aws

AI and Machine Learning on AWS

Confessions of a Data Scientist

Table of Contents

Foreword

If the 21st century has taught us anything, it's that **data is our most valuable resource**. Using insight-driven innovation, data is helping solve huge challenges across every industry.

As this world of data has developed, data scientists are **transforming humanity's future with ones and zeroes**. As always however, a professional is only as good as their tools.

More machine learning happens on AWS than anywhere else, with over 100,000 customers choosing AWS for its powerful compute, high-speed networking, and scalable storage options for any ML project or application.

With access to managed, modular services like Amazon SageMaker, a full suite of pre-trained AI Services—plus the ability to create a fully customizable AI/ML stack—customers use AWS to deploy models faster, improve productivity, reduce costs, and much more.

In this eBook, we will explore how some **data scientists are leveraging flexible, cost-effective technology**, frameworks, and infrastructure to streamline their AI and ML processes on AWS. Let's dive in…

Confessions of a Data Scientist

"When you start a business — especially a startup and machine learning base — you need collaborators that will run with you. Everyone will say 'no that can't be done', but the ideal partner will tell you that it can… AWS were exactly that type of partner."

Roy Cohen
CO-FOUNDER, NEUROSCIENTIST AND CONSUMER BEHAVIOUR EXPERT AT BEHAVIDENCE

"Initially there was no data. We needed the data, we needed a repository, we needed scripts and compute to run and assess the data. If we had to go off and build it on-premises with the budget we had, we wouldn't have been able to accelerate this fast."

Girish Srinivasan
CO-FOUNDER & CTO AT BEHAVIDENCE

behavidence

Transforming mental health with digital phenotyping & ML

Behavidence is a mental health navigation app, created to help users better understand their personal wellbeing while facilitating an unbiased approach to diagnosis and treatment. The app uses pre-trained digital phenotyping models — essentially the picture data paints of a user — to deliver easy-to-follow daily mental wellbeing scores.

Using AI/ML to enhance the future of mental healthcare

With one click, the Behavidence app gives users daily feedback based on the data behind their digital behaviour. They aim to empower users with the insights they need to live better — validating personal feelings while facilitating early mental health screening at a clinical level. With the right tools to drastically reduce false-positive diagnosis, Behavidence are using AWS to accurately recognise indicators for depression, anxiety, and Attention Deficit Hyperactivity Disorder.

Changing the way we understand mental health

The main challenge that Behavidence set out to solve stemmed from the conventional approach to psychiatric diagnosis and treatment. In traditional clinical settings, some researchers have observed a bias to overdiagnosis. For patients, the stress of the psychiatric appointment may magnify behaviour traits. This is compounded by the potential for flaws in questionnaire-focused diagnostic tools.

Given these limitations, Behavidence found that AI/ML algorithms have a useful role to play in offering another approach to predicting behaviour. So, they decided to leverage the universality of mobile phones, in combination with the security and scalability of AWS, to change the way we understand mental health.

Behavidence's goals

Behavidence wanted to become a clinical diagnostic tool within 5 years. To achieve this, they knew they'd need a platform that could handle storage, analysis and management of vast amounts of smart device data, enabling them to sort through the noise and generate robust psychiatric insights. For this reason, they chose AWS.

behavidence

Working with AWS

A framework for rapid development and evaluation

With an integrated pipeline of services including data orchestration, analysis, and storage, AWS offered Behavidence the tools they needed to optimise and scale their operations — forming a solid base to build their ML models.

Credits for prototyping and PoC development

Behavidence started out with a very limited budget, and AWS credits were critical to rapidly developing their tech stack and gaining traction as a business.

Secure development support with 24/7 service expertise

As a startup with key data privacy and security concerns, Behavidence worked with AWS to design an effective and efficient architecture that met all user and compliance regulations.

cainthus

Confessions of a Data Scientist

"AWS really helped us to develop quickly in the computer vision and ML space because of tools and services like SageMaker and AWS Step Functions that integrate with Metaflow."

Laura O'Malley

DATA SCIENTIST AT CAINTHUS

"The build versus buy conundrum is something that we faced a lot at the beginning. We wanted to focus on the challenges that were unique to our business, and with the great portfolio of services on AWS there was always something that we could choose instead of building a solution of our own."

