

AWS Announces Five New Machine Learning Services and the World's First Deep Learning-Enabled Video Camera for Developers
Wednesday, November 29, 2017 07:11:00 PM (GMT)

Amazon SageMaker makes it easy to build, train, and deploy machine learning models

AWS DeepLens is the world's first deep learning-enabled wireless video camera built to give developers hands-on experience with machine learning

Amazon Transcribe, Amazon Translate, Amazon Comprehend, and Amazon Rekognition Video allow app developers to easily build applications that transcribe speech to text, translate text between languages, extract insights from text, and analyze videos

NFL, Intuit, Thomson Reuters, DigitalGlobe, Hotels.com, ZipRecruiter, Washington Post, Motorola Solutions, Infor, Elementum, iSentia, RingDNA, and City of Orlando are among the many customers using these new services

Today at AWS Re:Invent, Amazon Web Services, Inc. (AWS), an Amazon.com company (NASDAQ: AMZN), announced five new machine learning services and a deep learning-enabled wireless video camera for developers. Amazon SageMaker is a fully managed service for developers and data scientists to quickly build, train, deploy, and manage their own machine learning models. AWS also introduced AWS DeepLens, a deep learning-enabled wireless video camera that can run real-time computer vision models to give developers hands-on experience with machine learning. And, AWS announced four new application services that allow developers to build applications that emulate human-like cognition: Amazon Transcribe for converting speech to text; Amazon Translate for translating text between languages; Amazon Comprehend for understanding natural language; and, Amazon Rekognition Video, a new computer vision service for analyzing videos in batches and in real-time. To learn more about AWS's machine learning services, visit: <https://aws.amazon.com/machine-learning.com>.

This press release features multimedia. View the full release here: <http://www.businesswire.com/news/home/20171129006097/en/>



Amazon SageMaker and AWS DeepLens make machine learning accessible to all developers

Today, implementing machine learning is complex, involves a great deal of trial and error, and requires specialized skills. Developers and data scientists must first visualize, transform, and pre-process data to get it into a format that an algorithm can use to train a model. Even simple models can require massive amounts of compute power and a great deal of time to train, and companies may need to hire dedicated teams to manage training environments that span



AWS DeepLens (Photo: Business Wire)

multiple GPU-enabled servers. All of the phases of training a model—from choosing and optimizing an algorithm, to tuning the millions of parameters that impact the model's accuracy—involve a great deal of manual effort and

guesswork. Then, deploying a trained model within an application requires a different set of specialized skills in application design and distributed systems. As data sets and variables grow, customers have to repeat this process again and again as models become outdated and need to be continuously retrained to learn and evolve from new information. All of this takes a lot of specialized expertise, access to massive amounts of compute power and storage, and a great deal of time. To date, machine learning has been out of reach for most developers.

Amazon SageMaker is a fully managed service that removes the heavy lifting and guesswork from each step of the machine learning process. Amazon SageMaker makes model building and training easier by providing pre-built development notebooks, popular machine learning algorithms optimized for petabyte-scale datasets, and automatic model tuning. Amazon SageMaker also dramatically simplifies and accelerates the training process, automatically provisioning and managing the infrastructure to both train models and run inference to make predictions using these models. AWS DeepLens was designed from the ground-up to help developers get hands-on experience in building, training, and deploying models by pairing a physical device with a broad set of tutorials, examples, source code, and integration with familiar AWS services to support learning and experimentation.

“Our original vision for AWS was to enable any individual in his or her dorm room or garage to have access to the same technology, tools, scale, and cost structure as the largest companies in the world. Our vision for machine learning is no different,” said Swami Sivasubramanian, VP of Machine Learning, AWS. “We want all developers to be able to use machine learning much more expansively and successfully, irrespective of their machine learning skill level. Amazon SageMaker removes a lot of the muck and complexity involved in machine learning to allow developers to easily get started and become competent in building, training, and deploying models.”

