Gartner Research

# Roles and Skills to Support Advanced Analytics and AI Initiatives

Zain Khan

2 August 2022

Gartner

# Roles and Skills to Support Advanced Analytics and AI Initiatives

Published 2 August 2022 - ID G00770015 - 44 min read

By Analyst(s): Zain Khan

Initiatives: Analytics and Artificial Intelligence for Technical Professionals

Data and analytics technical professionals need to define their roles and work together as part of an AI team. This research aims to define core AI and ML roles, skills and responsibilities, thereby helping to align the right skills to the required roles in advanced analytics initiatives.

## Overview

### Key Findings

- The proliferation of, and continuous development in, AI has created the need for roles and functions to help combat the challenges around data complexity and access, ML model ownership, fairness, and explainability.
- Advanced analytics professionals working in isolation without central ownership and management lack strategic insights, thereby limiting the effectiveness of the AI solution.
- “Build once and forget” approaches result in the inability to retain key engineering design patterns and best practices, thereby limiting reusability and hampering AI maturity within organizations.

### Recommendations

Technical professionals looking to work in the advanced analytics domains should:

- Focus on acquiring and strengthening skills around data management and AI use case determination to overcome the most pressing challenges faced in AI implementations.

- Explore the emerging roles, and key responsibilities, of the model validator and model owner, and look to acquire skills on the model monitoring, testing, explainability and ownership front.
- Define key roles catered to each phase of the ML development life cycle, aligning business goals with long-term ML growth and working together as part of an AI team to achieve greater strategic success in advanced analytics implementations

## Strategic Planning Assumptions

By 2025, a scarcity of data scientists will no longer hinder the adoption of data science and machine learning in organizations.

By 2024, 40% of all organizations will offer or sponsor specialized data science education to accelerate upskilling initiatives, up from 5% in 2021.

Through 2023, the machine learning engineer will be the fastest growing role in the AI/ML space, with open positions for ML engineers half (50%) that of data scientists, up from less than 10% in 2019.

## Analysis

### Introduction

Artificial Intelligence (AI) is maturing at a rapid pace. According to Gartner's AI in Organization's Survey 2021, AI usage increased from 35% in 2019 to 52% in 2021. However, data complexity and accessibility, difficulty measuring AI success, and lack of skills of staff remain the top barriers to AI implementation (see Figure 1). As a result, the demand for a highly skilled and diverse AI role continues to soar.

**Download All Graphics From This Material**

## Figure 1. Top 3 Barriers to AI Implementation

**Sum of Top 3 Rank Barriers to AI Implementation**


n = 698; Base: Excludes Not sure

Q: What are or will be the top 3 barriers to the implementation of AI techniques within your organization

Source: Gartner P-21023 AI in Organizations Survey 2021

Note: The percentages have been rounded off

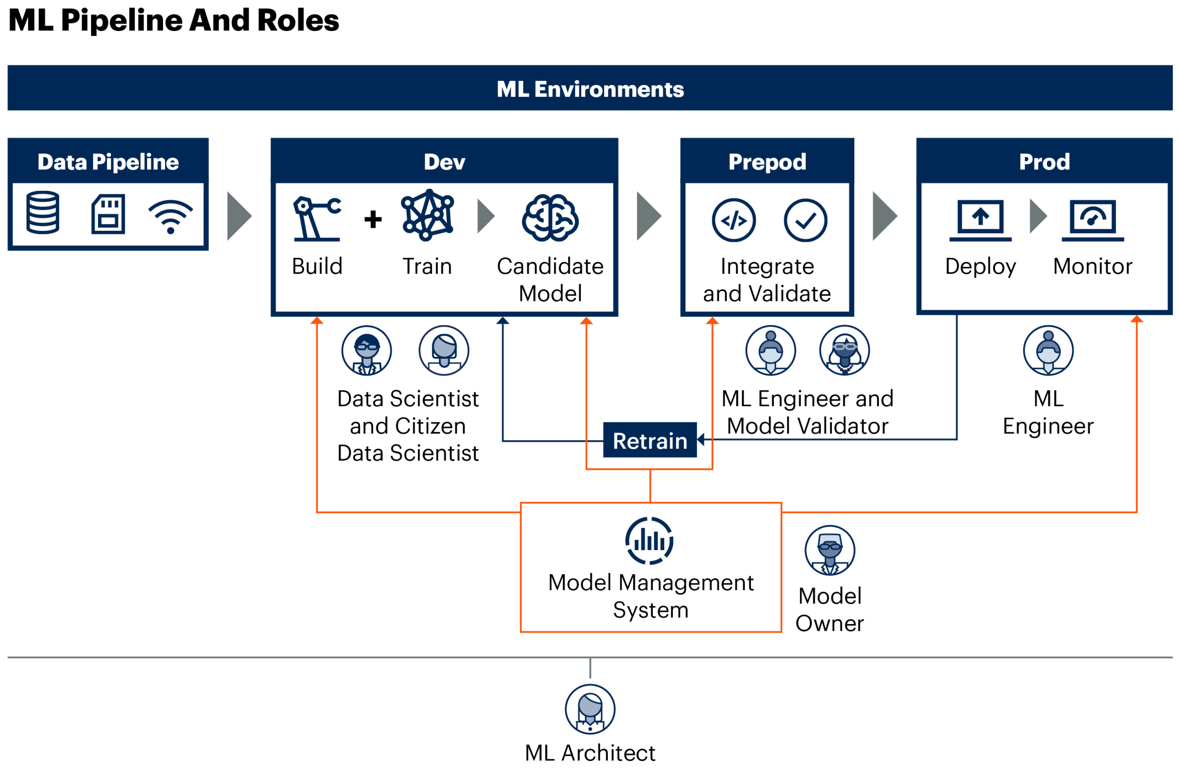
770015_C

The technical barriers shown in Figure 1 form essential areas of growth and progress for data and analytics professionals looking to work in AI initiatives. This research defines the core and emerging roles and skills for technical professionals in the ML/AI space. ML is a subset of AI and constitutes the dominant method in creating AI solutions. For more information on the differences between AI, ML and deep learning, read Go Beyond Machine Learning and Leverage Other AI Approaches.

The roles discussed in this paper include data scientist, citizen data scientist, ML engineer, ML architect, model owner and model validator. It should be noted that these roles are not exhaustive, but are a core combination of key and emerging roles and the overall AI solution needs input from other professionals as well. For more details, read What Are Must-Have Roles for Data and Analytics?

The data scientist, citizen data scientist, ML engineer and ML architect role will be discussed first. They are the current key roles within the AI space, whereas model owners and model validators are key emerging roles and will be discussed later. Figure 2 shows how these roles are assigned to each stage in a typical ML development workflow.

## Figure 2. ML Pipeline and Roles




Source: Gartner
770015_C

The team should focus on synergy and not just be a sum of its parts. Citizen data scientists can work on business use case evaluations with ML model owners and often create prototypes and proofs of concept (POCs) using self-serve SaaS platforms. Data scientists can put these POCs to work by concentrating on the technicalities and working with open-source platforms and frameworks to build and train the model. Once the model has been developed, ML engineers optimize it and place it in production. Model validators perform quality assurance and testing to ensure the model remains explainable to model owners, and model owners can track ML models using model observability tools. All this work happens under the design blueprint and framework outlined by the ML architect, who lays down the architecture, rules and processes, and ensures the privacy and compliance frameworks are also followed. Each of these roles is discussed in later sections.

In the last section, this paper recommends that these core roles work together as part of an AI team to achieve greater success in AI initiatives.

## Data Scientist

### Role and Responsibility

Data scientists stand at the center of any advanced analytics initiatives and have remained the most popular persona within this space. The nature of the role and responsibilities of a data scientist can vary based on their experience, the size and analytics maturity of the enterprise, and project complexity. As such, their responsibilities will vary and include:

- **Machine learning development and tuning.** This involves ML model learning and training, hyperparameter configuration and fine tuning the ML model. This is the core responsibility of data scientists and should be a collaborative and consultative effort in participation with senior data scientists.
- **Researching AI and ML use cases** catered to the different business domains and defining success criteria (in consultation with the ML architect). They should assess the viability of ML to achieve tangible business outcomes and determine whether the business problem even calls for an ML/AI solution.
- **Selecting the correct algorithms for ML development.** Depending on the use case, data scientists will spend time researching the correct ML algorithms and techniques and selecting between supervised, unsupervised or reinforcement learning. For more details, read Machine Learning Playbook for Data and Analytics Professionals.
- **Data selection and management.** Data complexity, quality and accessibility are the top barriers for AI implementation. Data scientists should work closely with data engineers by serving as advisors in the building of data lakes, warehouses and lakehouses. Data scientists should define use-case-specific domain and ML transformation and data selection rules for data engineers and define the ease of data accessibility. For instance, they should identify batch versus streaming or structured versus unstructured data or, within a lakehouse implementation, stage the “silver” or “gold” layer in object storage as delta/parquet files versus staging in a data warehouse. For more details, read Essential Skills for Data Engineers and Data Engineering Essentials, Patterns and Best Practices.