Enrico Marchesin

PRINCIPAL SOFTWARE ARCHITECT AT CAINTHUS

cainthus

Confessions of a Data Scientist

"We designed most of our systems around the concept of "elasticity". From the very beginning we used solid decoupling mechanisms like SNS/SQS and Kinesis to interconnect independent parts of our processing pipelines. This has not only made the overall system more resilient, it also allowed development teams to move faster with greater autonomy and focus."

Jane Cummings

CTO AT CAINTHUS

Enhancing animal husbandry with AI-driven insights

Cainthus is an Ag-Tech startup who use computer vision and AI to continuously monitor nutritional, behavioural, and environmental activity on dairy farms. Cainthus' offerings, ALUS Nutrition and ALUS Behaviour, are designed to help farmers drastically increase profitability and sustainability at scale — all without losing a drop of milk.

Bringing a traditional industry into the future

AI/ML and Data Science are at the core of Cainthus' offerings, harnessing raw visual data from cameras deployed on edge devices in remote locations and processing it with AWS Cloud services. They use Deep Learning models and a variety of Computer Vision tools and processes alongside sophisticated AI techniques to continuously observe herd activity, transforming 24/7 observations into actionable real-time insights.

Helping farmers and livestock in equal measure

Balancing labour shortages, farm profitability, and feed management with animal health and welfare has always been a challenge for large scale dairy operations. Finding the right balance is a necessity, with farmers turning to Cainthus to increase milk production per cow by as much as 4.5% — ultimately driving significant reductions in waste and land use.

The vast majority of Cainthus' customers are located in remote rural areas that are prone to internet downtime, and so connectivity and reliability are key. With unstable connections that threaten to interrupt core functionality, the resilience of the AWS infrastructure is essential to ensure that both farmers and cattle stay on the AI/ML-powered grid.

Cainthus' goals

With significant requirements for collecting and storing data, training models, and continuously monitoring model performance, Cainthus needed as much consistency in their technology stack as possible. With this in mind, they built their service on AWS — using cloud-native tools and services to continuously develop and optimise their AI solutions.

cainthus

Working with AWS

Maintaining focus on their core solutions

Using consistent and reproducible patterns based on AWS services enabled Cainthus to increase overall productivity — allowing the creation of shareable assets across different AI/ML projects and accelerating the pace of innovation and delivery of features.

Scaling resources based on changing demands

Sudden spikes in data ingestion volume were to be expected with intermittent rural connectivity on farms, and so Cainthus looked to AWS for an immediately scalable solution.

Essential performance, storage, and connectivity

As a startup with a solution that spans Edge/IoT computing and the cloud, Cainthus rely on AWS to meet their significant data processing and storage needs.

Confessions of a Data Scientist

"As a start-up company looking to find product market fit, it's always been vital for us to be able to ship and iterate quickly. And that's something where having the reliability and the breadth of services that AWS offers has been essential for us."

Morgan Bruce

VP OF ENGINEERING AT ONFIDO

onfido

Enabling remote identification using computer vision

How do you remotely verify real people in the digital world? Onfido helps companies to see real identity — the humans behind the screens — using best-in-class AI monitored by world-class security experts. Customers can prove their identities with just a government-issued ID and selfie, so businesses can verify users without compromising on experience, privacy, or security.

Setting a new standard for digital access

As an automated identity verification platform, AI and ML are critical to Onfido's business model. Every time users take a photo or video of themselves to verify their identity, Onfido runs a complex series of automated AI/ML tasks, including: document verification, optical character recognition, biometric and face matching, and even tests for genuine "liveness".

Verifying identity in a remote world

Fraud is an ever-evolving threat, with complex evasion techniques ranging from spoofed selfies to deepfakes and 3D latex masks. With this in mind, maintaining the reliability of results while also ensuring a frictionless user experience is of paramount importance. Together with AWS, Onfido is able to guarantee the security that customers expect.

With the increasing need for remote identification, Onfido uses computer vision and AI/ML on AWS to streamline user experiences while securely exposing fraud. These technologies enable Onfido to detect whether documents, selfies, or even videos have been tampered with — matching users with their ID's to protect over 800 companies worldwide.

Onfido's goals

Using AWS as the platform for their application from the beginning, Onfido found that building their AI/ML and business intelligence features on the same infrastructure was natural fit. As a fast-growing global scale-up, they needed the ability to scale quickly and reliably — finding unmatched value in the breadth and depth of infrastructure and managed services that AWS provides.

onfido

Working with AWS

A scalable platform for innovation

By iteratively building their platform on top of AWS services like Amazon S3 and Amazon EKS, Onfido has been able to reliably scale their ML training in a way that would have been impossible on-premises.

