AI you can trust

An end-to-end ML development pipeline for trustworthy AI

HPE trustworthy AI principles[1]

Privacy-enabled and secure — Respect individual privacy and be secure

Human focused — Respect human rights and be designed with mechanisms and safeguards to support human oversight and prevent misuse

Inclusive — Reduce harmful bias and support equal treatment

Responsible — Be designed for transparent and accountable use, inform an understanding of the AI, and enable outcomes to be explained

Robust — Be engineered to allow for quality testing, include safeguards to maintain functionality, and help minimize misuse and the impact of failure

Trust and quality integrated into an end-to-end MLOps architecture for compliant AI

Arising challenges for artificial intelligence

Trust

Artificial intelligence (AI) that makes statements and decisions is now used in almost all industries to support people and businesses. Applications range from image classification in healthcare to predictive maintenance in manufacturing. At the same time, the widespread use of AI technologies leads to new challenges in terms of security, privacy, robustness, explainability, and inclusion. Customers, suppliers, management, and product owners demand transparency and want to understand the AI models being used.

Compliance

The European Union is currently working on a regulation on AI rules. Transparency, trust, and explainability are important criteria for the European Union institutions, especially for the use of high-risk AI.[2]

[1] "What is Trustworthy AI and why is everyone talking about it?" HPE blog, 2021

[2] "Regulatory framework proposal on artificial intelligence," European Commission

Benefits

End-to-end MLOps for faster ML development
HPE Machine Learning Data Management Software reproducible training pipelines work hand in hand with the ability of HPE Machine Learning Development Environment to configure training environments through industry standard XML files. It allows ML developers to achieve faster time-to-market while working in a defined process.

Improved resource usage
HPE Machine Learning Development Environment manages available resources in an efficient way, resulting in reduced costs. Hardware pools can be established to balance compute performance and collaboration within the cluster. IT administrators can set up, manage, secure, and share AI compute clusters.

Fast and reliable explainability
With TruEra, ML developers get the broadest and deepest view into model performance. TruEra reduces troubleshooting time by pinpointing exactly where the problem is by reliable analyses. This helps data teams to comply to regulatory guidelines such as the AI Act of the European Union.

Integration with popular frameworks and platforms
Interoperability is very important as the ML landscape is highly fragmented and fast evolving. Data scientists can use their favorite frameworks and libraries. This architecture can be deployed on the Kubernetes provider of your choice with the help of HPE Services.

Trust by design
The architecture tackles the previously mentioned trust and ethical concerns by establishing a secure quality management for data and AI models. Secure data storage, monitored AI training, and documented training pipelines build trust through transparency and holistic analyses. The results of analyses can be evaluated in an intuitive way leading to informed and correct actions of the ML developers.

Explainability

Many powerful AI applications rely on complex but opaque models, which include neural networks. These models are not transparent, and their decision-making process is not traceable out of the box. Therefore, there are various methods that make models and their decisions explainable for data scientists, decision-makers, and users. In addition, the model and data must be evaluated to detect bias or other issues and be able to take countermeasures.

Operationalization

Modern AI development takes place in automated MLOps pipelines, where the process steps are orchestrated from raw data storage to deployment and retraining. To develop trustworthy and responsible AI models, an automated and scalable end-to-end training pipeline with integrated explainability and continued assurance of data and model quality with respect to the above criteria is required.

Introducing the end-to-end ML development pipeline for trustworthy AI

The following architecture enables machine learning (ML) developers to build ML models in a streamlined and continuously trusted way. Multiple proven components for different tasks along the ML development process are integrated and tested in a coherent architecture to simplify configuration and management. The whole architecture is designed to be consumable as a service on HPE GreenLake.

The entry point for the data is **HPE Machine Learning Data Management Software**. HPE Machine Learning Data Management Software is used for both data processing and workflow management. It offers automated data versioning and data processing pipelines. Furthermore, it triggers training steps automatically for increased productivity and reduced risks. Data-driven MLOps pipelines enhance end-to-end reproducibility yielding more trust.

After each training, **TruEra** Diagnostics checks the model to detect model quality issues regarding generalization, bias/fairness, robustness, performance, reliability, and many more. Plus, it makes opaque-box models explainable. The analysis results are presented in drill-down dashboards. But AI monitoring does not end when models are deployed. With TruEra Monitoring, AI models continue to be monitored when they are in production. Real-time analysis and alarms help data scientists to take the right action quickly.