- **Data curation and feature engineering.** This involves labeling and annotating the data from the refined data stores and adding further enhancements that add nuance to the input data for ML algorithms. For details, read Feature Stores for Machine Learning (Part 1): The Promise of Feature Stores.

- **Data exploration and visualization.** Data scientists should spend time exploring data and observing patterns and anomalies from the refined data collected. This not only helps them to understand key metric behaviors but will also shed light on anomalies. This task is usually carried out by junior data scientists as they seek to gain an understanding of the data landscape.

Most AI initiatives fail because there is a lack of care around postdeployment productionizing, maintaining and scaling the ML solution. As such, data scientists often work hand-in-hand with ML engineers and under the supervision of ML architects to deploy the ML solution.

Data scientists should also take on the task of spreading data literacy and explaining the benefits of adopting advanced analytics to aid in decision making. They can help dispel myths around AI explainability and fairness and help educate business users on the myriad AI use cases that can aid and enhance business decision making.

### Skills Required

There are many certifications and training programs offered by almost all major technology companies that offer a good combination of critical thinking and machine learning skills. Examples include IBM Data Science Professional Certificate and Stanford's Machine Learning programs. Some cloud vendors, like Amazon Web Services (AWS) have launched interactive platforms, such as DeepRacer, that provide hands-on training and ML development in a gaming environment.

### Technical Skills

Technical professionals looking to work as a data scientist should:

- Possess a quantitative background with undergraduate or graduate degrees in computer science, physics, statistics, engineering, mathematics or economics. However, this is not strictly necessary as more disciplines become analytical. Degrees in biology, chemistry, business and psychology also impart critical thinking and reasoning skills.

- Have strong formative understanding of statistical and mathematical concepts, theories and applications such as linear algebra, probability theory, calculus, algorithms and data structures.
- Have a strong understanding of machine learning use cases, algorithms and techniques, including differentiating between supervised, unsupervised and reinforcement learning.
- Have a strong understanding of algorithms such as linear regression, logistic regression, regularization, decision trees, clustering algorithms, and matrix factorization techniques.
- Understand the steps involved in the machine learning life cycle, including:
  - Data selection and preparation (on-premises data stores versus cloud, batch versus streaming, files versus database, synthetic versus real data)
  - Feature engineering (imputation, handling outliers, binning, log transform)
  - Model training
  - Model selection
  - Model testing (cross-validation, A/B testing)
  - Model interpretation
  - Inference
- Be proficient in programming languages such as Python, R and MATLAB and be familiar with development environments such as Jupyter Notebook, RStudio, SAS Studio, Microsoft's Visual Studio, and PyCharm and open-source machine learning libraries such as TensorFlow, Keras and PyTorch.
- Have working knowledge of cloud computing and ML platforms and tools. For instance, Amazon SageMaker, Microsoft (Azure Machine Learning), Google's Vertex AI and IBM Watson. For a detailed list of the rankings, see Magic Quadrant for Data Science and Machine Learning Platforms.

- Understand the data management architecture that feeds their ML algorithms, whether on-premises or in the cloud. This involves understanding the concepts and usage around data warehouses, data lakes or lakehouses. As such, they should be comfortable using SQL because it is considered the dominant programming language when it comes to interacting with analytical data stores.

- Possess strong data visualization skills using mainstream business intelligence tools such as Microsoft Power BI, Tableau or even using Python libraries such as seaborn and matplotlib.

- Have an understanding of Machine Learning Operations (MLOps) practices, including DevOps principles such as IaC, containerization and CI/CD pipelines.

### Nontechnical Skills

Technical skills alone do not define a data scientist. Personality fit, communication skills and business acumen are key skills as well. Technical professionals looking to set themselves up for success as a data scientist should:

- Possess strong communication skills. They should be comfortable in explaining technical concepts in common business terms to the business community and technical professionals from different backgrounds.

- House deep domain expertise within their functional areas around the terms, metrics and the overall business function. This is essential because it will help them in designing effective use cases for ML and AI that are catered to the respective business units.

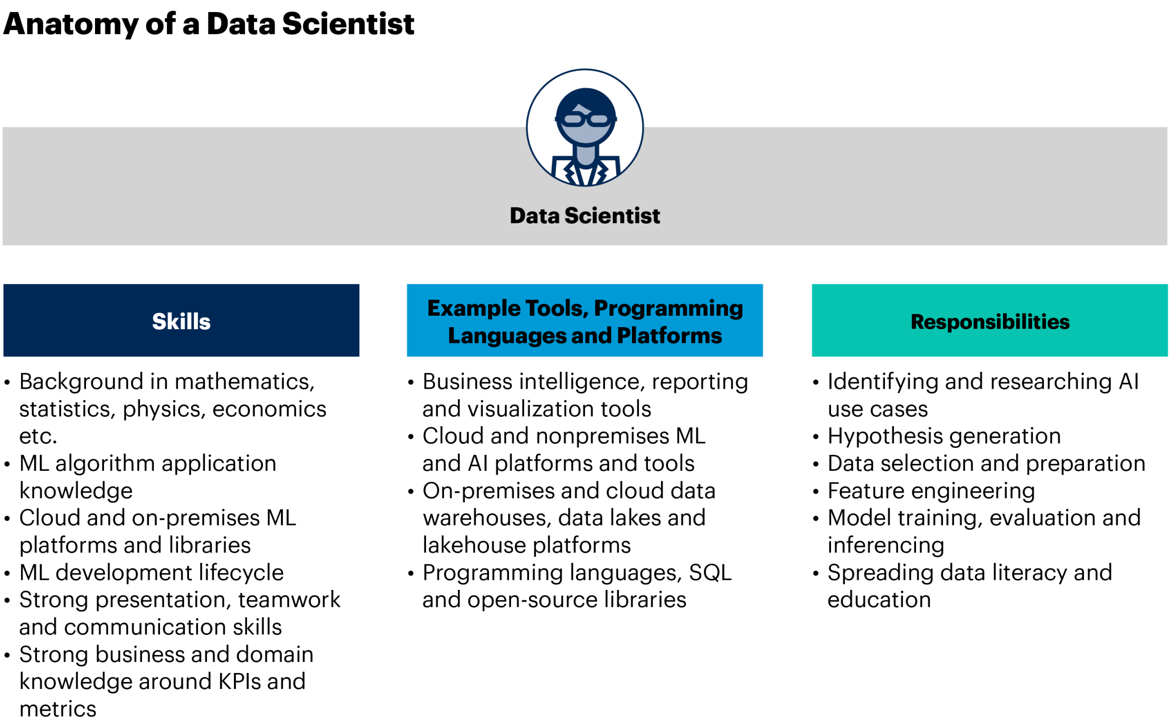
- Enjoy working in a collaborative, team environment. Data science projects involve technical professionals from data management, DevOps, business intelligence and business experts, and it is essential to develop and maintain positive relationships with these professionals.

- Have a curiosity mindset that is always open to researching new use cases and possibilities.

- Help drive data literacy within the organization. They should explain the benefits of ML and AI and help assuage the fears the business community may have around the ML/AI use cases with a focus on ethical AI.

Figure 3 summarizes the current skill set required of data scientists.

Figure 3. Anatomy of a Data Scientist




Source: Gartner
770015_C

## Upskill Path

Data scientists looking to upskill have multiple options in this regard:

- Technical growth toward a hybrid role by learning ML engineering
- Moving up within the data science practice

The hybrid pathway involves learning ML engineering and moving toward the Architect role. It requires knowledge of machine learning deployment and automation, performance tuning, infrastructure and integrating machine learning models into business applications. Google's Professional Machine Learning Engineer focuses on ML operationalization and is a good option for gaining certification within this space. Pursuing the ML architect upskilling path requires more robust training and experience in software engineering, quality assurance, system design, security, user experience design and integration. For more details on these roles, read the appropriate sections within this paper.

