Canonical

Make better decisions with open source big data and AI solutions

How to build a smarter enterprise with a secure, integrated open source stack

Executive Summary

What does it mean to be a smarter enterprise, powered by artificial intelligence (AI) and big data? The possibilities reach far beyond today's chatbots, self-driving cars, and product recommendation engines. McKinsey research suggests that 70% of companies will be using AI by the year 2030, adding trillions of dollars in value to the global economy.[1] AI will allow these organisations to speed up innovation and improve resource allocation. The pace of change promises to be breathtaking.

Behind the scenes, more sophisticated AI algorithms are expected to drive tremendous change across almost every industry. In retail, AI will be used for fully self-service stores, dynamic pricing, and virtual fitting rooms. In healthcare, AI will be a standard tool for diagnostic work; in finance, AI will transform everything from financial planning to wealth management; in government, AI will streamline processes for smarter cities; and in telco, every service will be personalised.

The magic is in the melding of AI and big data. Data of incredible volume, velocity, and variety is fed into the AI engine, making the AI smarter. Then, less human intervention is needed for the AI to run properly. Finally, the AI can deliver deeper insights—and strategic value—from the ever-increasing pools of data, often in real time.

But getting to this AI-driven future is not without significant challenges. With concerns ranging from skills gaps to tooling complexity to time-to-market pressures, organisations increasingly need an efficient, integrated way to experiment with AI—and move the projects into production. The underlying hardware, application layer, and data fabric all need to work together as efficiently as possible to deliver business value. That's where open source solutions, supported by trusted vendors, come in.

This white paper will explore the growing importance of big data in the modern enterprise, and how it's increasingly used with AI to solve the toughest problems. We'll look at how data scientists and AI developers are working together to build AI applications in different industries worldwide. Finally, you'll learn how a secure, end-to-end open source stack can help your organisation deploy AI efficiently and drive business growth.

1 https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-organization-of-the-future-enabled-by-gen-ai-driven-by-people

Contents

Big data in the modern enterprise

For years, organisations have been putting data to work to make smarter decisions. But what does “big data” really mean in the modern era? As computing power and storage capacity have increased in efficiency and decreased in price, companies have been able to store, analyse, and leverage more data than ever before. In fact, what was considered “big data” just a few years ago might not be as big and daunting now.

A good rule of thumb is that “big data” refers to data that exceeds the capacity of a single computer, both in terms of computational power and disk storage. Similarly, the types of data have also increased dramatically. In the past, businesses mainly worked with tabular data; now, companies are developing insights from camera footage, text data, geospatial positioning data, and online customer behaviour data, just to name a few.

The modern enterprise can use AI algorithms and models to learn from these treasure troves of big data, and make predictions or decisions based on the data without being explicitly programmed to do so. What’s more, the AI models grow more accurate over time.

It all adds up to AI and data strategies needing to be more interconnected than ever. According to an MIT Technology Review survey, 78% of CIOs say that scaling AI to create business value is the top priority of their enterprise data strategy, and 96% of AI leaders agree. Nearly three out of four CIOs also say that data challenges are the biggest factor jeopardising AI success.[2]

Online retail is one of the most advanced sectors when it comes to big-data analytics and AI. Today’s consumers expect a personalised experience, complete with tailored product recommendations and a streamlined shopping journey.

2 https://www.databricks.com/lp/mit-cio-industry-summary

In a recent survey, Salesforce found that 73% of consumers expect companies to understand their unique needs and expectations, and 56% expect all offers to be personalised.[3]

Meanwhile, outside of e-commerce, many in the business world are not aligned on the best way to collect data and use it in AI initiatives, or how to manage customers' personal data in a secure and compliant manner. The regulatory environment will likely evolve by 2030. But well before then, organisations need a way to empower data scientists and AI developers to collaborate effectively, iterate faster, and produce smarter models to create lasting value.

