto-populates training details like input datasets, training environment, and training results directly into Amazon SageMaker Model Cards. Practitioners can also include additional information using a self-guided questionnaire to document model information (e.g., performance goals, risk rating), training and evaluation

results (e.g., bias or accuracy measurements), and observations for future reference to further improve governance and support the responsible use of ML.

- **Amazon SageMaker Model Dashboard provides a central interface to track ML models:** Once a model has been deployed to production, practitioners want to track their model over time to understand how it performs and to identify potential issues. This task is normally done on an individual basis for each model, but as an organization starts to deploy thousands of models, this becomes increasingly complex and requires more time and resources. Amazon SageMaker Model Dashboard provides a comprehensive overview of deployed models and endpoints, enabling practitioners to track resources and model behavior in one place. From the dashboard, customers can also use built-in integrations with Amazon SageMaker Model Monitor (AWS's model and data drift monitoring capability) and Amazon SageMaker Clarify (AWS's ML bias-detection capability). This end-to-end visibility into model behavior and performance provides the necessary information to streamline ML governance processes and quickly troubleshoot model issues.

To learn more about Amazon SageMaker governance capabilities, visit aws.amazon.com/sagemaker/ml-governance.

Next-generation Notebooks
Amazon SageMaker Studio Notebook gives practitioners a fully managed notebook experience, from data exploration to deployment. As teams grow in size and complexity, dozens of practitioners may need to collaboratively develop models using notebooks. AWS continues to offer the best notebook experience for users with the launch of three new features that help customers coordinate and automate their notebook code.

- **Simplified data preparation:** Practitioners want to explore datasets directly in notebooks to spot and correct potential data-quality issues (e.g., missing information, extreme values, skewed datasets, and biases) as they prepare data for training. Practitioners can spend months writing boilerplate code to visualize and examine different parts of their dataset to identify and fix problems. Amazon SageMaker Studio Notebook now offers a built-in data preparation capability that allows practitioners to visually review data characteristics and remediate data-quality problems in just a few clicks—all directly in their notebook environment. When users display a data frame (i.e., a tabular representation of data) in their notebook, Amazon SageMaker Studio Notebook automatically generates charts to help users identify data-quality issues and suggests data transformations to help fix common problems. Once the practitioner selects a data transformation, Amazon SageMaker Studio Notebook generates the corresponding code within the notebook so it can be repeatedly applied every time the notebook is run.

- **Accelerate collaboration across data science teams:** After data has been prepared, practitioners are ready to start developing a model—an iterative process that may require teammates to collaborate within a single notebook. Today, teams must exchange notebooks and other assets (e.g., models and datasets) over email or chat applications to work on a notebook together in real time, leading to communication fatigue, delayed feedback loops, and version-control issues. Amazon SageMaker now gives teams a workspace where they can read, edit, and run notebooks together in real time to streamline collaboration and communication. Teammates can review notebook results together to immediately understand how a model performs, without passing information back and forth. With built-in support for services like BitBucket and AWS CodeCommit, teams can easily manage different notebook versions and compare changes over

time. Affiliated resources, like experiments and ML models, are also automatically saved to help teams stay organized.

- **Automatic conversion of notebook code to production-ready jobs:** When practitioners want to move a finished ML model into production, they usually copy snippets of code from the notebook into a script, package the script with all its dependencies into a container, and schedule the container to run. To run this job repeatedly on a schedule, they must set up, configure, and manage a continuous integration and continuous delivery (CI/CD) pipeline to automate their deployments. It can take weeks to get all the necessary infrastructure set up, which takes time away from core ML development activities. Amazon SageMaker Studio Notebook now allows practitioners to select a notebook and automate it as a job that can run in a production environment. Once a notebook is selected, Amazon SageMaker Studio Notebook takes a snapshot of the entire notebook, packages its dependencies in a container, builds the infrastructure, runs the notebook as an automated job on a schedule set by the practitioner, and deprovisions the infrastructure upon job completion, reducing the time it takes to move a notebook to production from weeks to hours.

To begin using the next generation of Amazon SageMaker Studio Notebooks and these new capabilities, visit aws.amazon.com/sagemaker/notebooks.

Automated validation of new models using real-time inference requests
Before deploying to production, practitioners test and validate every model to check performance and identify errors that could negatively impact the business. Typically, they use historical inference request data to test the performance of a new model, but this data sometimes fails to account for current, real-world inference requests. For example, historical data for an ML model to plan the fastest route might fail to account for an accident or a sudden road closure that significantly alters the flow of traffic. To address this issue, practitioners route a copy of the inference requests going to a production model to the new model they want to test. It can take weeks to build this testing infrastructure, mirror inference requests, and compare how models perform across key metrics (e.g., latency and throughput). While this provides practitioners with greater confidence in how the model will perform, the cost and complexity of implementing these solutions for hundreds or thousands of models makes it unscalable.

Amazon SageMaker Inference now provides a capability to make it easier for practitioners to compare the performance of new models against production models, using the same real-world inference request data in real time. Now, they can easily scale their testing to thousands of new models simultaneously, without building their own testing infrastructure. To start, a customer selects the production model they want to test against, and Amazon SageMaker Inference deploys the new model to a hosting environment with the exact same conditions. Amazon SageMaker routes a copy of the inference requests received by the production model to the new model and creates a dashboard to display performance differences across key metrics, so customers can see how each model differs in real time. Once the customer validates the new model's performance and is confident it is free of potential errors, they can safely deploy it. To learn more about Amazon SageMaker Inference, visit aws.amazon.com/sagemaker/shadow-testing.