However, many data scientists may opt to move up in seniority toward a senior data scientist and then toward the principal data scientist and eventually aim for the chief data scientist position. For more details, read The Chief Data Scientist Role Is Key to Evolving Advanced Analytics and AI.

## Citizen Data Scientist

### Role and Responsibility

Citizen data scientists are business professionals who have the technical inclination and interest in business intelligence (BI) and ML. Citizen data scientists are increasingly becoming important where in-depth technical skills required of data scientists are either not available (as evidenced in Figure 1) or the organization is in an infancy stage regarding ML/AI maturity. Working with ML engineers and data scientists, citizen data scientists might take on the role of a data science educator who helps bridge business with advanced analytics. Or, as a solo analytics specialist, they may use low-code AI SaaS platforms to develop AI solutions. In some other cases, they may also develop the AI solution and hand it off to data scientists for finalizing. Technical professionals looking to work as citizen data scientists:

- Use their domain expertise in **researching effective ML business use cases** and help develop the project objectives.

- Help bridge the skills gap by using **Augmented ML and AI functionalities** and low-code SaaS applications. These capabilities automate the different steps in the development of AI systems and applications using drag and drop interfaces and including feature engineering, algorithm selection, model training and hyperparameter optimization.

- Acquire, explore and establish **data requirements.** This step involves collecting and preparing relevant data to be used in machine learning models. Often, this will involve the use of self-serve data transformation and feature engineering tools like AWS Glue DataBrew and Microsoft (Power Query).

- Extend the functionalities and use cases of traditional descriptive analytics in business intelligence applications by **advocating the use of AI-powered features** such as natural language querying (NLQ). Most BI tools on the market offer these functionalities. Examples include ThoughtSpot and Microsoft Power BI. A detailed list can be found in Magic Quadrant for Analytics and Business Intelligence Platforms.

- **Use pretrained models** to build AI and ML applications. Pretrained models have already been designed, tuned and trained for certain capabilities and do not need feature engineering or manual model selection.

- **Educate businesses on the benefits of AI** because they are also SMEs from their respective functional areas. They can be excellent sources of expanding data literacy, educating business leaders on the benefits of AI and filling knowledge gaps in more mature analytical setups.

- **Lay the groundwork for future AI** development. Organizations in the infancy stage or lacking technical skills may often hire citizen data scientists to start ML initiatives. As such, citizen data scientists can provide valuable expertise down the line when expert technical professionals are hired.

- **Glue IT and business together.** They have the opportunity to present the analytics perspective to senior business leadership and model owners and to present the business perspective to data scientists and ML engineers.

### Skills Required

Citizen data scientists are being considered to fill in a wide variety of skill sets, and with access to many tools and platforms, the skill requirements keep expanding. Although this document provides an overview of the skills required, cloud vendors have started offering training programs for citizen data scientists. Examples include Teradata's Citizen Data Scientist and C3 AI's Citizen Data Scientist programs.

### Technical

Business or technical professionals looking to work as citizen data scientists should have:

- A strong grasp of AutoML and integrated-ML and AI tools, pretrained models, and data warehouses with augmented AI/ML capabilities. There are many products that specialize in the AutoML and integrated-ML space and provide GUI-based self-serve AutoML functionalities such as: Azure Machine Learning Designer (Azure Machine Learning Studio), Amazon SageMaker Canvas,Google Vertex AI (AutoML), IBM AutoAI, DataRobot and H2O.ai. More can be found in Magic Quadrant for Data Science and Machine Learning Platforms.
- End-to-end ML development life cycle knowledge. This includes data acquisition and preprocessing, model building and training, and model testing and deployment.
- Strong understanding of key metrics and KPIs. This is necessary in order to analyze data and construct ML models that answer key business questions.
- Proficiency in the use of business intelligence tools, such as Microsoft Power BI, Tableau, Looker and ThoughtSpot. A complete list can be found within Magic Quadrant for Analytics and Business Intelligence Platforms. These tools enable self-serve descriptive analytics and often provide a launching pad for predictive analytics as more and more AI capabilities are being added.
- Experience using self-serve data transformation tools to prepare data for ML use cases. They should be able to access data, whether stored in data warehouses, data marts or data lakes. Modern ML platforms (e.g., Amazon SageMaker Data Wrangler) have self-serve data preparation built into them.

### Nontechnical Skills

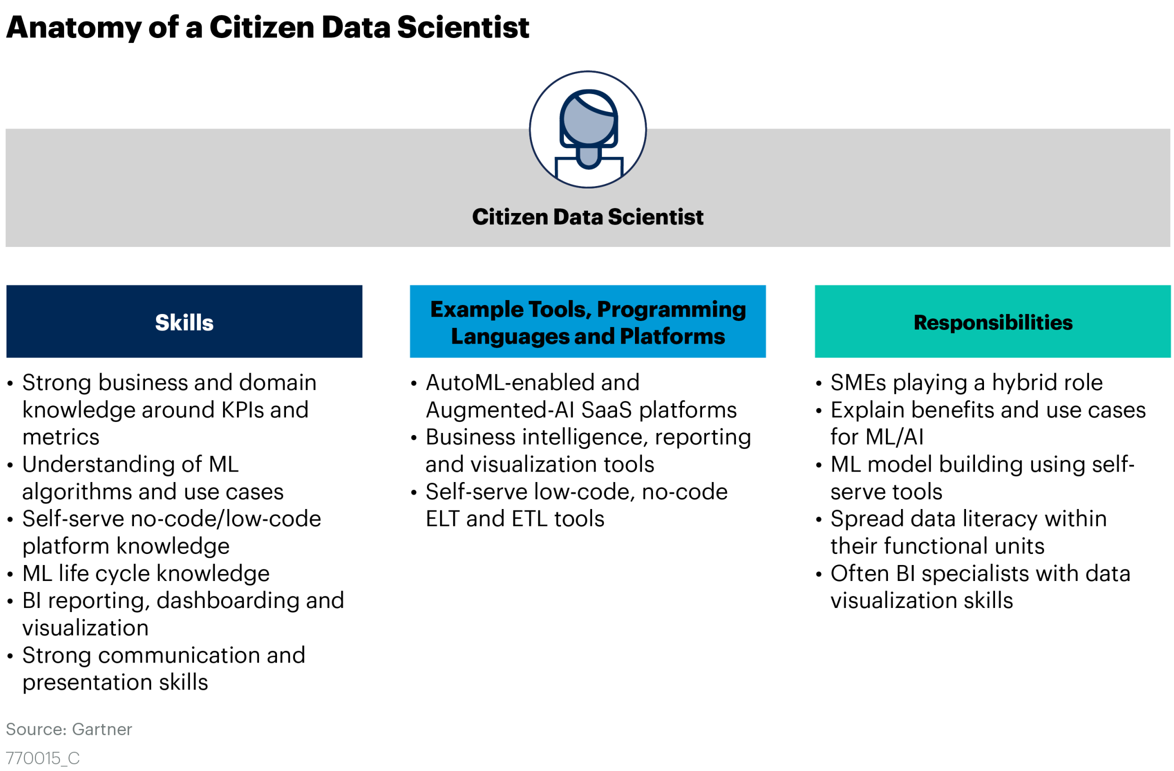
Business or technical professionals looking to work as citizen data scientists should:

- Possess a very strong understanding of the business functional units within an organization. Because they typically emerge from SME roles, they are aware of the business logic, mapping and processes and can provide subject matter expertise to data scientists and ML Engineers.
- Be excellent communicators and enjoy presenting and explaining technical jargon to business users.

- Be data literate — they understand the data lineage and metadata of the data in question. They can map key attributes to business definitions, describe the format of data (structured versus unstructured) and know the velocity of data (batch versus streaming).
- Excel in teamwork and collaboration. Because they occupy the boundary between technical and business users, they should be able to work with senior data scientists.
- Possess a natural inclination toward research and independent learning. This is important because self-serve tools, ML techniques and AI platforms continue to expand. Because they might be serving the role of a lone expert, they are expected to have a strong grasp on the latest tools and AI development in the market.

Figure 4 shows an overview of the citizen data scientist.