The perfect match: Big data and AI models

Today's companies have immense volumes of data at their fingertips to power AI applications that accelerate growth, fuel innovation, and improve customer experiences. The speed of innovation even comes down to the data itself. Gartner predicts that by 2025, the use of synthetic data will reduce the volume of real data needed for machine learning by 70%.[4]

Often, data scientists will systematically interact with datasets to obtain better results and increase the accuracy of their machine learning (ML) applications. This may involve using different data samples for training or increasing the size of the dataset. As a result, enterprise AI projects often involve many petabytes of data. Big data technologies are a natural fit for large-scale data preparation—alignment, cleansing, labelling, and so forth—that is necessary to ensure high-quality input data for ML models.

By using big data with their ML initiatives, organisations can:

- Accelerate training times for ML models
- Improve overall efficiency of model development
- Generate predictions from very large, pre-existing batch datasets
- Deliver insights in near-real time from continuous streams of data
- Respond faster to change with more informed decisions

Barriers to big data and AI adoption

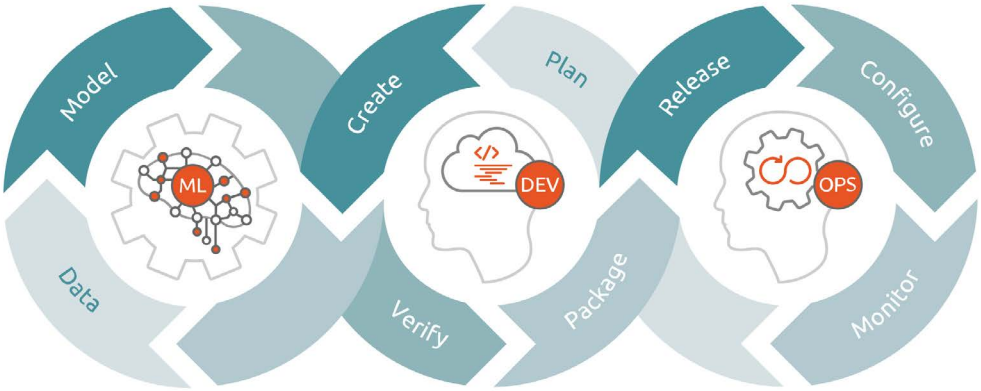
Despite the many benefits of using big data with AI, companies are still faced with significant barriers to adoption. These chiefly relate to the operational challenges of deploying emerging technologies into today's enterprise IT environments, where flexibility, transparency, and robust security measures are a business imperative. Key barriers include:

- **Skills gaps:** AI projects require a multifaceted team of data scientists, ML engineers, software developers, IT operations staff, and solution architects. The rate of technological change can rapidly make in-house skillsets obsolete. It can also be difficult to attract and retain personnel with deep knowledge of the platform services of specific cloud service providers.

3 https://www.salesforce.com/eu/resources/research-reports/state-of-the-connected-customer/
4 https://www.gartner.com/en/insights/generative-ai-for-business

- **Ecosystem complexity:** Big data management has traditionally been supported by open source initiatives based on Apache Hadoop. But recently, the role played by Apache Hadoop has been increasingly filled by cloud-ready solutions and public cloud services that are less cumbersome. Meanwhile, many enterprise users now find themselves struggling to deliver on the promise of cloud-based operational efficiency, information security, and cost predictability. Data scientists and ML engineers just need access to the right tools to do their job, without needing expertise in DevOps platforms.

- **Time to market:** Many organisations look to cloud deployments for convenience, ease of use, and faster time to market for AI initiatives. Cloud service providers eliminate the upfront cost and lead times for acquiring the hardware needed for big data and AI. But what happens if you want to move successful AI experiments into production on-premises? Cloud platforms often have proprietary features that are "sticky" and tie you to a specific vendor. To speed time to market and ensure portability, adopting a multi-cloud strategy with related technologies can be the perfect solution.



