



AI-DRIVEN MANUFACTURING: AN ENTERPRISE APPROACH TO DISCOVER POSSIBILITIES

Abstract

Enterprises are building machine learning (ML) capabilities to augment their digital business scenarios. However, machine learning projects implemented by manufacturing enterprises register a high failure rate due to various reasons, including: setting wrong goals, ignoring business principles, improper investment of funds and resources, lofty expectations of quick success, inadequate project governance, deviation of talent development from the actual needs, flawed project roadmaps, adopting the wrong technologies, lacking understanding of the data required for the scenarios, poor data quality, etc.

In this point of view, Infosys experts analyze how common issues affect machine learning projects and recommend an enterprise architecture (EA) approach to prevent failure. Our experts explain general and federated machine learning with use cases from the capital goods industry. In addition, they discuss how shifts in technology such as generative AI are influencing architecture evolution as well as the enterprise architecture methodology at the algorithm level of machine learning.

Common issues in ML projects

Enterprises are gaining tangible value by adopting ML-related solutions, including use cases of existing ML services such as product recommendation, facial recognition for multi-factor authentication, vehicle registration number recognition for parking service, Chatbots, and emerging industry specific applications based on generative AI.

1. Setting wrong goals: AI technologies replace human effort in many tasks, but AI models need to be trained. Data collection and preprocessing for training AI / ML models is labor-intensive. In addition, human effort is needed to fine-tune the systems for proper functioning. So, enterprises that set the goal of saving labor costs may not achieve it.

2. Ignoring architecture principles: The expectation of high production rates and significant cost savings leads to quality loss and safety risks. Similarly, leakage of sensitive data and practices that violate moral customs or cause legal issues, such as improper use of private data and social prejudice or discrimination, adversely impact customer trust and corporate brand value.

3. Improper investment of funds and resources: When capital investment is inadequate or there is no long-term goal for cultivating capabilities, the development and deployment of AI platforms and capacity building are difficult, resulting in defects in capabilities, and the expected output cannot be met in the long run.

4. Over expectation of a speedy success: Enterprises that set machine learning models as short-term investments or pursue unreasonably quick results fail to effectively understand and utilize the potential knowledge and skills of machine learning models, thus negatively affecting investment results.

5. Diverting talent development from actual needs: Enterprises start the ML capacity planning without identifying clear business drivers and goals which align to the digital transformation strategy.

6. Flawed project roadmaps: Enterprises that embark on projects without effective analysis of the target architecture and planning, waste time and resources on ineffective tasks such as data collection and framework establishment.

7. Lack of project governance: Failure to promptly detect deviations from principles and compliance mandates, lack of effective measurement methods to monitor achievement of established goals and requirements, and the inability to dynamically track the latest trends and implementation of adjoining technologies affect ML projects.

8. Adopting incorrect technologies: When the adopted technology platforms don't match with the trends or the enterprises' internal IT environment, especially in complex situations such as B2B integration, the investment in ML project leads to loss of time, money, or business opportunities.

9. Inadequate understanding of the data required for developing scenarios: Before starting the projects, research on the data should be carried out to validate if the data contains the patterns to support the scenario. Otherwise, the desired results will not likely come out.

10. Poor data quality: Ineffective data governance leads to subpar data quality, which prevents enterprises from achieving the defined targets of business scenarios.



An enterprise approach for building ML capacity

An enterprise architecture approach can effectively mitigate common issues in ML projects. This approach spans analysis of the enterprise business strategy, value stream and architectural principles (reference: TOGAF Standard, Version 9.2). It enables enterprises to find answers to the following questions: What are the business problems to be resolved? How can end-to-end value be generated? What are the gaps in business and IT capabilities? What are the costs and benefits? What is the roadmap? What are the available technical solutions?

Let us take a use case from the manufacturing industry to understand the EA approach.

Prism Alloy Machinery (PAM) is a machinery and equipment manufacturer. Its clients are from all over the world, and most of them use maintenance services offered by PAM. Three years ago, PAM started to use machine learning technology to augment value and reduce operating costs.

After several failures, PAM adopted the EA methodology to ensure that the capacity building strategy was in the right direction. The CIO authorized the architect team to start the 'Smart Work' program to identify opportunities and develop plans to transform the enterprise architecture for realizing business goals. The architects developed a plan in collaboration with Infosys AI-ML experts:

At the outset, establish, review, and update architectural principles and the business strategy.