**Figure 4. Anatomy of a Citizen Data Scientist**




Gartner

## Upskill Path

Citizen data scientists looking to upskill themselves have excellent options because they already possess a sound understanding of ML concepts and business domains. A natural path is becoming a full-fledged data scientist. For this, they need to learn relevant programming languages such as Python, R and MATLAB and become proficient in programming. This can be challenging, but they have the advantage of already working within the data science domain and can learn on the job by either shadowing senior data scientists or by enrolling in online courses. In today's world, online education platforms such as Coursera, edX, Udemy and Udacity provide education that is low cost and high quality.

They also have the option of pursuing leadership positions within their respective domains and providing an analytics perspective to business decision making. This can involve taking the role of a model owner, which will be discussed later in this paper.

## ML Engineer

### Role and Responsibility

ML engineers are software engineers with a focus on machine learning. They put into production what data scientists design, experiment and build. Technical professionals looking to work as ML engineers should expect to:

- **Performance tune and scale** out the ML models data scientists have developed. Data scientists are not software engineers, and ML engineers will be expected to refactor their Python or R codes into production-ready code and ensure the models are scalable.

- **Productionize the ML models** developed by data scientists. Examples include refactoring Python code written on Jupyter Notebook to PySpark.

- **Develop AI and ML pipelines for continuous operation,** feedback and monitoring of ML models leveraging best practices from the CI/CD vertical within the MLOps domain. This can include monitoring for data drift, triggering model retraining and setting up rollbacks.

- **Optimize AI development environments** (development, testing, production) for usability, reliability and performance.

- Have a strong **relationship with the infrastructure and application development** team in order to understand the best method of integrating the ML model into enterprise applications (e.g., transforming resulting models into APIs).

- **Work with data engineers** to ensure data storage (data warehouses or data lakes) and data pipelines feeding these repositories and the ML feature or data stores are working as intended.

- Evaluate open-source and AI/ML platforms and tools for **feasibility of usage and integration from an infrastructure perspective.** This also involves staying updated about the newest developments, patches and upgrades to the ML platforms in use by the data science teams.

### Skills Required

ML engineers bring software engineering best practices to ML and AI development, but also serve as key members of an overall data science team. Their skills can also be divided into technical and nontechnical skills

### Technical Skills

Technical professionals looking to work as ML engineers should:

- Be well-versed in software engineering (C++, Java, Python), infrastructure provisioning and DevOps principles, whether they are related to infrastructure as code, microservices architecture or CI/CD automation.

- Be able to evaluate the performance and monitoring characteristics of the ML model. These include model size (what is the size of the model), inference performance (speed at which results are returned for inference), memory consumption (how much memory will be consumed once in production), model observability and drift.

- Know ML operationalization and orchestration (MLOps) tools, techniques and platforms. This includes scaling delivery of ML Models (MLOps), managing and governing AI Models (ModelOps), and managing and scaling AI platforms (platform ops for AI). For more information on MLOps, read Demystifying XOps: DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI.

- Be well-versed in choosing the correct approaches around integration, deployment and infrastructure requirements for the ML/AI model. This includes augmented AI and pretrained or open-source models. It involves identifying the correct integration point and interface and correct deployment mode specifying special hardware requirements and the frequency of model updates.

- Be familiar with ML algorithms, AI use cases and applications. They should also have an understanding of, but not expertise in, open-source high-code frameworks like PyTorch or TensorFlow, augmented AI and ML platforms, pretrained ML models and integrated AI PaaS tools. Examples include Azure Machine Learning Studio, Google's Vertex AI, IBM Watson Studio, Amazon SageMaker and open-source tools like Kubeflow. For a detailed list, see Magic Quadrant for Data Science and Machine Learning Platforms.

- Have knowledge about data engineering concepts, tools and automation processes (DataOps) since data pipelines and architectures provide the base for building AI solutions. Examples include MPP data warehouses like Snowflake and Amazon Redshift and all-in-one Apache Spark platforms like Databricks.

#### Nontechnical Skills

Technical professionals looking to work as ML engineers should have the following nontechnical skills:

- Strong collaboration skills. In mature organizations, they might be part of a team of data scientists, model owners and an ML architect, and it is important to work as a team to productionize the AI solution.

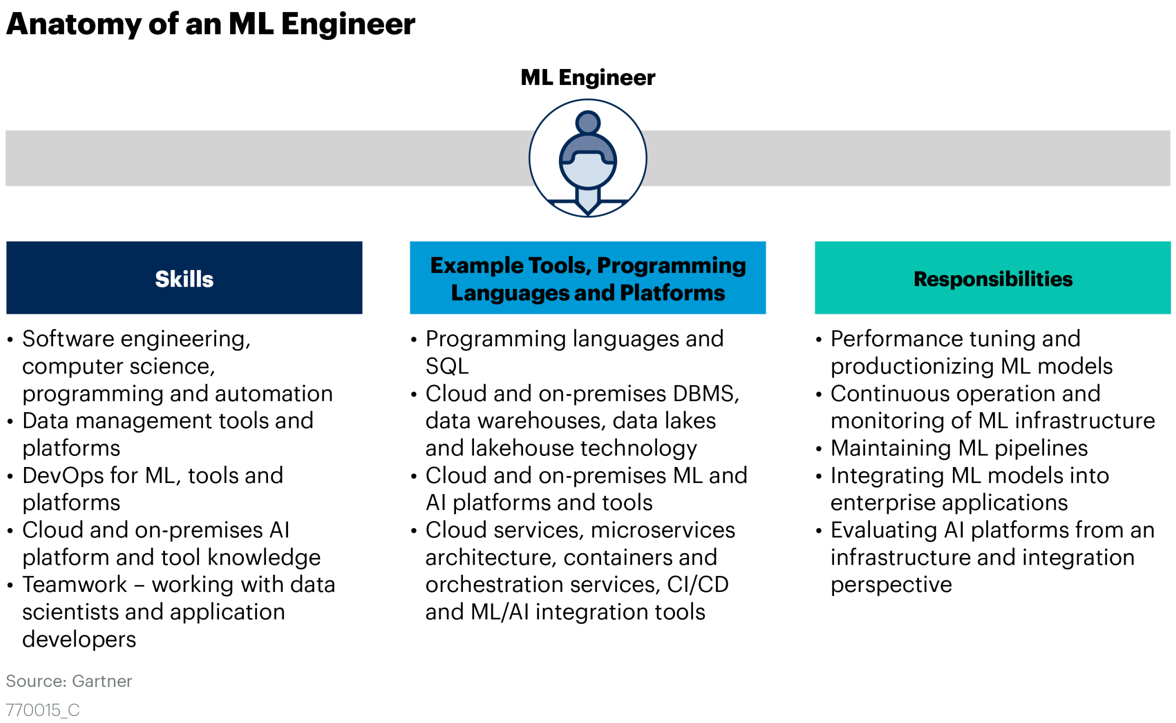
- AI strategy development. They should help devise, along with data scientists and ML architects, the long-term AI growth plan, keeping in mind the scalability and availability of resources.

- Possess an open willingness to learn. This will be crucial in keeping abreast of new developments within the AI and ML space as more and more vendors offer AI orchestration and integration tools. For example, they can learn Agile development and gain an understanding of using Scrum. This can allow for quick creation and prototyping for AI initiatives.

- Be able to explain the AI development process with technical professionals from the software domain to aid in better integration of AI applications into mainstream enterprise applications.

Figure 5 provides an overview of the role and skills required of ML Engineers.

**Figure 5. Anatomy of an ML Engineer**




## Upskill Path

ML engineers looking to upskill have a tremendous opportunity to transition to a hybrid data scientist role or work toward attaining the AI or ML architect position. Because ML engineers already possess an understanding of ML algorithms, software engineering and DevOps principles, they only need to focus on strengthening theoretical knowledge around statistics and mathematics, in addition to learning ML model development and research. Because they work in close collaboration with data scientists, they can shadow and learn domain-specific ML expertise.

Depending on experience, they can also transition to the Architect role. The experience required can range from five to 10 years, and they can opt for cloud certifications in solution architecture with the major cloud vendors, such as AWS, Microsoft Azure and

Google Cloud Platform (GCP).

## ML Architect

### Role and Responsibility

ML architects are the chief technical professionals that are responsible for designing, building and overseeing the overall AI and ML solution in organizations. They are the central point that provides expertise on ML complexities around interconnectivity, integrations, privacy, security, scalability and operation.

