A CTO's Guide to Top Artificial Intelligence Engineering Practices

Published 29 October 2021 - ID G00744628 - 15 min read

By Analyst(s): Arun Chandrasekaran, Farhan Choudhary, Erick Brethenoux

Initiatives: Digital Future; Artificial Intelligence

AI initiatives are increasing across organizations, but enterprise architecture and technology innovation leaders continue to face major challenges in moving them to production. This research outlines top engineering practices that can accelerate operationalization of AI to reap business value.

Additional Perspectives

- Summary Translation: A CTO's Guide to Top Artificial Intelligence Engineering Practices
  (24 November 2021)

Overview

Key Findings

- AI projects are characterized by high failure rates and take a long time to move from pilot to production. Slightly more than 50% make it from pilot to production, and those take an average of nine months.
- Most enterprises view artificial intelligence (AI) projects mostly as a model development exercise. Hence, far less time and less budget are spent on data preparation, model operationalization and integration with core enterprise infrastructure and applications.
- There is a huge knowledge gap in understanding how to foster cohesive collaboration between data science, data engineering, IT and other critical teams needed to succeed with AI projects.
- The AI ecosystem is dynamic and fast moving, with numerous technology advancements. Staying abreast of these advancements and leveraging their business value is a daunting exercise due to lack of a rigorous scanning process and perceived maturity risks.

Recommendations

Enterprise architecture and technology innovation leaders including CTOs driving business transformation through technology innovation should:

- Foster an AI engineering practice, and enable industry best practices across data, analytics and software engineering (i.e., DataOps, ModelOps and DevOps).
- Nurture a high-performance team from the start. Fill critical roles for AI, develop a training and/or hiring program with HR, ensure responsibilities are well-defined among teams, and define common KPIs that emphasize agility, accuracy, performance and responsible AI.
- Harness the potential of composite AI by ensuring the right alignment between business use cases and the appropriate AI techniques.
- Experiment with new developments in the field of AI, such as generative AI, which can generate models and data to solve critical business challenges.

Strategic Planning Assumptions

Through 2025, close to 50% of AI projects will face delays in moving to production due to lack of organizational collaboration and IT process immaturity.

By 2025, 75% of enterprises will use cloud to operationalize AI to benefit from elasticity, integration, reduced operational complexity and the cognitive APIs of cloud providers.

By 2025, use of synthetic data and transfer learning will reduce the volume of real data needed for AI by more than 50%.

Introduction

AI continues to be a top priority for enterprise architecture and technology innovation (EA&TI) leaders; however, reaping its business value has been elusive and difficult for many enterprises. Innovation leaders need to create a strong and cohesive AI engineering discipline within their organization to usher governance, agility and scalability of AI projects.

AI engineering is a discipline focused on the governance and life cycle management of a wide range of operationalized AI and decision models. AI engineering methods enable better governance and consistency in reusing, retraining, rebuilding, interpreting and explaining AI models. These methods aim to provide an uninterrupted flow between the development, operationalization and full maintenance of AI models.