EA principles focus on compliance with regulations, protection of private data, data quality, safe production, and safeguarding corporate reputation. Activities such as value stream analysis, demand analysis, and architecture design should abide by EA principles.

One of the principles require technology solutions to ensure reliable and efficient production. This principle often guides decisions to adopt MLOps architecture.

These principles also require enterprises to track technology trends and adopt appropriate technology routes to achieve goals. PAM used a federated AI model training platform to support the equipment maintenance scenario. It is a classic example of using architectural principles to design technical solutions, while protecting customer data.



Second, conduct value stream analysis to ensure that project funds and resources fetch maximum return on investment.



The value stream diagram revealed that equipment maintenance services contributed 30% of revenue at PAM; followed by spare parts sales, which contributed 16% of revenue. The combined revenue surpassed equipment sales.

After a series of discussions between the architect team and the operations team, a catalog of requirements was prepared. The core requirements included use of AI technology and IoT data to reduce maintenance costs and production disruptions.

Value stream analysis effectively avoided the problem of improper investment of funds and resources and helped formulate a relatively accurate budget with priorities. At the same time, specific requirements were established for enhancing capabilities of personnel and platforms.

Table 1 is part of the catalog of requirements initiated by the architect team.

No.	Category	L1 requirement	Persona	Description
MS001	Intelligent maintenance service	Inspection planning	Maintenance Service Unit	Inspection planning enhanced by AI
MS002	Intelligent maintenance service	Spare parts planning	Maintenance Service Unit	AI-powered spare part procurement and inventory management
MS003	Intelligent maintenance service	Maintenance planning	Maintenance Service Unit	AI-powered maintenance service planning
MS004	Intelligent maintenance service	Service evaluation	Maintenance Service Unit	Using metrics to report performance and improvements

Table 1: Requirement catalog for the Smart Work program

Third, use the architectural analysis method from business requirements to downstream flow - data architecture, application architecture, and technical architecture.

The architects used insights from a gap analysis to update the application and data architecture. The analysis revealed specific technical requirements: optimize the service plan by leveraging accurate forecasts of parts failure, reduce the time for offline inspection, reduce frequency of parts replacement, and accelerate the response to incidents. Meanwhile, the requirements also specified to extend the life of parts at large scale and in continuous production and to ensure the maintenance plans are economic and risk-free. The architect team proposed a technical architecture based on the requirements (Figure 1).

IoT data from sensors and application-level data helped the team identify and address issues: What are the patterns of failure? How do parameters such as abrasion, pressure, vibration, load,

frequency, voltage, and temperature contribute to malfunction?

PAM required two critical capabilities: Data science and a mechanism to continuously optimize the model, algorithm, and application (ML Ops).

Given the specific technology context, the data scientists chose deep learning-based Intelligent Fault Diagnosis (IFD) technology considering: the volume of equipment and IoT data, data value density, and distribution of data sources.

As a feature enhancement, the data scientists also established unsupervised learning mechanisms to find out unknown associations between features and failure patterns. The team of architects also planned data governance initiatives to improve the quality of relevant data.