Global availability and world-class reliability

The AWS global infrastructure consisting of multiple Availability Zones in different Regions across the globe was crucial element in meeting the expectations for guaranteed global uptime as well as compliance and regulatory needs of Onfido's customers.

The AWS well-architected framework

AWS' broad knowledge base of best practices, collected throughout the years and published through the Well-Architected Framework, helped Onfido to realise the full benefits of cloud computing, defining and building their ML stacks on a truly dependable infrastructure.

CLEO
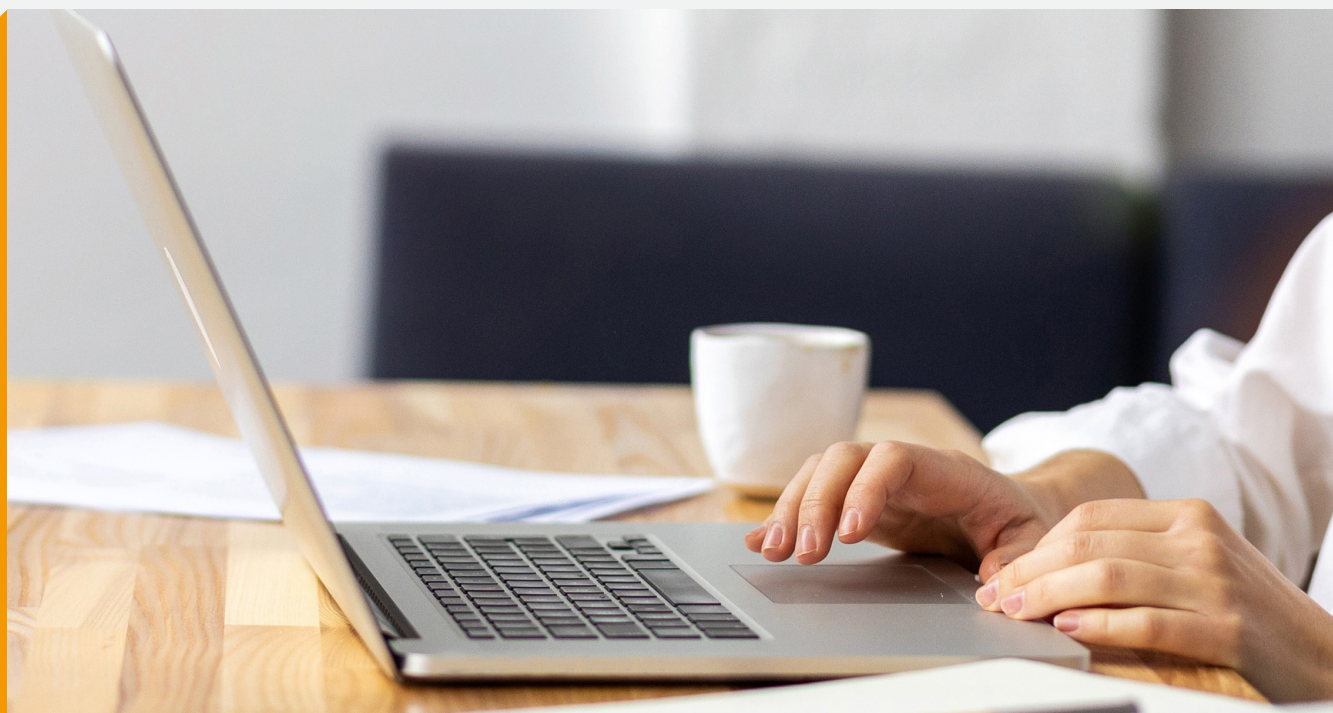
Confessions of a Data Scientist

"Scaling isn't just about having more computing power, its about choosing the right technology at the right time. [AWS] has that vast portfolio of technologies to choose from while you're growing as a company — and we've certainly taken advantage of that".

Sam Taylor

VP DATA AT CLEO

CLEO

A money app that's got your back

Cleo has global ambitions to fight for the world's financial health. Using simplicity and humour, Cleo has helped four million people improve their relationship with money. They're empowering users to make better financial decisions: spending well, building credit and saving what they can, while providing a holistic view of their finances.