With Amazon SageMaker developers can:

- **Easily build machine learning models with performance-optimized algorithms:** Amazon SageMaker is a fully managed machine learning notebook environment makes it easy for developers to explore and visualize data they have stored in Amazon Simple Storage Service (Amazon S3), and transform it using all of the popular libraries, frameworks, and interfaces. Amazon SageMaker includes ten of the most common deep learning algorithms (e.g. k-means clustering, factorization machines, linear regression, and principal component analysis), which AWS has optimized to run up to ten times faster than standard implementations. Developers simply choose an algorithm and specify their data source, and Amazon SageMaker installs and configures the underlying drivers and frameworks. Amazon SageMaker includes native integration with TensorFlow and Apache MXNet with additional framework support coming soon. Developers can also specify any framework and algorithm they choose by uploading them into a container on the Amazon EC2 Container Registry.
- **Fast, fully managed training:** Amazon SageMaker makes training easy. Developers simply select the type and quantity of Amazon EC2 instances and specify the location of their data. Amazon SageMaker sets up the distributed compute cluster, performs the training, outputs the result to Amazon S3, and tears down the cluster when complete. Amazon SageMaker can automatically tune models with hyper-parameter optimization, adjusting thousands of different combinations of algorithm parameters to arrive at the most accurate predictions.
- **Deploy models into production with one click:** Amazon SageMaker takes care of launching instances, deploying the model, and setting up a secure HTTPS end-point for the application to achieve high throughput and low latency predictions, as well as auto-scaling Amazon EC2 instances across multiple availability zones (AZs). It also provides native support for A/B testing. Once in production, Amazon SageMaker eliminates the heavy lifting involved in managing machine learning infrastructure, performing health checks, applying security patches, and conducting other routine

maintenance.

With AWS DeepLens, developers can:

- **Get hands-on machine learning experience:** AWS DeepLens is the first of its kind: a deep-learning enabled, fully programmable video camera, designed to put deep learning into the hands of any developer, literally. AWS DeepLens includes a HD video camera with on-board compute capable of running sophisticated deep learning computer vision models in real-time. The custom-designed hardware, capable of running over 100 billion deep learning operations per second, comes with sample projects, example code, and pre-trained models so even developers with no machine learning experience can run their first deep learning model in less than ten minutes. Developers can extend these tutorials to create their own custom, deep learning-powered projects with AWS Lambda functions. For example, AWS DeepLens could be programmed to recognize the numbers on a license plate and trigger a home automation system to open a garage door, or AWS DeepLens could recognize when the dog is on the couch and send a text to its owner.
- **Train models in the cloud and deploy them to AWS DeepLens:** AWS DeepLens integrates with Amazon SageMaker so that developers can train their models in the cloud with Amazon SageMaker and then deploy them to AWS DeepLens with just a few clicks in the AWS Management Console. The camera runs the models, in-real time, on the device.

"We've deepened our relationship with AWS, adding them as an Official Technology Provider of the NFL and are excited to use Amazon SageMaker for our next-generation stats initiative," said Michelle McKenna-Doyle, SVP and CIO, National Football League. "With Amazon SageMaker in our toolkit, our developers can stop worrying about the undifferentiated heavy lifting of machine learning, and start adding new visualizations, stats, and experiences that our fans will adore."

As the world's leading provider of high-resolution Earth imagery, data and analysis, DigitalGlobe works with enormous amounts of data every day. "DigitalGlobe is making it easier for people to find, access, and run compute against our 100PB image library which is stored in the AWS cloud in order to apply deep learning to satellite imagery," said Dr. Walter Scott, Chief Technology Officer of Maxar Technologies and founder of DigitalGlobe. "We plan to use Amazon SageMaker to train models against petabytes of earth observation imagery datasets using hosted Jupyter notebooks, so DigitalGlobe's Geospatial Big Data Platform (GBDX) users can just push a button, create a model, and deploy it all within one scalable distributed environment at scale."

[Hotels.com](https://www.hotels.com) is a leading global lodging brand operating 90 localized websites in 41 languages, "At [Hotels.com](https://www.hotels.com), we are always interested in ways to move faster, to leverage the latest technologies and stay innovative," says Matt Fryer, VP and Chief Data Science Officer of [Hotels.com](https://www.hotels.com) and Expedia Affiliate Network. "With Amazon SageMaker, the distributed training, optimized algorithms, and built-in hyperparameter features should allow my team to quickly build more accurate models on our largest data sets, reducing the considerable time it takes us to move a model to production. It is simply an API call. Amazon SageMaker will significantly reduce the complexity of machine learning, enabling us to create a better experience for our customers, fast."