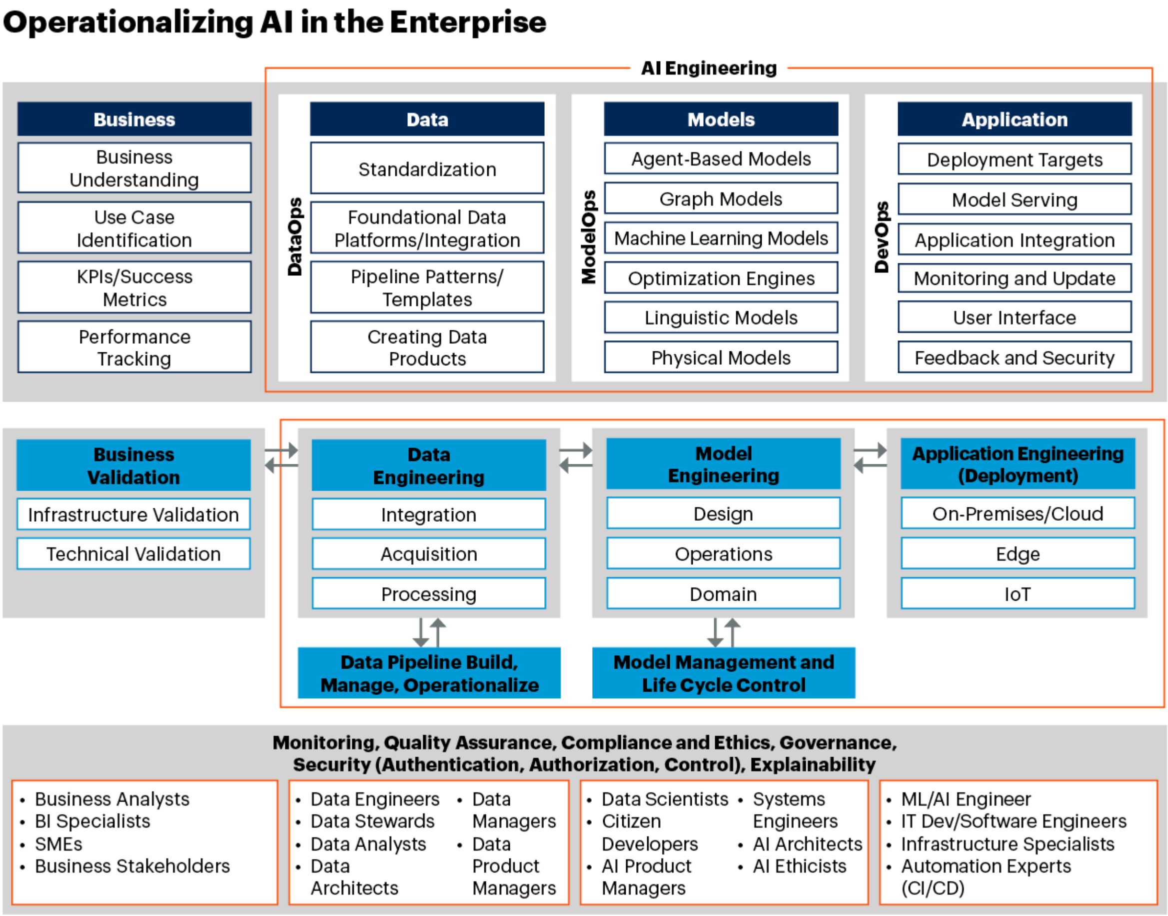
AI engineering stands on these vital foundational pillars — DataOps, [1] DevOps [2] and ModelOps, [3] with responsible AI [4] practices integrated within. While these are foundational pillars, AI engineering practices need to constantly evolve by leveraging composite AI techniques as well as by utilizing emerging, generative AI technologies (such as synthetic data) that can augment AI processes in resource-constrained environments.

Analysis

Create an AI Engineering Practice and Enable Industry Best Practices Across Data, Analytics and Software Engineering

Operationalizing AI in the enterprise requires multiple stakeholders, personas and practices to come together to realize the value of AI initiatives. Figure 1 highlights the disciplines, objectives and personas associated with it.

Figure 1: Operationalizing AI in the Enterprise



Source: Gartner
744628_C

The business stakeholders must stay involved in all aspects of an AI initiative. These include gathering business goals and aligning the end objective to the strategic vision, use case identification and shortlisting (see Toolkit: Discover and Prioritize Your Best AI Use Cases With a Gartner Prism), defining success metrics and identifying KPIs that will be directly impacted, and doing cost-benefit analysis and performance monitoring of AI projects. This is critical because business must be directly involved with the infrastructure, data engineering, model engineering and application engineering teams.

Data engineering involves making appropriate data accessible and available to various data consumers, such as data scientists and business analysts, at the right time, while also informing the business stakeholders of the data situation at the organization. It involves collaboration across business and IT, and often is a centralized practice aimed at establishing the core foundations of data management solutions. One of the primary tasks of a data engineering practice is to automate the data pipeline creation, by, for example, building metadata-driven data pipeline patterns (see How to Build a Data Engineering Practice That Delivers Great Consumer Experiences).