Their core responsibilities include:

- **Architecting, designing, directing and leading** AI and ML solutions that are scalable, fault-tolerant and low-bias and that can be integrated and operationalized using available tools and resources.

- **Defining the processes, standards, frameworks, prototypes and toolsets** in support of AI and ML development, monitoring, testing and operationalization. This includes decisions around leveraging open-source, augmented ML, pretrained ML models or integrated AI PaaS solutions.

- **Establishing operational efficiency** in AI initiatives by becoming the liaison with and developing a feedback loop between data science development and business use cases to ensure delivery of business value.

- **Evaluating ML explainability toolkits and product features** and aligning them with business use cases. This includes adherence to regulatory frameworks, security and privacy, explaining the inner workings of the model and interpreting the outcome of ML models to the business community.

- **Providing technical expertise and advisory services** in the development of AI strategy with senior leadership, whether to promote integrated-ML solution development, introduce agile methods to accelerate delivery of ML initiatives or develop long-term plans for ML directives.

- **Leading research around new use cases** of, and educating business about, ML and AI. They attend conferences, webinars and seminars from vendors and major cloud platforms to keep abreast of new developments and bring back this information to senior management.

### Skills Required

#### Technical Skills

Because ML Architects operate at the crossroads of software engineering, data science, systems architecture and deep domain expertise, their skill set is varied. On the technical front, technical professionals looking to work as ML Architects should have:

- Expertise in data science and ML development life cycle and the tools and platforms around it. They should be familiar with open-source programming languages like Python, R, SAS and MATLAB. They should know the differences between augmented AI/ML platforms, pretrained models and integrated ML and AI solutions.

- Proficiency in software engineering, cloud computing, DevOps and system design. They should, ideally, have experience in distributed fault-tolerant systems, streaming applications and microservices architecture and should understand the nuances around on-premises, cloud and multicloud application architecture.

- Strong understanding of data management principles and architectures. This includes knowledge of data warehouses (Snowflake, Amazon Redshift), data lakes and lake houses (whether on-premises or cloud; e.g., Apache Hadoop, Apache Spark, Amazon S3, Azure Data Lake Storage Gen2) as well as understanding Lambda, Kappa and Delta Lake architecture and the extract, transform, load/extract, load, transform (ETL/ELT) tools around them (e.g., Databricks, Azure Data Factory, AWS Glue). For more information, see Exploring Lakehouse Architecture and Use Cases and Assessment of Databricks as a Data and Analytics Platform.

- Experience in integrating applications, such as ML models or AI applications, into mainstream enterprise architecture. This involves knowledge around networking, security, privacy and infrastructure (compute, storage, memory and servers).

#### Nontechnical Skills

ML architects are expected to have strong soft skills as well, and these include:

- Strong communication, teamwork and collaboration skills. Because ML architects stand at the top of the ML food chain, they will be working closely with software architects, data architects, data science teams and senior leadership such as the CIO and CAO. As such, they need to have strong interpersonal skills so they can promote AI within the enterprise architecture.

- Leadership, delegation and technology management skills. Data scientists, citizen data scientists and ML engineers will often look up to the ML architect to provide technical guidance and roadmaps for implementing ML and AI solutions.

- Contract management and end-user license agreements. The architect should also be skilled at managing vendors and their services, contract agreements and end-user licensing agreements.

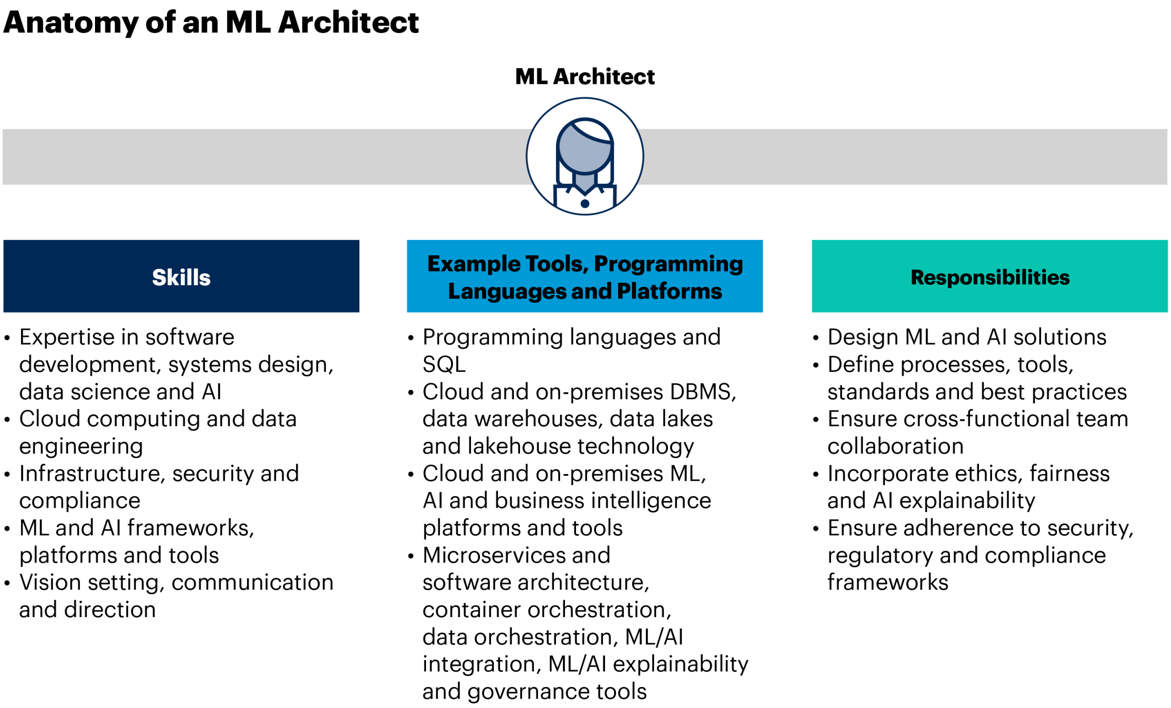
- Systems design thinking from a holistic perspective. Because their chief role is to design scalable ML applications that can integrate into the company's systems, they need to architect and design ML systems that can easily integrate.

- Strong strategizing and governance skills. ML architects should be skilled in planning long-term goals in partnership with senior management. This will enable business goals to be aligned with ML and AI development.

- Domain experience. ML architects should have in-depth business knowledge on the different functional areas because they will routinely collaborate and talk with business leaders in designing ML applications. This will also help in providing insights to business needs and understanding long-term growth plans.

Figure 6 provides an overview of an ML architect.

## Figure 6. Anatomy of an ML Architect




Source: Gartner
770015_C

### Upskill Path

ML architects occupy the most senior technical position within the ML/AI space. They are ideally placed to move into senior leadership and management positions. However, this will be highly dependent on the organization they work in. Mature organizations can offer different pathways, depending on the organizational structure

Chief data scientist is the most senior advanced analytics role within an organization, and those in this position prioritize how ML and AI can contribute to strategic business initiatives. They focus on the ethical implications and manage risks associated with ML and AI development. They help develop and motivate a data-driven culture, with workshops and sessions, to explain AI and data science. As such, they serve to provide inspiration for growth and management of data science teams and AI initiatives. For more information, read What Are the Top 3 Priorities for Chief Data Scientists?

## Model Owner

### Role and Responsibility

Model owners are operational decision makers providing ML model validation and signoff from the business perspective. They are the chief decision makers and subject matter experts from the respective business domains providing business rules and ML model feedback to data scientists and ML Engineers. They may report to the senior leadership of their business units. In some cases, they may also report to the chief data scientist. They are akin to data stewards operating on the data management side.

---

*Model owner is not a strictly defined title. The role can be assumed by domain experts assigned to ML and AI development catered in their respective business units, or it can be an enterprise role with a single person held responsible for all ML models.*

---

Analytics and business professionals looking to take the role of a model owner:

- **Own the ML model from the business perspective.** They decide what business use case the model serves and ensure business value is maintained as time progresses.

- **Do not need to have a separate working title.** They can work part-time and can hold official titles, such as director, manager or even business analyst and report to the respective business heads for which the AI initiative is being developed.