Simplifying personal finance using ML

With the view that traditional banking apps can lack transparency and be difficult to navigate, Cleo takes a list of virtually indistinguishable transactions and categorises them using Machine Learning. Users get the insights they need to make better money decisions, delivered through a devastatingly funny chat interface, powered by Natural Language Processing (NLP).

Financial freedom at your fingertips

Trust is fundamental to Cleo's relationship with users, and (sometimes brutal) honesty has helped to build that.

Cleo found they could use Machine Learning to help users feel like they were talking to a friend, and they continuously iterate on the highly intuitive chat interface. This presents interesting language processing challenges. Users need an equally appropriate response, whether they're submitting informal, idiomatic phrases or more formal, financial queries.

Today, Cleo uses AWS' AI/ML services to improve customer insights, identify lending opportunities, and protect their users against organised fraud.

Cleo's goals

As a venture capital backed startup, Cleo needed to move quickly and start providing value to customers as soon as possible. To this end, AWS' comprehensive portfolio of services for AI/ML and data analytics enabled them to scale their business with agility and speed.

CLEO

Working with AWS

A platform for iterative business growth

As a rapidly growing business, Cleo has leveraged AWS' comprehensive portfolio of technologies and services in order to scale up and continuously improve their services.

Years of proven technical expertise

With a deep knowledge base of well-documented best practices, as well as the vast experience of trusted AWS Solutions Architects, AWS helped Cleo to shift focus from remedying technical issues to making the best possible decisions for their business.

Industry leading tools for innovation

One of Cleo's core business values is to always iterate with data. They use AWS extensively to store, transform and query their data to deliver better experiences to their customers.

Confessions of a Data Scientist

"We chose AWS because it really suits our platform. Data science tools are additional, so something that's easy to set up and integrate with our platform was always going to be key for us."

Jonas Kiiver

CTO AT MONESE

"The industry has seen a significant rise in cybercrime and fraud over the past year. So the focus across the industry is as always to protect legitimate customers while stopping those who want to cause harm… and our biggest weapon in that is data."

Alexandra Rowbottom

HEAD OF COMMUNICATIONS AT MONESE

monese

Driving financial freedom with data

As a leading pan-European fintech, Monese empowers ambitious people with untraditional work and income patterns. Offering easily accessible and personalised financial services, they aim to make it easy for customers to manage their money — all made possible with data science.

A data-driven approach to global finance

By adopting a data-driven and digital-first approach, Monese ensures that onboarding, account opening, payments, and transfers are safe and secure — while also meeting their core goals of accessibility and personalization. In doing so, they are able to deliver the efficient, quick, and low friction experiences that their customers have come to expect.

Balancing security with ease of access

Data science allows Monese to meet a two-sided challenge: on one hand it enables them to enhance ease of use and deliver valuable insights, using data to build a rich picture of customer finances both at home and abroad. It also plays a key role in their business differentiation, helping them to develop their pan-European and highly portable service.

On the other hand, Monese's open approach to finance creates new avenues for exploitation. As such, data and machine learning have become essential tools for them to combat fraud, helping to identify key markers for genuine customer behaviour. These markers and habits ultimately inform Monese's modelling — all of which is done on AWS.

Monese's goals

Many of Monese's data science challenges relate directly to personalisation and risk, this is so they can reduce friction and offer more meaningful services and features to their customers. Monese needed a cloud-based platform to integrate their data science models to ensure a high level of personalisation. With positive customer experiences remaining their number one priority, AWS tools and services have been key to making it work.

monese

Working with AWS

Faster, simplified deployments

Working with AWS helped Monese to integrate, deploy and execute their data infrastructure at a fast pace — freeing their infrastructure team to focus on new projects.

