

# Driving Business Innovation

with Machine Learning and AI on AWS









# **Table of contents**

Introduction	3
Getting started in machine learning	3
Well-curated data is the foundation of success	5
Implementing AI in practice	6
Use cases: how ML looks in action	8
Conclusion	9
About AllCloud	9







# Introduction

Machine learning (ML) and artificial intelligence (AI) are already well known for their ability to significantly improve business efficiency. The problem has always been when and how to implement these advanced technologies in business operations.

In contrast to the last few years, new cloud-based ML services are now making it easier for companies to get started – even if they do not have any in-house data science resources.

# **Getting started in machine learning**

Many companies already know that using machine learning and AI algorithms can deliver additional efficiency potential – even in companies that are already highly digitalized and automated. However, the integration of this technology is still only slowly progressing. There are many reasons for this: the lack of clarity on the market, the scarcity of data scientists, and the huge variety of approaches and tools.

When the AI trend began a few years ago, it was often said that companies should just start by experimenting. Yet some companies trying to use AI soon found out that it's not as easy as they may have thought. Trying it out is vastly different to actually implement the technology and use it to its maximum advantage.

Thankfully, this is no longer the case. There are now scalable, flexible cloud platforms and solutions for machine learning services that make it significantly easier to implement ML in practice. Yet ML still needs the right processes and qualified service providers to implement applications successfully and select the right tool at the right time from the wide array of choices available. Managed service providers relieve the company IT department's workload so that its budgets and resources can be used for innovation. With cloud-based AI solutions, as with others, companies need to focus on smart cost optimization and control.





## **Definitions: AI, machine learning, deep learning**



Artificial intelligence (AI)

Al is technology that teaches a computer to imitate human behavior in some way.



### Machine learning (ML)

Machine learning is a subgroup of AI. ML includes any data-driven approaches to pattern recognition that can extrapolate forecasts based on historical facts.



Deep learning is a subgroup of machine learning and is based on artificial neural networks, such as those used in image and speech recognition or in robot control.

In this guide, we are using ML as an umbrella term for algorithms and models that are trained using underlying data; these often learn additional information on their own as well.

# **Avoiding typical pitfalls**

Due to their layered complexity, AI projects can easily fail if the necessary preparation is lacking. The first thing to do is carefully define the use case and clarify the question of whether it can actually be resolved using data. Things can often get difficult if there is anything wrong with the underlying data – and many companies have some catching up to do in this area. So-called "data tastings" or "data discovery" by specialists can help companies at the beginning by clarifying the data's condition.

One general rule-of-thumb is that an agile project methodology is best suited to the iterative approach common in the development of ML applications. Start with the problem at hand: identifying the biggest pain points and clarifying whether the relevant data is available, rather than starting from the Al approach as an end in itself – that would be the wrong way around.



## Well-curated data is the foundation of success

One thing all ML use cases have in common is that they need to be based on clean, error-free, complete digital data.

A good data foundation is not only the first hurdle but also the most difficult to overcome before machine learning can deliver important findings from the data across a range of different contexts. The good news: once this hurdle has been cleared, it gets much easier to implement additional plans going forward.



# Hunters and collectors: keeping everything in the data lake

A sensible strategy could certainly be to collect the information in a data lake and not "throw any away". However, this will only work if information is available as to exactly how and when the data was added and by which method. Unfortunately, there is no escaping the task of curating and preparing the data. Against this backdrop, it is particularly important to appoint a data owner with a view to data governance.

# It won't work without data governance

If a company's data has not been prepared properly and data governance is lacking, its data scientists will have a lot of work to do to curate it. It is not uncommon for 80 percent of the project effort to be invested in data preparation. Unfortunately, algorithms can't help here as its work depends on data quality and correctly annotated data. In practice, that remains one of the biggest challenges. In addition, data governance is not a just technical matter but a strategic-cultural one. Departmental silos still rule the roost in almost all industries. There must be a companywide recognition that data is not a by-product of processes but rather a central asset to be handled responsibly.

Data catalogs have become more popular, especially in the last two years, because they clearly show relevant data from other departments or external sources. The data is packaged up virtually, described, and placed in a sort of 'market square' to which access is regulated. For example, this enables users in the development department to access production data.



## Implementing AI in practice – quickly and efficiently

# Begin transparent data processes with a concrete use case

It regularly happens that companies already have a concrete idea about a problem that could be solved with AI. A specialized cloud service provider can help them get a feel for their data and set up a data governance process and the necessary quality measures. It makes good sense to have a specific use case as a starting point, which can also serve as a blueprint for additional application scenarios.

Based on AWS, companies can get started quickly by experimenting and testing a hypothesis with no additional investments. By means of an iterative process, they can then reach the stage of a minimal viable product (or model) – and if it doesn't work, they can quickly choose a different AI approach.