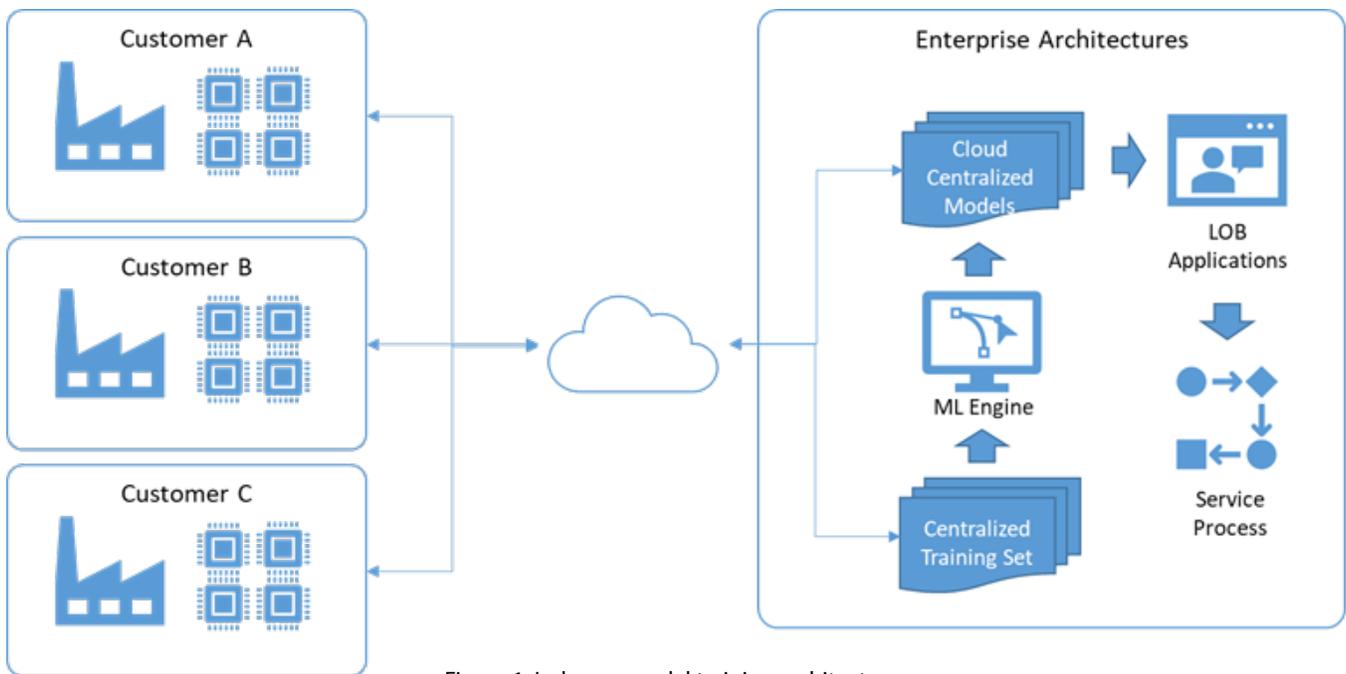


Figure 1: In-house model training architecture

Fourth, use the established architecture design to create an implementation roadmap.

Work packages are identified, and projects are planned for quick-win scenarios. PAM adopted the quick-win scenario planning method of enterprise architecture, while focusing on timelines. By planning quick-win scenarios, the architecture design could be verified at the lowest cost and in the shortest period. It gave users and enterprise decision makers the confidence to invest in additional resources. Notably, it enabled the project team to deliver a minimum viable product even before implementing the engineering platform.



Fifth, implement the engineering platforms and build a technical team.

PAM established a mechanism for data labeling (by identifying various fault or risk situations) and model training. The team defined the scope of the application solution: service planning, reporting and error analysis.

Although cloud-based platforms offer ready solutions for specific industry applications, none of them supported PAM's solution building blocks. The AI team needed a platform with PaaS capabilities such as IoT, big data computing, data mining, AI model development, deployment, and maintenance (Figure 2).

The options for PAM included Microsoft Azure AI Platform, and HiAI from Huawei. The architecture methodology, iterative requirement management solution, project plans, change management, and execution framework was setup accordingly. ML capability building was planned from diverse perspectives, such as business, application, data, technology, and personnel. It ensured that time and resources were spent efficiently across activities.