Recommendations:

- Treat data as a corporate asset, and ensure there is a centralized data management team (consisting of data engineers, data stewards, database administrators and others) to define standards for data preservation, data quality, master data management and adherence to regulatory compliance.
- Adopt a DevOps approach to data through automated build, test and/continuous deployment automation to manage the full life cycle of data pipelines.
- Empower line of business teams so they may manage their data and AI products by adapting DataOps practices within targeted use cases.
- Focus on reusability of data assets throughout the organization to enable free flow of ideas, collaboration and trust, and promote a data-driven culture.

Model engineering involves creating the right mix of models that can interact with or benefit from one another while working with the business, data and IT teams. The end objective of any model should be operationalization in such a way that it directly impacts business KPIs and improves success metrics defined by the business in the shortest span of time. Model engineering is central to AI engineering because it converges various AI artifacts, platforms and solutions, while ensuring reusability, scalability and governance of the AI models.

Recommendations:

- Avoid looking at model engineering as a monolith of just model development. Don’t create a best-in-class model without an end goal in mind. Model engineering needs to ensure operationalization and scalable value generation.

- Ensure that you have a centralized ModelOps solution that encompasses continuous model deployment, validation checks, reusability, real-time monitoring, ongoing compliance and comprehensive change management capabilities.

- Evolve your ModelOps to make your model engineering practice more future resilient, such as ensuring quality checks at each step of the process, explainability, security and privacy, while paving the way for composite and generative AI.

Agile software engineering is critical in creating a flexible yet consistent software pipeline for AI projects. AI projects are similar to other software development projects in mirroring nonproduction (development and testing), preproduction (staging) and production environments. However, AI projects are different from software projects due to the rapid iteration of models and experimentation with both model code and data. DevOps tools that can aid in tight application packaging, model service granularity, consistent code versioning and automated software delivery should be part of the AI development life cycle.

Some key DevOps tools that can aid include:

- Code repositories that help with source code management and container build and management tools, which can simplify model packaging, scanning, orchestration and distribution.

- Continuous integration/continuous delivery (CI/CD) pipeline tools that can automate and accelerate software delivery. Model-serving tools streamline the deployment of AI models to a target environment such as Kubernetes.

- Security tools for build, runtime scanning, secrets management and workload protection.

Recommendations:

- Ensure there is a high degree of consistency across your nonproduction, preproduction and production environments to accelerate the time to market of your models.

- Take an agile approach to AI. Ensure continuous delivery of software, deliver working software frequently, create small and self-organizing teams, and encourage developers to closely work with business people and/or data scientists.

- Deploy a container-/Kubernetes-based architecture to bring all the tools and methodologies together to solve your model packaging, heterogeneity and infrastructure challenges.
- Choose an end-to-end integrated DevOps toolchain to bring standardization, governance, security and automated workflow into the AI development life cycle.

Nurture a High-Performance AI Team

To successfully operationalize and scale AI initiatives, organizations need to build diverse AI roles and skills. The core roles that most organizations need are AI engineers, data engineers, data scientists and business domain experts. The core AI team members are often part of the office of the CTO (or CIO) within analytics, innovation or engineering functions. However, as organizations start expanding their use cases and mature and scale their AI practice, a number of other additional roles might become critical.

All of these roles are not mandatory. Determine which roles you want to fill within your AI practice based on the job descriptions in this research. Figure 2 summarizes some of the core (current) and emerging roles in the AI domain.