- **Provide the business perspective** and can provide functional requirements for the ML model, thereby playing the dual role of a business SME. They can define the business rules and definitions and provide clarity on the nuances around data, including data quality, testing and validation. They are responsible for ensuring the business rules, definitions and process updates are continuously documented and communicated to the data science team.

- Can **help define model monitoring and measurement framework** to create drift alerts based on business KPIs. This is essential to prevent model performance degradation and value dissipation.

- **Define the rules for, and provide, signoff on ML models.** They ensure the ML models meet the business requirements as part of the requirements submitted

- **Address model access, governance and privacy concerns.** They are the decision makers when it comes to providing access to business owners from other departments or even to data scientists or ML engineers.

### Skills Required

Model owners need to possess enough technical skills to understand data science, but are not required to be technical experts in this domain. Their skills can be divided into technical and nontechnical as well.

#### Technical Skills

Analytics-oriented business professionals looking to fulfill the role of a model owner should have:

- A thorough understanding of the ML development life cycle: Data acquisition and quality, ML model research and use case selection, feature engineering, model development and inference, and model monitoring and production.
- A strong understanding of model monitoring to ensure they can identify model drift from the business perspective and alert the data science team accordingly. They will, most often, use model observability tools like IBM Watson OpenScaleand Amazon SageMaker Model Monitor to observe the ML models. For more details, see Case Study: Monitoring the Business Value of AI Models in Production (Georgia Pacific).
- A strong understanding of, and ability to address, ML challenges on fairness, ethics, privacy and governance. For more information, read Incorporate Explainability and Fairness Within the AI Platform.
- An understanding of ML use cases and types and the ability to understand the correlation with the business use case as selected by data scientists.

#### Nontechnical Skills

In addition to technical skills, model owners should have the following soft skills:

- Excellent communication and collaboration skills. Because the model owners serve as business experts on the ML models, they need to have strong teamwork skills so they can work with data scientists and serve to glue business and ML domains together.

- Expertise in business domain knowledge and understanding of KPIs and metrics. This is to ensure that key metrics remain the priority for the ML model and that any changes in business processes, definitions and data values are communicated to the data science team.

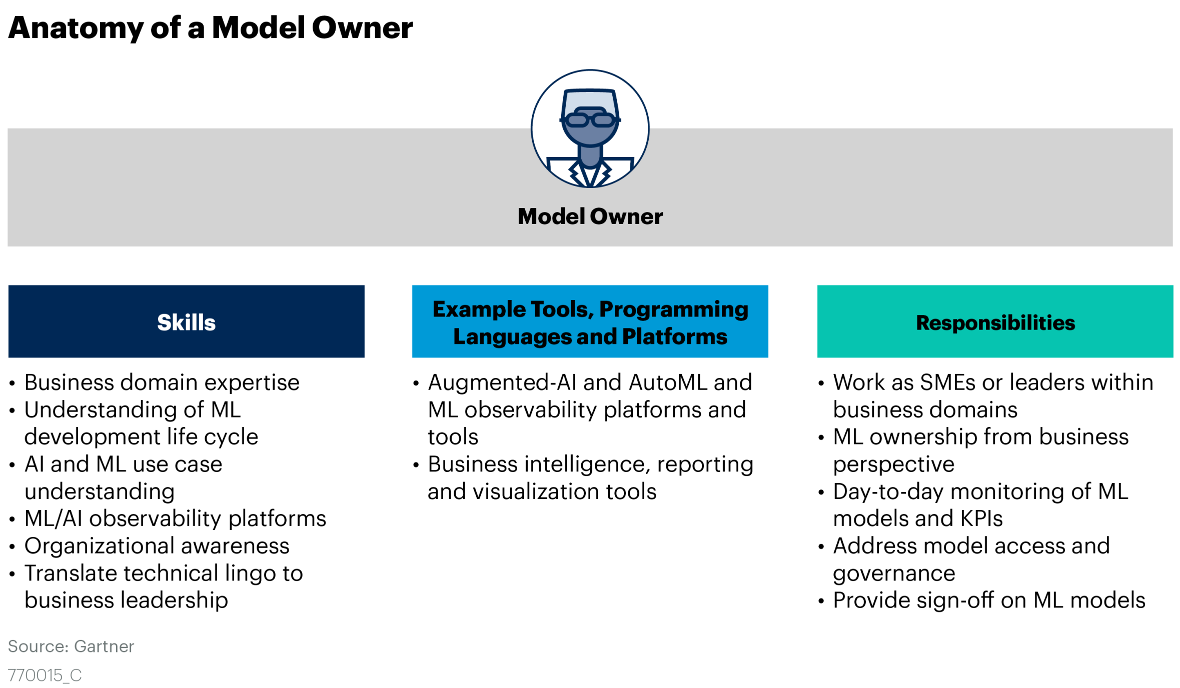
- A willingness to learn new technologies, platforms, algorithms and approaches to ML and AI. This will enable them to keep themselves updated on the newest developments within the AI space so they can keep their business units up to date. This can help drive data literacy within the organization.

- Presentation skills. This skill is key and encompasses the use of presentation slides or even charts to help explain and understand the evolution of the ML model.

- Organizational knowledge. This will help model owners to understand the overall enterprise domain they operate in and what the long-term goals of their respective business domains look like. As such, they can offer valuable feedback to senior leadership on AI strategy and governance.

Figure 7 provides an overview of the model owner.

## Figure 7. Anatomy of a Model Owner



### Upskill Path

Because model owners will be business experts within their functional groups, their upskill path has to be catered accordingly. If they are operating at the analyst level, then they can add data science skills to their repertoire by taking on the work of citizen data scientists. This jump does not require learning in-depth technical data science tools such as open-source tools, frameworks and programming languages, but rather focuses more toward self-serve analytics. Because model owners already understand the business use case of the respective ML models, acquiring some technical skills can result in the formation of an effective citizen data scientist. For more details, refer to the Citizen Data Scientist section of this document.

Model owners, who function as managers or senior leaders within their functional groups, can look to position themselves as analytics leaders within the enterprise and should look at the head of AI or the chief data analytics officer (CDAO) role. CDAO is an emerging role within data and analytics and focuses on executive accountability to drive data mandates. CDAOs help organizations become data-driven by aligning business strategies with data and analytics initiatives and providing an analytics focus to business decision making at the executive level. For more details, read The Chief Data and Analytics Officer's Journey to Business Success.

Model Owners can also emerge from the risk management domains and can bring the risk perspective to AI solutions. They can provide guidance and help improve AI solutions by incorporating security and cybersecurity best practices. According to How Organizations Manage AI Information Risk Today, risk management is still very limited within the AI domains. However, CISOs rank it as the second most urgent digital technology trend to secure in the near future. Adding AI and ML knowledge to information risk officers can help them better plan and integrate enterprise risk management processes into AI solutions. Upskill paths can involve moving toward the CISO or CRO roles.

## Model Validator

### Role and Responsibility

As ML has matured, it has started to borrow best practices from the software engineering domain. This has included DevOps, and the application of DevOps to ML is known as MLOps. Another domain within software engineering is quality assurance. ML/AI models are often considered black boxes, and the inability to explain them increases risk and mistrust in the usage of ML models. However, it is hard to test and explain the ML models because, unlike software testing, they are difficult to break into different components (unit testing), and the outputs are usually probabilistic and nondeterministic. Apart from business explainability, regulatory requirements, such as GDPR, HIPAA and U.S. Federal Reserve SR 11-7 are also crucial driving factors to develop explainability in AI/ML work. For more information, read Video: Why Is Responsible AI Important for Data and Analytics Professionals?

Model validators can be data scientists who take on the function of a QA engineer and work independently of data scientists that are developing the ML solution. They will operate under a shorter period of time, and their primary function is to evaluate and test the ML model. They will focus on model testing, interpretation and explainability. Some of the responsibilities they are expected to have are as follows:

- **Validating the relevancy of data.** This involves verifying data's completeness, integrity, appropriateness (removing bias, quality assurance) and ensuring that preprocessing has been standardized on both training and test data.

- **Use case verification.** Validators will understand and ensure the ML solution has been built according to business requirements as established from the model owner or the functional unit.

- Testing the ML model. They can conduct scenario analysis to **ensure the model is resistant to any severe events.** Apart from this, they should also work with ML engineers to set up testing environments with the availability of testing data.