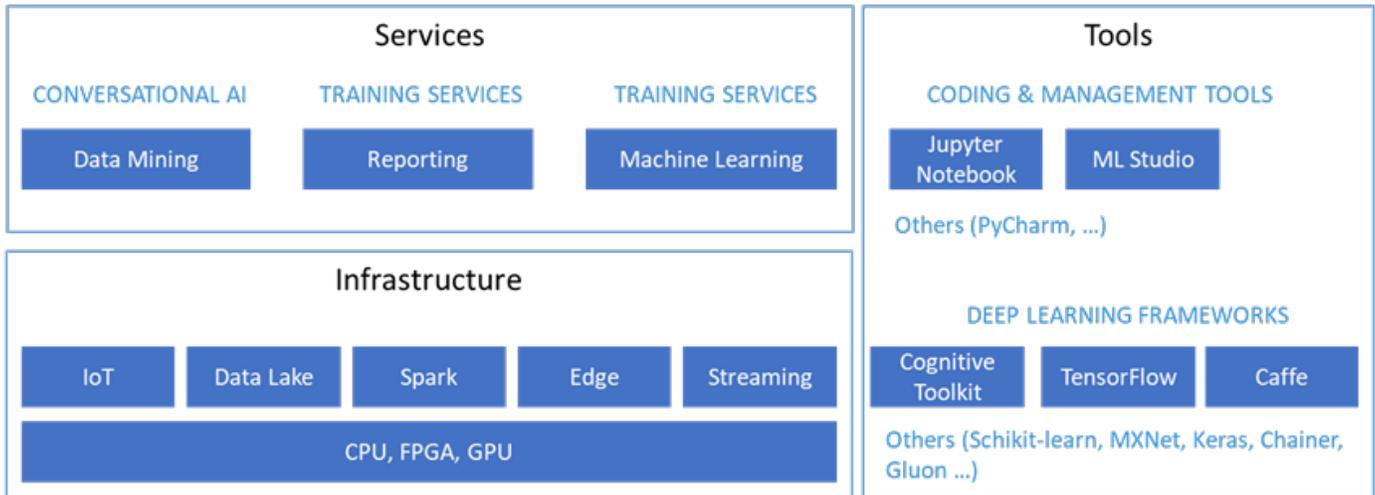


Figure 2: An engineering ML platform

Finally, use architectural governance methods to monitor and govern projects.

A robust governance model provided a checklist for the EA team: establish measurement methods, regularly check whether the project can achieve the established goals and requirements, track

the latest trends and technologies implemented, and ensure compliance with architectural principles and regulations for data protection.



Technical design choices for the PAM Smart Work program

During a technical review of the solutions with stakeholders, the service director raised a critical concern: most of the solutions required customers to share their data, which raised security concerns. Specifically, it could lead to potential violation of business principles related to data privacy.

This problem can only be solved with a program change, which is again governed by architect team. The architect team classified the original solution as an 'internal solution' and used federated learning to solve the problem of user data protection in B2B business scenarios.

Federated learning is a collaborative ML approach that does not require centralized training data. It trains an algorithm across multiple decentralized edge devices or servers, using local data samples. It allows multiple actors to build a common, robust machine learning model without sharing data (Figure 3).

This technology is appropriate for situations where machine learning relies on data that is difficult to extract due to critical issues such as data privacy, data security, data access rights, and technical challenges such as bandwidth for data transfer.

For instance:

- Medical institutions need to gather data from other institutions to meet information requirements.

- Elevator manufacturers need to analyze failure patterns of a variety of subsystems and components, without gathering installation data from the clients.
- Municipal departments need to train models to adjust traffic light control systems to enable better traffic throughput, but beaming video streams of automobiles and passengers consumes immense bandwidth.

Often, enterprises face other issues that lead to solutions based on federated learning:

- IT constraints: cost-intensive infrastructure transformation, such as computational power, cost of transferring large datasets, etc.
- Technology debt: sending and receiving huge volumes of training data in a legacy IoT environment.

Some other applications of federated learning include:

Elevator service management: data privacy and bandwidth issues due to the lack of awareness of risk patterns and the relationship between different types of incidents and parameters such as vibration, load, frequency, voltage, and temperature.

Car navigation service: data privacy issues while developing smart software with capability for dynamic route planning based on the season, occasion, day of the week, day of the month, and time of the day.