Figure 2: AI Roles in the Enterprise

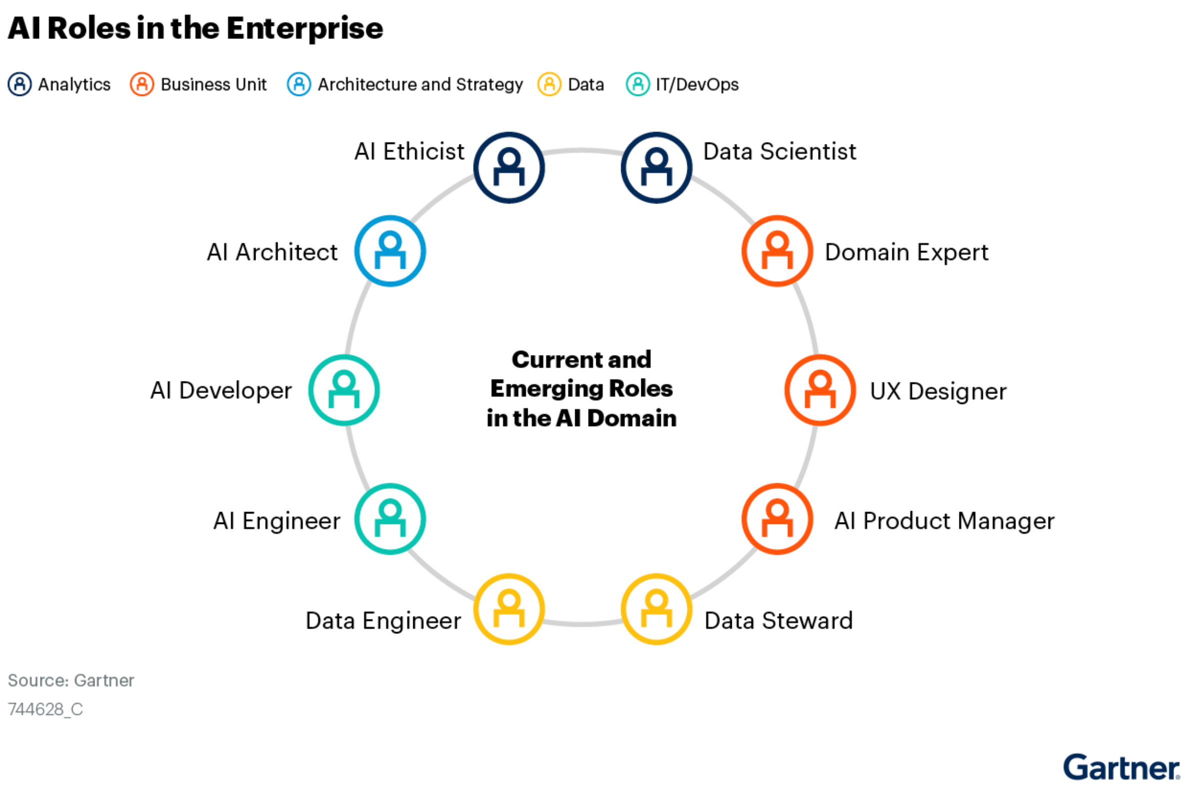


Table 1 describes the key roles and their responsibilities.

Table 1: Key Roles and Their Responsibilities

(Enlarged table in Appendix)

Role	Responsibilities
AI Architect	- Integrate AI into overall IT infrastructure. - Gather requirements from stakeholders, and map them to processes and products that need to be implemented.
AI Developer	- Design and develop AI applications. - Develop APIs to help integrate data products and code into applications.
AI Engineer	- Collaborate with data scientists, and streamline delivery of AI models into production. - Optimize AI models for performance, scalability and reliability with a focus on continuous improvement.
AI Ethicist and/or AI Trust Officer	- Analyze the business ethics, and align AI use cases and decisions with company mission. - Ensure trustworthy AI through organizationwide training and collaboration that puts data privacy, security, explainability, and algorithmic and data bias at the center.
AI Product Manager and/or Domain Expert	- Understand and prioritize market opportunities for AI use cases. - Serve as an interface between customer needs and AI development and deployment teams.
Data Engineer	- Make the appropriate data available for data scientists. - Work on implementing projects with a focus on collecting, parsing, managing, analyzing and visualizing large sets of data.
Data Scientist	- Identify analytics challenges that offer the greatest opportunities, and determine appropriate datasets and algorithms. - Experiment with and build AI models.
Data Steward	- Audit and classify data assets. - Create data policies that balance self-service needs with compliance, security and IP risks.
Domain Expert	- Have domain knowledge, and bring in the context and consequence of the use cases. - Assess value versus risk, and work with data scientists and AI ethicists on acceptable AI outcomes.
UX Designer	- Conduct UX testing of AI products, and provide suggestions on how to differentiate AI products. - Bring human-centric thinking to AI product design, and provide appropriate user feedback to data science teams.