How to reap the benefits of big data and AI

With well defined practices and end-to-end tooling, enterprise teams can spend less time on operational tasks and focus on rapidly creating and testing the data-driven AI models that bring value to the business. In fact, research shows that organisations are putting substantially more AI models into production—growing by 411% between 2022 and 2023—while also increasing their ML experimentation by 54%.[5]

Open source software has been a popular solution for enterprise data management for years, and now, more ML algorithms and software libraries are available to developers at no cost. So, developers can experiment freely with big data and AI open-source projects, with no software licence fees. This is especially beneficial if an organisation runs multiple data processing environments at scale. What's more, strong communities often develop around different open source software projects, reducing the risk of undetected flaws and expanding the opportunities to share expertise.

Fully integrated, open source platforms are also accelerating the acceptance of machine learning—and reducing the perceptions of complexity. Developers are more likely to experiment with free software because they have virtually nothing to lose in trying it out. Meanwhile, the integrated solutions reduce the cost of operations, improve IT security, and enhance support.

5 https://www.databricks.com/resources/whitepaper/mit-cio-vision-2025

As mentioned earlier, a multi-cloud approach can both speed time to market for AI initiatives, while also ensuring portability across cloud environments—whether public clouds or private on-premises solutions. The multi-cloud strategy dissociates the underlying infrastructure from the software. As a result, it minimises dependency on a single vendor while maintaining choice and flexibility.

Industry use cases for big data and AI

Organisations across every sector are deploying big data and AI initiatives. While there are common elements across these use cases, there are also key differences in their goals and applications of the AI/ML technology.

Thanks to the hype around ChatGPT, organisations in just about every industry are exploring new ways to use large language models (LLMs) to deliver value. LLMs are deep learning models that specialise in understanding natural language. They are pre-trained on massive datasets and can speed up all kinds of internal processes, from repetitive clerical tasks to code generation to basic copywriting. Of course, LLM-based chatbots are the hottest use case, as organisations look to create their own private ChatGPT. It almost goes without saying that LLM-based chatbots are far more advanced than standard chatbots. To achieve better performance, they are trained on much larger datasets. And they are designed to understand the context behind "chunks" of language. So, the LLM-based chatbot can maintain a brand's personality and tone, respond to customers in different languages, and engage with them in a multi-touch dialogue—such as by providing multi-step troubleshooting guidance.

Government

In the public sector, AI-driven automation can streamline a wide range of workflows, from approvals of government payments to online permit processing to public health inspections. Operational efficiency is one of the leading priorities for government agencies of all sizes. But AI can also help with proactive equipment maintenance, smart traffic management, sustainable energy distribution, and public safety resource allocation.

When it comes to LLMs, government agencies can use them to extract knowledge from virtual mountains of government paperwork, whether it's for ensuring regulatory compliance, supporting the intelligence community, or summarising reports for the general public. Custom LLMs can help agencies in modernising their legacy code bases. Plus, LLMs can serve as a virtual instructor for employees, assist in human resources processes, and much more.

The use of LLMs and other emerging systems in government must go hand-in-hand with the use of strong data security controls. On-premises or private cloud-based deployments are required to keep sensitive information safe. These so-called sovereign clouds help reduce the risk of data breaches or unauthorised access.

In addition to citizen data, smart cities rely upon the data generated by Internet of Things (IoT) devices and sensors, often enriched by AI and ML applications, to predict maintenance needs to drive efficiency. They can do things like synchronise stoplights to reduce traffic congestion, guide drivers to available parking spaces, and optimise energy consumption to improve air quality across an entire region.

AI enables governments to simplify and optimise complex operations, while focusing resources on the tasks that matter most. For example, governments can leverage AI to provide more effective health services, from monitoring the spread of disease to tracking patients' health to optimising the distribution of medicine and patient care.

Finance

Digital innovation is also transforming finance. Advances in financial technology, such as mobile money, peer-to-peer (P2P) or marketplace lending, robo-advice, and insurance technology (InsurTech) are reshaping a wide range of areas—from everyday payments to wealth management. Data-driven AI solutions can be leveraged across all of these activities to improve how financial institutions respond to changes in customer needs and expectations.

In the financial sector, the classic use of AI is for credit risk analysis, which involves detecting cash flow and managing the risk associated with loans. Predictive analytics can help financial institutions evaluate a company's historical data to better understand the current and future health of the company. LLMs are also invaluable in the fight against financial crime. They can analyse past fraudulent activities, understand the context, and predict future fraud schemes.