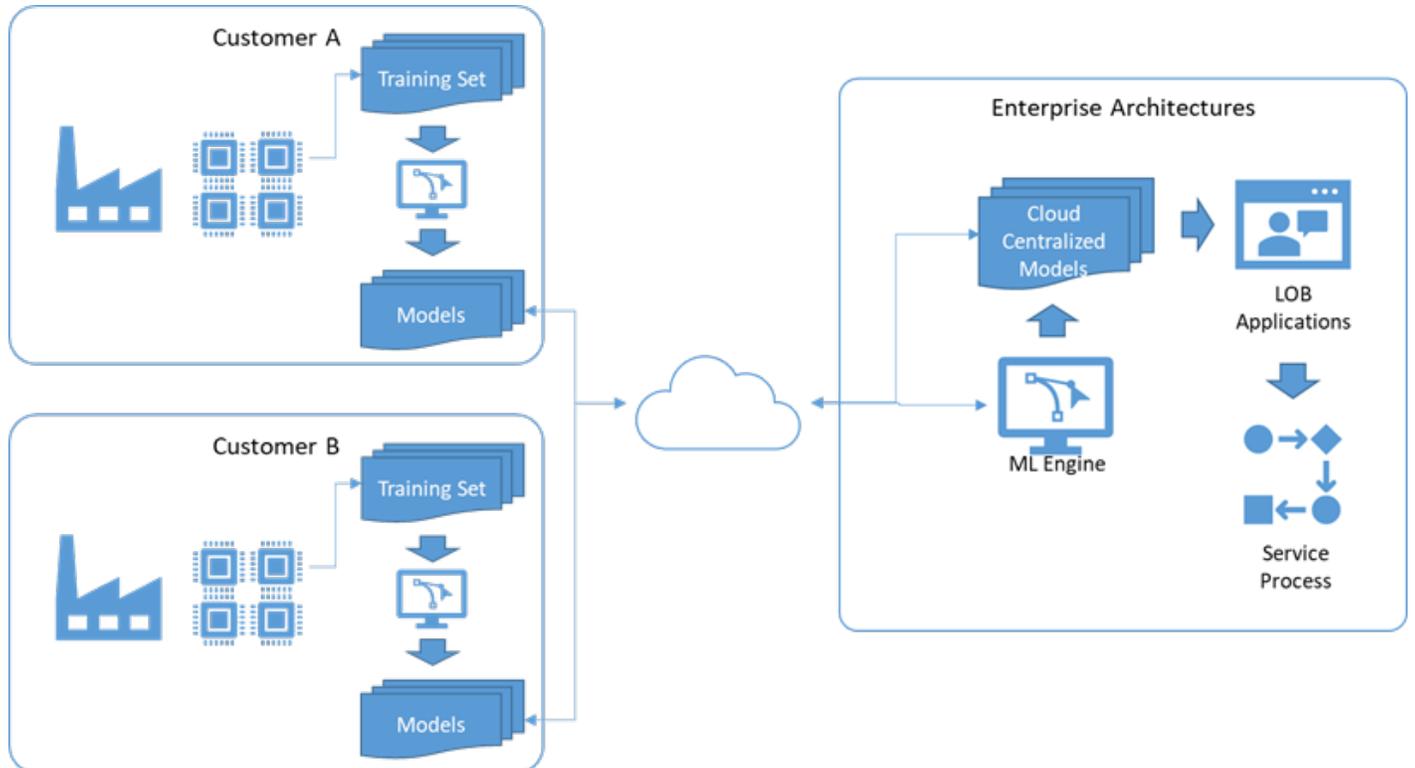


Figure 3: A solution based on federated learning

Cognitive options for technical design of the Smart Work program

Infosys consultants and the team of architects at PAM explored the use of Large Model AI including ChatGPT in the manufacturing industry, and found that it can be applied in the following business cases:

- Predictive manufacturing: Large Language Model can be useful in predicting the production output of chemicals, raw material requirements for specific manufacturing workflows, and pathways to improve the efficiency of equipment.
- Fault diagnosis: Large Model AI helps diagnose faults in complex manufacturing systems and expand relevant knowledge.

- Quality management: Helps to analyze the root cause of quality issues and provides recommendations for effective resolution.
- Manufacturing process optimization: Helps optimize workflows, which improves operational efficiency and quality of production.

Private deployment of Large Model AI is a feasible option to address specific business needs. The optimal technology approach is to fine-tune or prompt general industry specific models using enterprise specific SOP as the corpus to enable staffs in various operations.

ML-powered future

Machine learning is an integral part of digital transformation capability. The methodology developed by Infosys for adoption of enterprise architecture practices helps manufacturers avoid common failure modes in ML projects. The EA approach decomposes relevant value streams into processes, roles,

requirements, applications, data, technical capabilities, and solution components. Further, iterative architectural design helps identify business and IT issues, resulting in bespoke technical solutions that capitalize on federated learning and Large Model AI.

Authors



Rajnish Sharma

Vice President, Head of Service Delivery and Business Operations, Infosys, China

He is responsible for overall Business Operations, Service Delivery and Client Success for Global and China based clients of Infosys in the Greater China region. With a passion for Cloud, AI and Automation led digital service offerings, he leads teams of professionals at Infosys that define and implement strategies for clients to navigate their digital transformation journey. He is based out of Shanghai and leads a cross functional team in areas including employee engagement, talent management, quality, capability development and presales functions. He has been with Infosys for over 20 years and has worked alongside clients in multiple industry sectors, across the globe including the US, Europe and Asia Pacific geographies.



Samxuan Wang

Principal Technology Architect, Infosys

Lead of Strategy Technology Group, Infosys, China

Samxuan is responsible for optimizing architecture capabilities at the enterprise, solution and technology level. He is part of a team that focuses on cloud migration, data analytics, and the business application of cutting-edge technologies such as machine learning, Web3, and blockchain.

For more information, contact askus@infosys.com



© 2023 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.