Source: Gartner (October 2021)

Recommendations:

- Map your use cases to the above roles. Determine which of these are critical roles that need to be filled, and create a timeline for filling them.
- Seek to fill these roles from within by creating a training and upskilling program in partnership with HR. Augment with massive open online courses (MOOCs) and other online programs, where feasible.

- Incentivize employees to be involved in open-source communities (see A CTO's Guide to Top Practices for Open-Source Software) and academic conferences. These can be great resources for new knowledge and talent hires.
- Encourage the team to adopt a lean-startup mindset. Validate ideas against business value, and focus on a minimum viable product to demonstrate feasibility and to reduce time to market.

Experiment With New Developments in the Field of AI

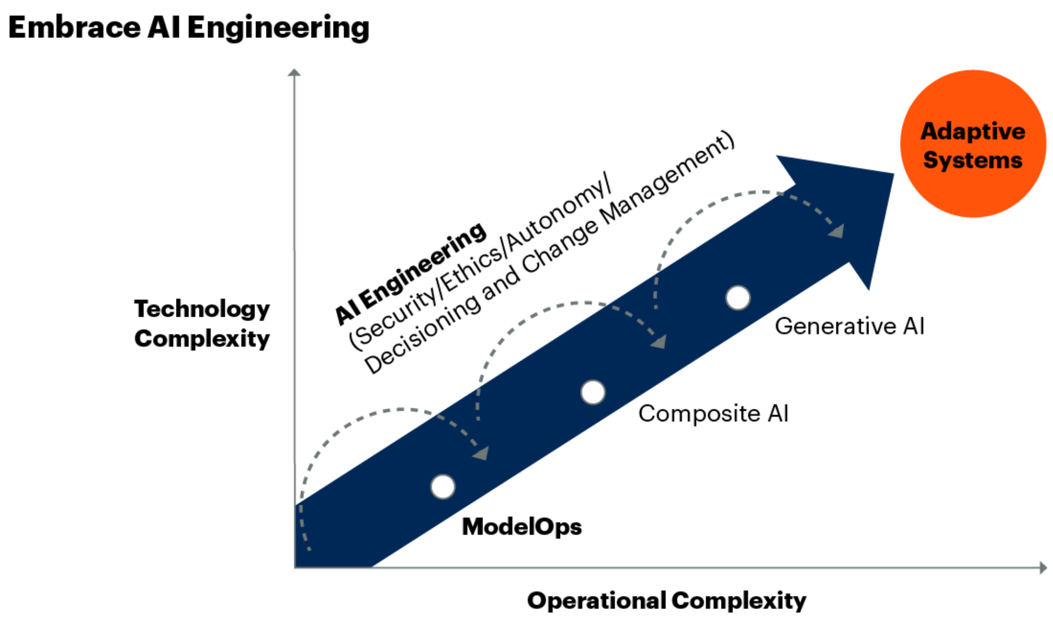
ModelOps and its integration within other Ops practices is the foundation of the AI engineering discipline. In a majority of cases, ModelOps still orchestrate the life cycle of a unique AI technique (e.g., MLOps for machine learning models). Even if there is no direct mapping from one technique to a specific business problem, there are categories of problems that can be mapped to a family of AI techniques. As an example, unveiling correlations in a large and multi-dimensional dataset usually involves machine learning techniques. Also, finding feasible courses of actions in a given amount of time, taking into consideration a set of constrained resources will require optimization techniques. However, in an increasing number of cases in the next three to five years, AI engineering will also deal with the combination of AI techniques (or composite AI).