Financial institutions are also able to use AI and big data for high-frequency trading, where large numbers of stock orders are executed within fractions of a second. Machine learning models use historic data and investment principles to predict the best times to place trades. They identify the situations when prices are most likely to increase; then, automated processes can place the trades accordingly.

Telco

In the telecommunications industry, AI solutions are accelerating innovation and driving down the costs of operations. LLMs can help service providers identify patterns of usage, predict network congestion, and enable proactive network management—ensuring a reliable experience for customers..

Predictive maintenance is also really important for telco operators, who get millions of alarms from their radio, transport, and core networks—often from various equipment types and vendors. The main promise of predictive maintenance is to allow convenient scheduling of corrective maintenance and prevent unexpected equipment failures. By understanding the actual condition of equipment, ML models forecast the degradation of an item and when intervention will be needed. In addition to cost savings, the predictive approach helps increase safety, reduce the environmental impact of accidents, and optimise spare parts handling.

In today's mobile networks, AI is also used a lot in a security context. For example, telco companies use AI to help prevent the fraudulent use of mobile subscriptions in SMS spamming, phishing attempts, or impersonations of other subscribers. In addition, telco companies use customer segmentation and behaviour analysis to help predict potential customer churn—and develop ways to mitigate it.

Retail

From brick-and-mortar stores to online marketplaces, retail companies are all increasing their investments in AI, in order to gain a competitive advantage, better understand their customers and solve some of their long-lasting problems. Unlike some other sectors, retailers have quickly moved towards a data-driven approach, using streams of data for improved speed, efficiency, and decision-making. Retailers can use LLMs and data-driven AI solutions to:

- Create enjoyable, personalised experiences to build stronger relationships with customers and increase loyalty.
- Get a better grasp of customer behaviour and expectations to improve forecasting, offer more attractive pricing, and optimise product placement.
- Automate inventory management with more visibility into products that are in stock, while also optimising interactions across the supply chain.

Large retailers are also using AI with market basket analysis, helping to uncover associations between items that occur frequently in transactions. By understanding what products are bought together, retailers can develop more informed promotions, and learn how to build excitement with the customer base.

Start your big data and AI journey with open source

To make the most of big data and AI initiatives, enterprises need dedicated, accelerated infrastructure designed for the intensive AI workloads. An effective ML stack needs to include hardware, software, and data fabric layers that are individually sophisticated, but also integrate and work together to unlock their respective potentials.

Hardware infrastructure

Hardware is at the foundation of a successful data-driven AI initiative, providing the compute, storage, and networking capabilities necessary to support the development and operation of ML models. ML projects are characterised by exceptionally large data volumes and computationally intensive workloads, which lead to unique hardware requirements across three key areas:

- **Compute:** Accelerated computing power is necessary for both efficiently training ML models and running them in production. The infrastructure must be capable of rapidly processing large datasets and performing highly complex mathematical operations, sometimes in real-time.
- **Storage:** To support the vast data volumes required to train accurate models, ML infrastructure needs cost-efficient memory, fast data access, and online storage systems optimised for large-scale data storage and retrieval.
- **Networking:** Moving data across different platforms within the AI stack and a company's wider IT ecosystem demands specialised high-speed networking and data transfer technologies.

Software infrastructure

The operating system (OS) provides the foundation for all the other elements in the ML stack, and acts as the bridge that enables the hardware and software layers to interact. The OS plays a key role in managing hardware resources; it supports the programming languages, libraries, and frameworks used in AI projects; and it is central to the overall security of the stack.

Ubuntu is the first-choice OS for countless AI developers due to its reliable release cadence, robust security, stable hardware drivers and widespread community adoption. Each Ubuntu LTS brings consistent and continuous bug fixes and security patches, with the option of 24/7 enterprise-grade support from Canonical.