The data is stored on **HPE Ezmeral Data Fabric**. This is a high-performance software-defined storage that is designed for analytics workloads on traditional and container-based environments. It enables trust and consistency by being able to unify different datastores with a single global namespace. It helps ensure global data access through various integrations including NoSQL queries and connectors.

The whole architecture is intended to run on state-of-the-art **HPE ProLiant Gen11** servers with **Intel®'s** 4th Gen Intel® Xeon® Scalable processors.
This hardware builds a strong foundation for secure, energy-efficient, and performant workload implementation. Sapphire Rapids CPUs provide a built-in AI acceleration engine delivering a significant leap in performance. This can be leveraged by the **Intel AI Analytics Toolkit** with optimizations on Intel CPUs for

various ML frameworks (including TensorFlow™, scikit-learn, PyTorch) resulting in faster ML development and inferencing. HPE ProLiant Gen11 servers support data-first modernization by bringing an intuitive cloud operating experience, security by design, and acceleration for next-gen workloads.

The **HPE Machine Learning Development Environment** is built upon the successful open-source Determined AI software. The HPE Machine Learning Development Environment lets ML engineers focus on building better models instead of managing IT infrastructure. This enables ML developers to:

- Train models faster using state-of-the-art distributed training, without changing their model code
- Automatically find high-quality models with advanced hyperparameter tuning
- Track and reproduce their work with experiment tracking that works out of the box, covering code versions, metrics, checkpoints, and hyperparameters

The enterprise machine learning platform for trustworthy AI

Component Stack

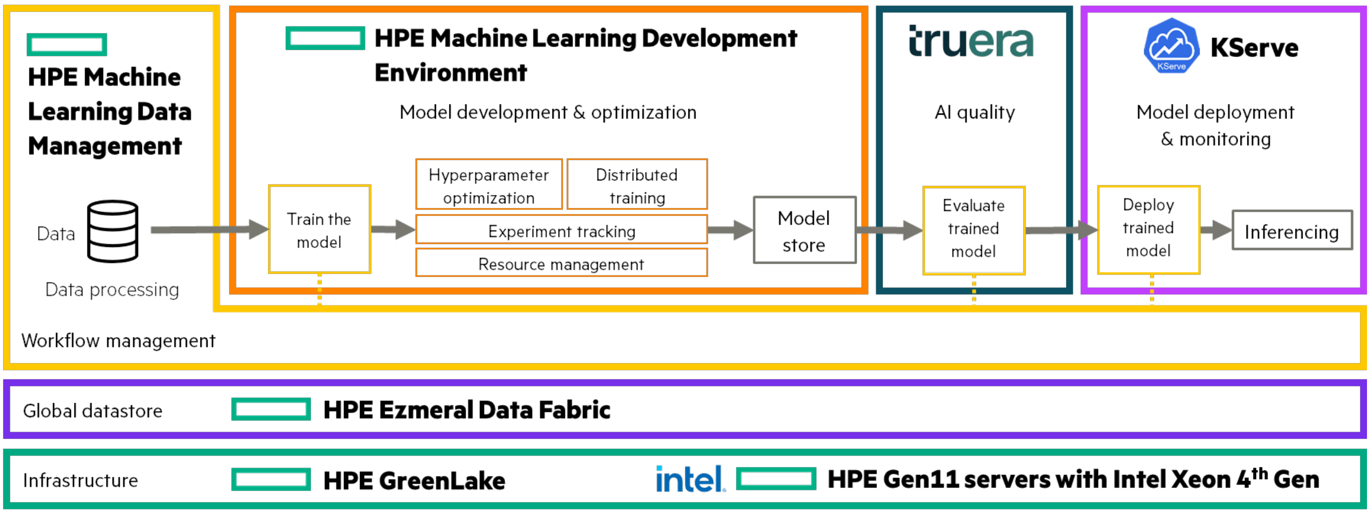

Figure 1. Architecture enabling end-to-end ML pipeline for trustworthy AI

Learn more at

HPE.com/info/ai

Chat now (sales)

Hewlett Packard Enterprise

a00130566ENW, Rev. 1