Intuit recognizes the enormous value and power of machine learning to help its customers make better decisions and streamline their work, every day. "With Amazon SageMaker, we can accelerate our artificial intelligence initiatives at scale by building and deploying our algorithms on the platform," says Ashok Srivastava, Chief Data Officer at Intuit. "We will create novel large-scale machine learning and AI algorithms and deploy them on this platform to solve complex problems that can power prosperity for our customers."

Thomson Reuters is the world's leading source of news and information for professional markets. "For over 25 years we have been developing advanced machine learning capabilities to mine, connect, enhance, organize and deliver information to our customers, successfully allowing them to simplify and derive more value from their work," said Khalid Al-Kofahi, who leads Thomson Reuters center for AI and Cognitive Computing. "Working with Amazon SageMaker enabled us to design a natural language processing capability in the context of a question-answering application. Our solution required several iterations of deep learning configurations at scale using the capabilities of Amazon SageMaker."

"Deep learning is something that our students find really inspiring. It seems like every week now it is leading to new breakthroughs in robotics, language, and biology. What I like about AWS DeepLens is that it seems

likely to democratize access to experimenting with machine learning,” said Andrew Moore, Dean of the School of Computer Science at Carnegie Mellon University. “Campuses like ours are going to be really excited to bring AWS DeepLens into our classrooms and labs to help accelerate the process of getting students into real-world deep learning.”

New speech, language, and vision services allow app developers to easily build intelligent applications

For those developers who are not experts in machine learning, but are interested in using these technologies to build a new class of apps that exhibit human-like intelligence, Amazon Transcribe, Amazon Translate, Amazon Comprehend, and Amazon Rekognition video provide high-quality, high-accuracy machine learning services that are scalable and cost-effective.

“Today, customers are storing more data than ever before, using Amazon Simple Storage Service (Amazon S3) as their scalable, reliable, and secure data lake. These customers want to put this data to use for their organization and customers, and to do so they need easy-to-use tools and technologies to unlock the intelligence residing within this data,” said Swami Sivasubramanian, VP of Machine Learning, AWS. “We’re excited to deliver four new machine learning application services that will help developers immediately start creating a new generation of intelligent apps that can see, hear, speak, and interact with the world around them.”

- **Amazon Transcribe** (available in preview) converts speech to text, allowing developers to turn audio files stored in Amazon S3 into accurate, fully punctuated text. Amazon Transcribe has been trained to handle even low fidelity audio, such as contact center recordings, with a high degree of accuracy. Amazon Transcribe can generate a time stamp for every word so that developers can precisely align the text with the source file. Today, Amazon Transcribe supports English and Spanish with more languages to follow. In the coming months, Amazon Transcribe will have the ability to recognize multiple speakers in an audio file, and will also allow developers to upload custom vocabulary for more accurate transcription for those words.
- **Amazon Translate** (available in preview) uses state of the art neural machine translation techniques to provide highly accurate translation of text from one language to another. Amazon Translate can translate short or long-form text and supports translation between English and six other languages (Arabic, French, German, Portuguese, Simplified Chinese, and Spanish), with many more to come in 2018.
- **Amazon Comprehend** (available today) can understand natural language text from documents, social network posts, articles, or any other textual data stored in AWS. Amazon Comprehend uses deep learning techniques to identify text entities (e.g. people, places, dates, organizations), the language the text is written in, the sentiment expressed in the text, and key phrases with concepts and adjectives, such as ‘beautiful,’ ‘warm,’ or ‘sunny.’ Amazon Comprehend has been trained on a wide range of datasets, including product descriptions and customer reviews from Amazon.com, to build best-in-class language models that extract key insights from text. It also has a topic modeling capability that helps applications extract common topics from a corpus of documents. Amazon Comprehend integrates with AWS Glue to enable end-to-end analytics of text data stored in Amazon S3, Amazon Redshift, Amazon Relational Database Service (Amazon RDS), Amazon DynamoDB, or other popular Amazon data sources.
- **Amazon Rekognition Video** (available today) can track people, detect activities, and recognize objects, faces, celebrities, and inappropriate content in millions of videos stored in Amazon S3. It also provides real-time facial recognition across millions of faces for live stream videos. Amazon Rekognition Video’s easy-to-use API is powered by computer vision models that are trained to accurately detect thousands of objects and activities, and extract motion-based context from both live video streams and video content stored in Amazon S3. Amazon Rekognition Video can automatically tag specific sections of video with labels and locations (e.g. beach, sun, child), detect activities (e.g. running, jumping, swimming), detect, recognize, and analyze faces, and track multiple people, even if they are partially hidden from view in the video.