- **Model documentation and reproducibility.** Validators should ensure the ML development documentation details the steps of data extraction, development strategy, model development and design, and model performance. It must be ensured the model can be reproduced with the given instructions.

- **Model explainability and fairness.** Because most ML models are considered black-box solutions, validators should be able to quantify and test the explainability and transparency of the ML model and help explain the model's behavior and reasons behind predictions. Validators will work with model owners and business teams to help set the definition for model fairness. For a detailed overview on AI explainability, read Incorporate Explainability and Fairness Within the AI Platform.

- **Model robustness.** Validators should work with ML engineers to ensure the model produces stable performance in the event any data or any relationships change. This involves testing for drift, noise and bias, as well as developing monitoring policies for deployed models.

- **Selection of AI explainability tools.** Validators should work with ML architects and data scientists in selecting the appropriate tools, platforms and methodologies to help in AI explainability and fairness. For details, see Cool Vendors in AI Governance and Responsible AI — From Principles to Practice.

### Skills Required

Model validators are expected to have similar expertise within the data science domain as regular data scientists and, as such, should possess both technical and nontechnical skills:

#### Technical Skills

- Strong background in statistics, mathematics, computer science with programming knowledge in Python, R, SQL, MATLAB; open-source frameworks such as TensorFlow and PyTorch; and development studios such as PyCharm and Spyder.

- Knowledge of ML/AI cloud platforms such as Amazon SageMaker, Azure Machine Learning Studio, Google Vertex AI and IBM Watson Studio.

- Design and development knowledge of the entire ML life cycle and a strong understanding of ML algorithms and their use cases.

- Proficiency using ML testing methodologies and associated tools and frameworks, explainability and interpretability. Examples of methods include evaluation metrics and data slices, manual error analysis, naive single prediction tests, directional expectation tests, and invariance tests. Snorkel is an open-source library used for data slice analyses, whereas Alibi focuses on providing code for black-box algorithms. LIME, SHAP, DeepLIFT and InterpretML (from Microsoft) are some of the tools used for ML model transparency and explainability. IBM's  AI Fairness 360 is an open-source Python toolkit catering to AI explainability as well.

- Data visualization and reporting skills to help in clustering and data exploration. This can help explain a model's input data behavior to data scientists and model owners. Examples include BI tools like Tableau and Microsoft Power BI and Python libraries such as seaborn.

### Nontechnical Skills

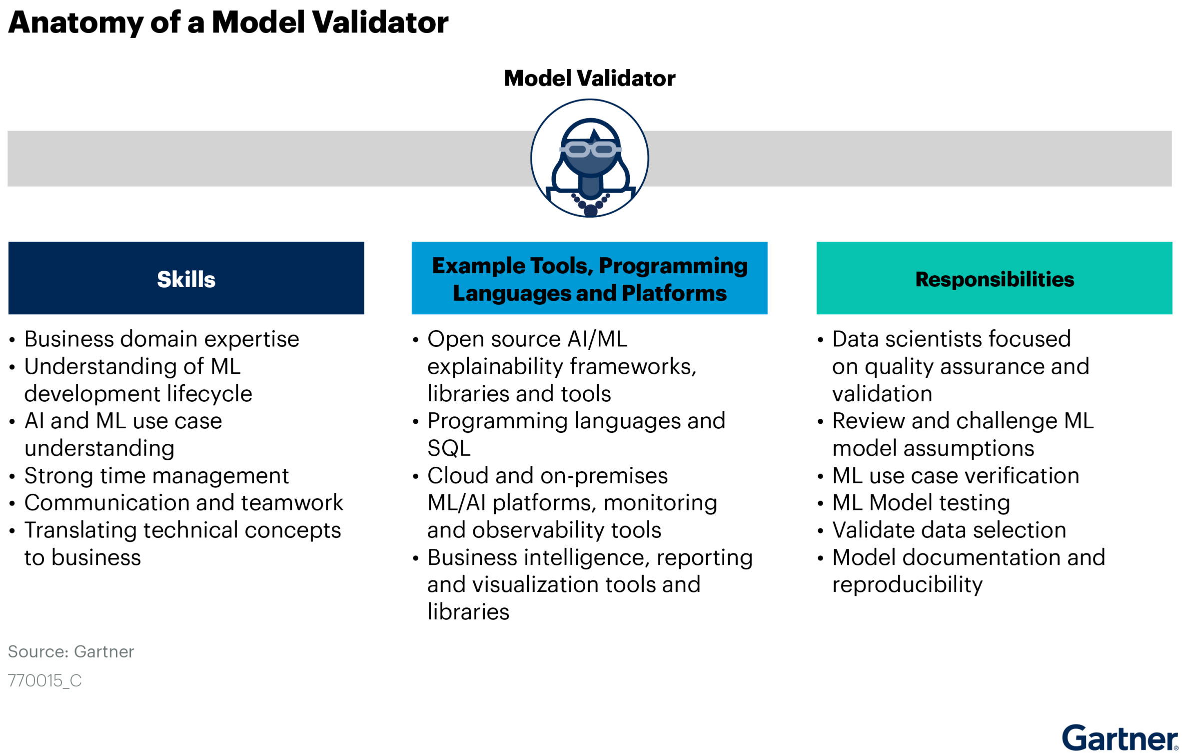
Model validators will need to have strong soft skills as well because they will take on the challenging task of explaining model behavior to the business. They should have the following skills:

- Strong communication, teamwork and interpersonal skills. Validators will have to work with model owners, business analysts and data scientists to help define the explainability and testing framework.

- Solid business domain knowledge. They should have an inherent understanding of how key metrics work, the definition of KPIs, and the driving factors behind patterns and observations.
- Ability to explain technical jargon to business users. Because validators will be explaining the behavior of the ML models, they should be able to detail, in layman's terms, why the model behaves the way it does.
- Time management. Because validators will often be operating under a shorter time span to complete testing and validation, they should have strong time and calendar management skills.

For an overview of the model validator, see Figure 8.

Figure 8. Anatomy of a Model Validator




### Upskill Path

Model validators can transition to a full-time data scientist role because they already possess a strong understanding of data science practices. They need to develop in-depth expertise on the ML development life cycle and strengthen coding and algorithm knowledge. Because they work in close proximity to data scientists, job shadowing can be an excellent way to gain skills while working

Another option can be switching to the software engineering domain and pursuing the traditional quality assurance role. The QA role has been long established and will require training on software development life cycle, authoring software test cases and automated testing as well as learning new tools and platforms like Jira Software and Confluence.

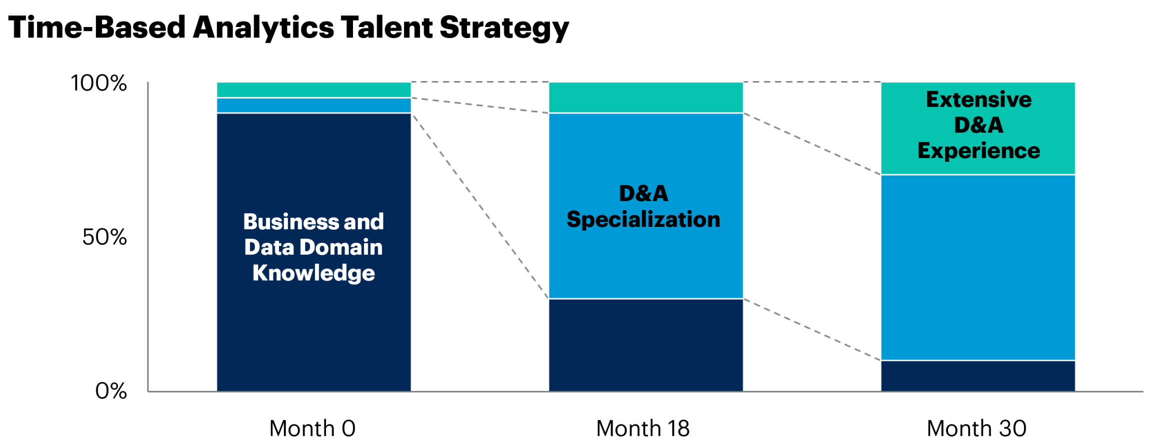
Validators should also pursue management roles within the data science team, such as manager for data science. Because they have been focused more on the testing and explainability aspect of ML development, they can play a crucial role in managing and developing AI solutions with a focus on transparency and ethics. They will need an aptitude for team management and delegation, vision setting, and leadership skills. Pursuing graduate school (MBA or MMA) can be an option, while a more cost-effective solution can also be to pursue executive education credentials such as Yale's Accelerated Management Program and Columbia Business School's Columbia Management Essentials.