New geospatial capabilities in Amazon SageMaker make it easier for customers to make predictions using satellite and location data
Today, most data captured has geospatial information (e.g., location coordinates, weather maps, and traffic data). However, only a small amount of it is used for ML purposes because geospatial datasets are difficult to work with and can often be petabytes in size, spanning entire cities or hundreds of acres of

land. To start building a geospatial model, customers typically augment their proprietary data by procuring third-party data sources like satellite imagery or map data. Practitioners need to combine this data, prepare it for training, and then write code to divide datasets into manageable subsets due to the massive size of geospatial data. Once customers are ready to deploy their trained models, they must write more code to recombine multiple datasets to correlate the data and ML model predictions. To extract predictions from a finished model, practitioners then need to spend days using open source visualization tools to render on a map. The entire process from data enrichment to visualization can take months, which makes it hard for customers to take advantage of geospatial data and generate timely ML predictions.

Amazon SageMaker now accelerates and simplifies generating geospatial ML predictions by enabling customers to enrich their datasets, train geospatial models, and visualize the results in hours instead of months. With just a few clicks or using an API, customers can use Amazon SageMaker to access a range of geospatial data sources from AWS (e.g., Amazon Location Service), open-source datasets (e.g., Amazon Open Data), or their own proprietary data including from third-party providers (like Planet Labs). Once a practitioner has selected the datasets they want to use, they can take advantage of built-in operators to combine these datasets with their own proprietary data. To speed up model development, Amazon SageMaker provides access to pre-trained deep-learning models for use cases such as increasing crop yields with precision agriculture, monitoring areas after natural disasters, and improving urban planning. After training, the built-in visualization tool displays data on a map to uncover new predictions. To learn more about Amazon SageMaker's new geospatial capability, visit aws.amazon.com/sagemaker/geospatial.

Capitec Bank is South Africa's largest digital bank with over 10 million digital clients. "At Capitec, we have a wide range of data scientists across our product lines who build differing ML solutions," said Dean Matter, ML engineer at Capitec Bank. "Our ML engineers manage a centralized modeling platform built on Amazon SageMaker to empower the development and deployment of all of these ML solutions. Without any built-in tools, tracking modelling efforts tends toward disjointed documentation and a lack of model visibility. With Amazon SageMaker Model Cards, we can track plenty of model metadata in a unified environment, and Amazon SageMaker Model Dashboard provides visibility into the performance of each model. In addition, Amazon SageMaker Role Manager simplifies access management for data scientists in our different product lines. Each of these contribute toward our model governance being sufficient to warrant the trust that our clients place in us as a financial services provider."

EarthOptics is a soil-data-measurement and mapping company that leverages proprietary sensor technology and data analytics to precisely measure the health and structure of soil. "We wanted to use ML to help customers increase agricultural yields with cost-effective soil maps," said Lars Dyrud, CEO of EarthOptics. "Amazon SageMaker's geospatial ML capabilities allowed us to rapidly prototype algorithms with multiple data sources and reduce the amount of time between research and production API deployment to just a month. Thanks to Amazon SageMaker, we now have geospatial solutions for soil carbon sequestration deployed for farms and ranches across the U.S."

HERE Technologies is a leading location-data and technology platform that helps customers create custom maps and location experiences built on highly precise location data. "Our customers need real-time context as they make business decisions leveraging insights from spatial patterns and trends," said Giovanni Lanfranchi, chief product and technology officer for HERE Technologies. "We rely on ML to automate the ingestion of location-based data from varied sources to enrich it with context and accelerate analysis. Amazon SageMaker's new testing capabilities allowed us to more rigorously and

proactively test ML models in production and avoid adverse customer impact and any potential outages because of an error in deployed models. This is critical, since our customers rely on us to provide timely insights based on real-time location data that changes every minute."

Intuit is the global financial technology platform that powers prosperity for more than 100 million customers worldwide with TurboTax, Credit Karma, QuickBooks, and Mailchimp. "We're unleashing the power of data to transform the world of consumer, self-employed, and small business finances on our platform," said Brett Hollman, director of Engineering and Product Development at Intuit. "To further improve team efficiencies for getting AI-driven products to market with speed, we've worked closely with AWS in designing the new team-based collaboration capabilities of SageMaker Studio Notebooks. We're excited to streamline communication and collaboration to enable our teams to scale ML development with Amazon SageMaker Studio."

About Amazon Web Services
For over 15 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud offering. AWS has been continually expanding its services to support virtually any cloud workload, and it now has more than 200 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, virtual and augmented reality (VR and AR), media, and application development, deployment, and management from 96 Availability Zones within 30 geographic regions, with announced plans for 15 more Availability Zones and five more AWS Regions in Australia, Canada, Israel, New Zealand, and Thailand. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—trust AWS to power their infrastructure, become more agile, and lower costs. To learn more about AWS, visit aws.amazon.com.

About Amazon
Amazon is guided by four principles: customer obsession rather than competitor focus, passion for invention, commitment to operational excellence, and long-term thinking. Amazon strives to be Earth's Most Customer-Centric Company, Earth's Best Employer, and Earth's Safest Place to Work. Customer reviews, 1-Click shopping, personalized recommendations, Prime, Fulfillment by Amazon, AWS, Kindle Direct Publishing, Kindle, Career Choice, Fire tablets, Fire TV, Amazon Echo, Alexa, Just Walk Out technology, Amazon Studios, and The Climate Pledge are some of the things pioneered by Amazon. For more information, visit amazon.com/about and follow @AmazonNews.