Rapid provisioning of new resources

With the ability to quickly provision resources on AWS, Monese was able to accelerate integration with their own platform and begin seamlessly run models on Amazon SageMaker.
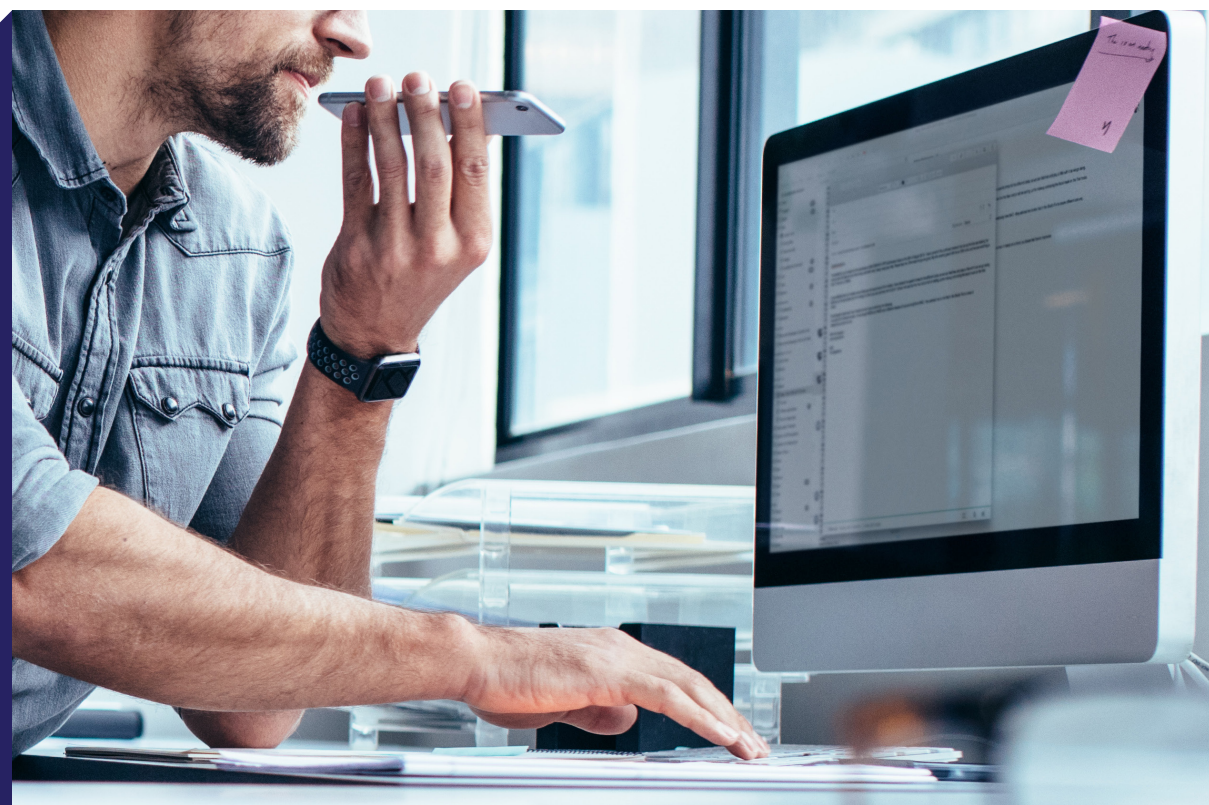
Closing Summary

Business transformation is driven by data

From revolutionising mental healthcare and improving agricultural productivity, to ensuring financial security and meeting the needs of a brand-new customer base, it is clear that data science is at the heart of a whole new wave of business transformation.

What links these businesses is their commitment to innovation, as well as the clear ability to grow and flourish on AWS. With access to a comprehensive portfolio of AI Services, ML infrastructure and powerful fully-managed services such as Amazon SageMaker — they are the pioneers of data-driven success.

AWS is proud to be part of this next generation and partner with a diversity of innovators, delivering flexible and highly reliable solutions across different industries and verticals.

AWS gives data scientists the AI/ML tools they need

Free to Try	Gain free, hands-on experience with the AWS ML services through our free tier and access the Founders tier which offers $1K for any bootstrapped startup.
Low Cost	With AWS, you only pay for the individual features you need, for as long as you use them, and without any long-term contracts or complex licensing.
Efficiency	AWS services improve team productivity by up to 10x because we automate all the heavy lifting required to manage and monitor ML infrastructure and security. Our ML services also offer collaboration tools and workflows.
Fast to Deploy	You can deploy ML models with a single click so you can start generating predictions quickly. Your ML applications are deployed onto auto-scaling Amazon ML instances across multiple Availability Zones for high redundancy.
Access to AWS Experts	AWS Activate is a free program for startups and early-stage entrepreneurs that offers tools and access to expertise so you can build using AWS best practices.
Proven	Companies around the globe like Lyft, Snapchat, Coinbase, and Stripe have launched and continue to grow on AWS.

Learn more about Machine Learning for Startups

If you have any questions or want further details, speak to a member of our team

Contact us

Join the AWS Startup Community

Visit the Startup Loft to get started and grow your startup with AWS

Sign up