“At Isentia, we built our media intelligence software in a single language. To expand our capabilities and address the diverse language needs of our customers, we needed translation support to generate and deliver valuable insights from non-English media content. Having tried multiple machine translation services in the

past, we are impressed with how easy it is to integrate Amazon Translate into our pipeline and its ability to scale to handle any volume we throw at it. The translations also came out more accurate and nuanced and met our high standards for clients,” says Andrea Walsh, CIO at Isentia.

“RingDNA is an end-to-end communications platform for sales teams. Hundreds of enterprise organizations use RingDNA to dramatically increase productivity, engage in smarter sales conversations, gain predictive sales insights, improve their win rate and coach reps to succeed faster than ever before. A critical component of RingDNA’s Conversation AI requires best of breed speech-to-text to deliver transcriptions of every phone call. RingDNA is excited about Amazon Transcribe since it provides high-quality speech recognition at scale, helping us to better transcribe every call to text,” said Howard Brown, CEO and Founder at RingDNA.

“The Post strives to give its nearly 100 million readers the best experience possible and relevant content recommendations are a key part of that mission,” said Dr. Sam Han (PhD), Director of Data Science at The Washington Post. “With Amazon Comprehend, we can leverage the continuously-trained NLP capabilities like Keyphrase and Topic APIs to potentially allow us to provide even better content personalization, SEO, and ad targeting capabilities.”

“Building intelligent applications to help customers drive their businesses is our entire focus,” said Manjunath Ganimasty, V.P. Software Development with Infor. “Amazon Comprehend allows us to analyze unstructured text within search, chat, and documents to understand intent and sentiment. This capability enables us to train our Coleman AI skillset, and also provide a truly focused and tailored search experience for our customers.”

“Natural language processing is hard. We’ve looked at everything from closed to open-source solutions to analyze and make sense of our data, but couldn’t find a practical solution that would allow us to stay agile, scalable, and cost effective. Amazon Comprehend provides a continuously-trained model allowing us to focus on our business and innovate in Supply Chain Management (SCM),” said Minh Chau, Head of Engineering at Elementum.

“The City of Orlando is excited to work with Amazon to pilot the latest in public safety software through a unique, first-of-its-kind public-private partnership,” said John Mina Police Chief., City of Orlando. “Through the pilot, Orlando will utilize Amazon’s Rekognition Video and Acuity technology in a way that will use existing City resources to provide real-time detection and notification of persons-of-interest, further increasing public safety and operational efficiency opportunities for the City of Orlando and other cities across the nation. ”

“The analytic features of Amazon Rekognition Video are impressive. They can, for example, help with search of historical and real time video for persons-of-interest, providing efficiencies and awareness by automating this typically human task,” Dan Law, Chief Data Scientist at Motorola.

About Amazon Web Services

For more than 11 years, Amazon Web Services has been the world’s most comprehensive and broadly adopted cloud platform. AWS offers over 100 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, and application development, deployment, and management from 44 Availability Zones (AZs) across 16 geographic regions in the U.S., Australia, Brazil, Canada, China, Germany, India, Ireland, Japan, Korea, Singapore, and the UK. AWS services are trusted by millions of active customers around the world—including the fastest-growing startups, largest enterprises, and leading government agencies—to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit <https://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. Customer reviews, 1-Click shopping, personalized recommendations, Prime, Fulfillment by Amazon, AWS, Kindle Direct Publishing, Kindle, Fire tablets, Fire TV, Amazon Echo, and Alexa are some of the products and services pioneered by Amazon. For more information, visit www.amazon.com/about and follow @AmazonNews.Amazon.com, Inc.

View source version on businesswire.com: <http://www.businesswire.com/news/home/20171129006097/en/>

--30-- APS/SE

Contact:

Amazon.com, Inc.
Media Hotline, 206-266-7180
Amazon-pr@amazon.com
www.amazon.com/pr

Copyright Business Wire 2017
1.2

Industries: Technology, Data Management, Internet, Software, Audio/Video

Languages: English

Primary Identifiers: 09TGVY-E, AMZN-US

Related Identifiers: 09TGVY-E, AMZN-US

Source: Amazon Web Services, Inc.

Subjects: Product/Service, Photo/Multimedia