Running AI on Ubuntu goes beyond your workstation. There is a wide variety of applications that are made simpler with open source. Using available tools and models, your teams can experiment quickly and take projects to production cost-effectively. For example, Kubernetes is an open source, cloud-agnostic platform designed for container orchestration. It's used for efficient resource allocation and scaling of ML workloads across public and private clouds.
With Canonical Kubernetes distributions, organisations can benefit from out-of-the-box cloud integration with the option of enterprise-grade support. MicroK8s from Canonical are a low-touch distribution of Kubernetes, designed for easy clustering, high availability, and streamlined upgrades. They are secure by default. Charmed Kubernetes provide extensive tooling for fully customisable deployments. Tools for AI, such as Kubeflow and Spark, can then run on top of Charmed Kubernetes for total cluster control at scale.

Databases

To start building data-driven AI applications, enterprises need to have a solid plan for managing the data. Canonical's data portfolio offers an enterprise-ready suite of popular open source data management solutions. The suite makes it easy to build, deploy, and run data-intensive applications with Apache Spark, Apache Kafka, MongoDB, OpenSearch, MySQL and PostgreSQL.

Each Canonical data solution is backed by 10 years of break/fix support, CVE patching, and a guaranteed response time SLA. Plus, they include advanced automation for common operational tasks, such as scaling, monitoring, upgrading, and securing the platform.

Canonical's data portfolio gives organisations an easy way to consume open source AI applications via Juju charms. Charmed Spark, for example, delivers a Canonical-supported distribution of Apache Spark to simplify batch data preparation and feature extraction, streaming data routing, and predictive inference at scale. In a similar fashion, Charmed Kafka for Apache Kafka simplifies the development of distributed, parallel data processing applications.

MLOps tooling

Machine Learning operations (MLOps) tools are designed to help enterprise AI teams work effectively at scale without being slowed down by low-value tasks. The strategic goal is to improve agility, accelerate outcomes, and reduce costs.

Built on top of Kubernetes, Kubeflow is one of the most widely used MLOps platforms, due to its comprehensive tooling, customizability, and large community of contributors.The Kubeflow suite of tools easily integrates with other open source ML tools. While there are many Kubeflow distributions, enterprise use cases require an enterprise-grade, trusted platform, such as Charmed Kubeflow from Canonical. In hybrid and multi-cloud deployments, Charmed Kubeflow delivers the same elasticity that enterprises have come to expect from public clouds. As a result, data scientists can effortlessly scale their AI experiments to thousands of jobs.

Find the right tools from Canonical

A secure, end-to-end open source AI stack enables enterprises to maximise AI efficiency and performance—moving models from concept to production with confidence. Technology integration means that the AI development ecosystem "just works" at scale, right out of the box.

For enterprise AI solutions, explore a plug-in architecture that can adapt to different needs, depending on the use case. Your teams need to be able to run workloads anywhere, including hybrid and multi-cloud environments. And don't forget the need for robust security, simplified operations with lifecycle management and automation, and simple per-node subscriptions

To innovate at speed with best-in-class open source AI solutions, consider the trusted technology from Canonical:

- **Charmed Kubeflow:** A production-grade MLOps platform to manage the end-to-end ML lifecycle.
- **Charmed Spark:** A security-maintained and fully supported solution for Apache Spark on Kubernetes.
- **Charmed MLFlow:** A platform used to manage ML workflows, used primarily for model registry.
- **Charmed OpenSearch:** Enterprise OpenSearch solution with support, security maintenance, and operations automation.
- **Charmed MongoDB:** Enterprise MongoDB solution with support, security maintenance, and operations automation.

Next steps

To learn more about building a smarter enterprise with open source AI solutions, please visit: ubuntu.com/ai and canonical.com/data

To discuss how we can support your team, please get in touch with us.

References

1. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-organization-of-the-future-enabled-by-gen-ai-driven-by-people
2. https://www.databricks.com/lp/mit-cio-industry-summary
3. https://www.salesforce.com/eu/resources/research-reports/state-of-the-connected-customer/
4. https://www.gartner.com/en/insights/generative-ai-for-business
5. https://www.databricks.com/resources/whitepaper/mit-cio-vision-2025

Canonical