### Image and video data processing – A smart yet simple approach to AI

Amazon Lookout for Vision is a no/low-code solution for anomaly detection based on computer vision in image and video data. This significantly accelerates quality management processes in particular, and enables the automation of inspection procedures previously conducted manually. Alongside defects in materials or surfaces, the technology also detects things like missing components in modules.

Companies start by designing a GUI or API for this and uploading the images. Users can then indicate which of the pictures shows good or bad quality. The model is then trained automatically based on this input. The service takes over the time-consuming decision as to which parameters and algorithms are best suited to the task. It also includes the integration of data pipelines and data transformation.

> Although this approach costs a little more, it often enables companies to do without a data scientist, who is then free to focus on more complex tasks.

The service also includes automatic maintenance and upgrades and new versions are integrated seamlessly. With some of the Lookout for Vision services, customers can start with as few as 30 images.





# Platform for internal data products



The goal of this approach is to obtain a platform for the company's data products. Every step on the journey contributes to barrier-free access to data across all departments. AWS provides a good foundation for the creation of this type of platform. And specialized service providers support customers with various concepts and templates that help them select the most suitable technologies for them and build a platform corresponding to their requirements. Some of the typical components include the following:

#### AWS S3

Serves as a 'hard disk' in the cloud where companies can store their raw data. An API gateway can ensure data access.

### **Amazon Kinesis**

Captures and buffers real-time information like video, audio and IoT telemetry data used for machine learning and analytics.

### **AWS Glue**

A serverless integration service that makes it easier to find, extract, and prepare data as well as combine and integrate it for machine learning.

### **AWS Database Migration Services**

Designed to integrate existing on-premises data to provide faster access to it. The original database remains entirely operational during the migration process.

### Centralized data catalog (part of Amazon Glue or via other third-party providers)

Often called a "data supermarket", it provides centralized access to data for all staff, including data owners and how often the data has been sampled.

### AWS SageMaker

This is the domain of the data scientist, a superstructure where data is inputted and analyzed. The tool's visual interface helps users quickly prepare, develop and provision complex machine learning models.



# Use cases: how ML looks in action

### Tower semiconductor: quality recognition in chip manufacturing

Tower Semiconductor Semiconductor uses photolithography to print chips onto silicon wafers. The manufacturing process involves highly complex production lines and strict digital quality control.

The company was using digital methods to collect data, analyzing it against a set of rules, but had not yet tried machine learning. The manufacturer's goal was to implement a self-learning solution with Amazon SageMaker, and to achieve this, Tower Semiconductor engaged AllCloud.

For the self-learning solution, the wafer photos were captured on the production lines and shown to the algorithm. ML algorithm results improve over time, which is why a data augmentation process is used. This involves a range of techniques with which the size of a data set can be artificially inflated by transforming the existing data, helping to improve the accuracy of the model. While process optimization at Tower Semiconductor reduced the error rate to 21 percent, the use of ML achieved a significant further reduction to just a 2 percent error rate.



### EyeControl: Improving accuracy and speed in medtech

EyeControl has developed a wearable personal communication device for people whose medical condition hampers their communication ability. The EyeControl device is a head-mounted infrared camera that tracks eye movements and turns them into communication. The camera takes a picture of the eye and sends it to a deep learning model that classifies each movement. The company was already using a machine learning (ML) model, with a 92% accuracy rate, however they knew they could achieve even better performance.

With the help of AllCloud, the cost of ML training was significantly reduced by using Spot Instances. Fast iterations enabled effective tuning of the model to reach very high evaluation metrics.

Through fine-tuning, AllCloud eventually found the best combination that gave a very high accuracy score, significantly improving on EyeCloud's current accuracy rate, with a model weight of just 100 MB and a fast inference time of 4 frames per second.



### Conclusion

Machine learning is an enormous field to navigate. Even with advanced services that have a large amount of inherent know-how, companies should opt to work with an experienced partner to ensure that all the components are combined in the right way. To put it succinctly: Lego is great but it's still possible to build really ugly castles and dysfunctional death stars with it.

Are you ready to harness the power of your data and start your machine learning journey? Our AllCloud AWS Data team will help your company map out its current and future data and begin to show a path to leveraging Al/ML in your business.







**Contact us today!** 

# **About AllCloud**

AllCloud is a global professional services company providing organizations with the tools for cloud enablement and transformation. Through a unique combination of expertise and agility, AllCloud accelerates cloud innovation and helps organizations fully unlock the value received from cloud technology and data and analytics.

As an AWS Premier Consulting Partner, a Salesforce Platinum Partner and Snowflake Select Partner, AllCloud helps clients connect their front office and back office by building a new operating model that allows them to harness the benefits of cloud technology and data and analytics. AllCloud is supported by a robust ecosystem of technology partners, proven methodologies, and well-documented best practices. Thereby, elevating customers by achieving operational excellence on the cloud, within a secure environment, at every milestone of the journey to becoming cloud first.

With over 13 years of experience and a portfolio of thousands of successful cloud deployments, AllCloud serves clients across the globe. AllCloud has offices in Israel, Europe and North America.