Organizations are combining different AI techniques to improve the efficiency of learning to broaden the level of knowledge representations and, ultimately, solve a wider range of business problems more efficiently. Combining AI techniques offers two main benefits in the short term:

- It brings the power of AI to a broader group of teams that do not have access to large amounts of historical or labeled data but do possess significant human expertise. Composite AI is one of the strategies to deal with "small data" (or lack of adequate data).
- It helps expand the scope and quality of AI applications because more types of reasoning challenges and required intelligence can be embedded in composite AI. Other benefits, depending on the techniques applied, include better interpretability and the support of augmented intelligence.

These benefits are possible if carefully crafted, maintained and governed through a purposeful adoption of an AI engineering practice orchestrating those efforts (see Figure 3).

Figure 3: AI Engineering Evolution



Source: Gartner
750904_C

Gartner

Among the emerging innovations in the field of AI, generative AI stands out as a high-business-impact trend. Generative AI refers to AI methods that can learn a representation of artifacts from the data, and use it to generate new, original, realistic artifacts that preserve a likeness to original data. Some emerging use cases for generative AI include:

- Content improvement
- Data science and business analytics — synthetic data
- Drug design
- Material science
- Creative content impacting media and marketing industries

In the long run, generative AI techniques will fundamentally change most industries. The benefits to society are already significant. Examples include support for the formulation of new drugs for unforeseen pandemics and the generation of synthetic data (artificial data) to overcome the “small data” problem.

Generative AI algorithms can be used to create models of things that do not exist in the real world. Generative models have the potential to impact many creative activities from content creation (e.g., art, stories, marketing materials) to many types of design (e.g., architecture, engineering, drug, fashion, industrial, process). Generative models can also be used for the inverse design of materials to have specific properties.

A robust AI engineering strategy will facilitate the performance, scalability, interpretability and reliability of AI models, while delivering the full value of AI investments throughout the evolution of AI techniques.

Evidence

This research is based on more than 500 phone inquiries and the authors' face-to-face interactions with Gartner clients.

Gartner's 2019 AI in Organizations Study was conducted online from November through December 2019 among 607 respondents from organizations in the U.S., Germany and the U.K. Quotas were established for company size and for industries to ensure the sample was representative. Organizations were required to have developed AI or intend to deploy AI within the next three years.

Respondents were screened to be part of the organization's corporate leadership or report into corporate leadership roles, and have a high level of involvement with at least one AI initiative. They had to have one of the following roles when related to AI in their organizations: determine AI business objectives, measure the value derived from AI initiatives or manage AI initiatives development and implementation.

The study was developed collaboratively by Gartner analysts and the Primary Research Team.

Results of this study do not represent global findings or the market as a whole but reflect sentiment of the respondents and companies surveyed.

[1] DataOps provides the foundational data operations for operationalizing AI models. It improves the flow of data to points of consumption in the business. It applies DevOps practices to data consumption by operationalizing data pipelines and workflow orchestration to specific consumer use cases.

[2] DevOps is a customer-value-driven approach to deliver solutions using agile methods, collaboration and automation. It emphasizes people, culture and collaboration among development, operations and other stakeholders to improve the delivery of customer value.

[3] AI model operationalization (ModelOps) is primarily focused on the governance and life cycle management of all AI and decision models. It aims to eliminate internal friction among teams by sharing accountability and responsibility.

[4] Responsible AI is an umbrella term for many aspects of making the right business and ethical choices when adopting AI that organizations often address independently. For example: business and societal value, risk, trust, transparency, fairness, bias mitigation, explainability, accountability, safety, privacy, and regulatory compliance. Responsible AI operationalizes organizational responsibility and practices that ensure positive and accountable AI development and exploitation.

Acronym Key and Glossary Terms

Small Data	The concept of “small data” indicates both the issue and approach on how to train AI models when small amounts of training data are enough or there is insufficient or sparse training data. There are a variety of strategies and data augmentation techniques to overcome the problem such as simulation, synthetic data, transfer learning, federated learning, self-supervised learning, few-shot learning and knowledge graphs.
Generative AI	Generative AI is a variety of AI methods that learn a representation of artifacts from the data, and use it to generate new, original, realistic artifacts that preserve a likeness to the training data, but do not repeat it. Generative AI can produce novel content (e.g., images, video, music, speech, text — even in combination), improve or alter existing content, and create new data elements.

Appendix

Table 2: List of OSS Projects and Commercial Vendors (Nonexhaustive)

(Enlarged table in Appendix)

DataOps Vendors/Projects	ModelOps Vendors/Projects	DevOps Vendors/Projects	Generative AI Vendors/Projects
Ascend.io	Algorithmia	Airflow	Adobe
Databricks	Domino	GitHub	AI.Reverie
DataKitchen	Hydrosphere.io	GitLab	DeepMind (Google)
Delphix	IBM	MLflow	Diveplane
Immuta	Iguazio	Red Hat (IBM)	Landing AI
Informatica	ModelOp	Kubeflow	IBM
Qlik	Modzy	Rancher (SUSE)	OpenAI
StreamSets	SAS	Jenkins	
Tamr	Seldon	Argo AI	
	Most DSML Vendors	VMware	

Source: Gartner (October 2021)

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Top Strategic Technology Trends for 2021: AI Engineering

Top Trends in Data and Analytics for 2021: From Big to Small and Wide Data

Hype Cycle for Artificial Intelligence, 2020

Innovation Insight for ModelOps

Predicts 2021: Artificial Intelligence Core Technologies

Magic Quadrant for Data Science and Machine Learning Platforms

Demystifying XOps: DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI

Implementing an Enterprise Open-Source Machine Learning Stack

Operational AI Requires Data Engineering, DataOps and Data-AI Role Alignment

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."

Table 1: Key Roles and Their Responsibilities

Role	Responsibilities
AI Architect	▪ Integrate AI into overall IT infrastructure. ▪ Gather requirements from stakeholders, and map them to processes and products that need to be implemented.
AI Developer	▪ Design and develop AI applications. ▪ Develop APIs to help integrate data products and code into applications.
AI Engineer	▪ Collaborate with data scientists, and streamline delivery of AI models into production. ▪ Optimize AI models for performance, scalability and reliability with a focus on continuous improvement.
AI Ethicist and/or AI Trust Officer	▪ Analyze the business ethics, and align AI use cases and decisions with company mission. ▪ Ensure trustworthy AI through organizationwide training and collaboration that puts data privacy, security, explainability, and algorithmic and data bias at the center.

AI Product Manager and/or Domain Expert	■ Understand and prioritize market opportunities for AI use cases. ■ Serve as an interface between customer needs and AI development and deployment teams.
Data Engineer	■ Make the appropriate data available for data scientists. ■ Work on implementing projects with a focus on collecting, parsing, managing, analyzing and visualizing large sets of data.
Data Scientist	■ Identify analytics challenges that offer the greatest opportunities, and determine appropriate datasets and algorithms. ■ Experiment with and build AI models.
Data Steward	■ Audit and classify data assets. ■ Create data policies that balance self-service needs with compliance, security and IP risks.
Domain Expert	■ Have domain knowledge, and bring in the context and consequence of the use cases. ■ Assess value versus risk, and work with data scientists and AI ethicists on acceptable AI outcomes.

UX Designer	■ Conduct UX testing of AI products, and provide suggestions on how to differentiate AI products. ■ Bring human-centric thinking to AI product design, and provide appropriate user feedback to data science teams.

Source: Gartner (October 2021)

Table 2: List of OSS Projects and Commercial Vendors (Nonexhaustive)

DataOps Vendors/Projects	ModelOps Vendors/Projects	DevOps Vendors/Projects	Generative AI Vendors/Projects
Ascend.io	Algorithmia	Airflow	Adobe
Databricks	Domino	GitHub	AI.Reverie
DataKitchen	Hydrosphere.io	GitLab	DeepMind (Google)
Delphix	IBM	MLflow	Diveplane
Immuta	Iguazio	Red Hat (IBM)	Landing AI
Informatica	ModelOp	Kubeflow	IBM
Qlik	Modzy	Rancher (SUSE)	OpenAI
StreamSets	SAS	Jenkins	
Tamr	Seldon	Argo AI	
	Most DSML Vendors	VMware	

Source: Gartner (October 2021)