## The AI Team

Roles discussed in this paper should work together as part of a core AI team and help drive advanced analytics initiatives. This is backed by Gartner's AI in Organization's Survey 2021, which indicates that 80% of organizations have a formal AI team and that business units' trust and readiness to use AI is higher with a formal AI team. The ability to get AI POCs into production is higher, and organizations are more likely to have a process to continuously evaluate AI initiatives with a formal team. Hence, an AI team with clearly defined roles and responsibilities can go a long way toward the strategic implementation of AI within organizations.

A question analytics leaders are often faced with is when to hire, which roles to prioritize and how to hire or train employees. There is no definite answer to this, and it will vary based on the organization's analytics maturity, availability of skilled personnel and project complexity. In Case Study: Internal Data Science Team Development (Eastman), Eastman developed employees internally by initially focusing on analytics-oriented domain experts. Eastman's analytics maturity and training timeline is shown in Figure 9. This is a sound approach because AI use case determination continues to be a major challenge as evidenced in Figure 1. Having domain experts morph into advanced analytics specialists (citizen data scientists) will lead to more nuanced AI development. Eastman initially focused training on basic statistical analyses of business processes, data quality improvement and BI.

## Figure 9. Time-Based Analytics Role Strategy




| Stages of Maturity | 1 Initial Team | 2 Basic Data Science | 3 Advanced Data Science |
|---|---|---|---|
| Deliverables | • Data Sourcing<br>• Data Quality<br>• Reports | • Generic ML<br>• Predictive Analytics Pilots | • Production Models<br>• Specialized Tasks |

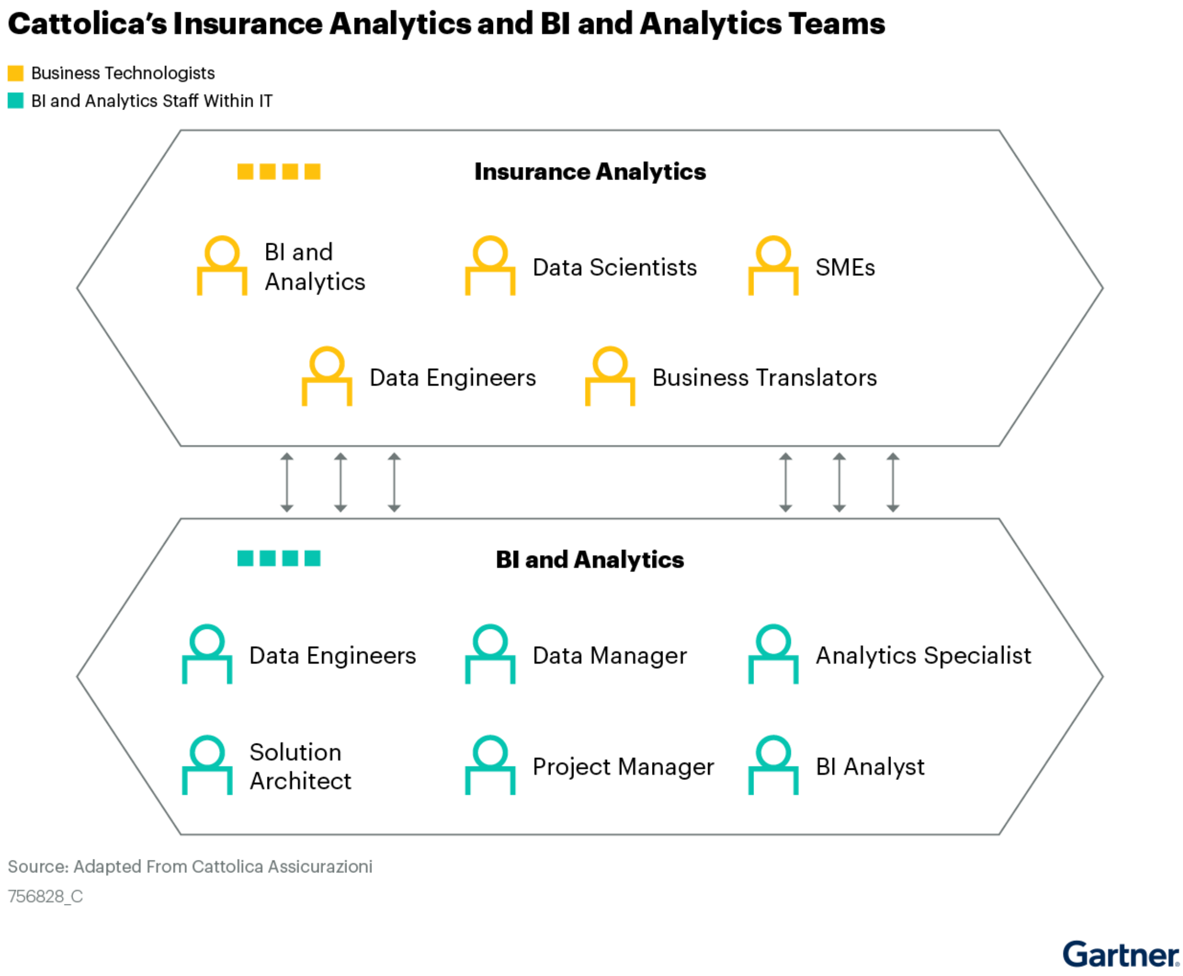
Source: Gartner
770015_C

Eastman gave around 1.5 years for its team to mature and develop expertise in analytics and then made its first external hire from a graduate data science program. Eastman could have hired this resource earlier, but realized the business was not ready for advanced analytics as yet. The new hire brought ML expertise and extended capabilities to include text analysis and forecasting. Around three years in, the team had matured enough to work on complicated NLP cases, and it began hiring for more experienced and skilled roles.

Companies starting on their journey can take a similar approach to Eastman and should incrementally add resources rather than hiring experienced, expensive professionals from the get-go. Initially, citizen data scientists can be developed internally, but as time matures, a skilled data scientist and then ML engineer and architect can be added. As AI matures and proliferates in the organization, model owners and validators should begin to be assigned. This staggered approach will ensure that data literacy propagates within the business units and that the organization becomes AI mature at a gradual and more nuanced pace.

An AI team can conduct centralized operations in the form of a center of excellence (COE). Alternately, its members can be decentralized and embedded within functional units or it can be a hybrid model where AI teams can be deployed in the form of SWAT teams but are centrally managed. For more details, read How to Create an Optimal Data and Analytics Organizational Model. In Case Study: An Approach to Build a D&A Core for Innovation, Cattolica employed a hybrid multidisciplinary fusion team that uses subject matter expertise to identify analytics use cases, thereby greatly increasing the speed of analytics delivery. This “insurance analytics” team consisted of SMEs from the business as well as data science and BI roles. This ensured the solution being developed was use-case driven with the team being a balance of both domain and analytics experts. Their team model can be seen in Figure 10.

### Figure 10. Cattolica’s Insurance Analytics Team Design




Organizations looking to leverage this model for advanced analytics can replace the role of SMEs and business translators with model owners. As this team matures, it can add more specialized roles like the ML Engineer, ML Architect and the model validator.

## Recommendations

Advanced analytics continues to surge in usage, and organizations face new challenges as use cases and analytics proliferation evolve and expand. Gartner recommends that technical professionals looking to work in advanced analytics initiatives should concentrate on upskilling and aligning their skills with the most pressing challenges around AI and ML development. These challenges revolve around data complexity and access, skills shortage, and building trust in ML and AI models.

Data scientists, ML engineers, citizen data scientists and ML architects should upskill to learn more about current trends and best practices in the data management discipline. This includes understanding the different data architecture approaches and differences in data lakes, data lakehouses and data warehouses. They should be active participants during the planning and development of these architectures and provide recommendations on how data will be used for ML and AI development. This will ensure that the data access layers will be prepared according to the data requirements for ML and AI development and will reduce the challenges in data accessibility and complexity.

Another area of concern has been on the AI explainability, testing and ownership front. ML models are often considered black box, and business trust in accepting the output of ML models has remained a challenge. Gartner recommends technical professionals explore roles catered to ML model explainability and ownership, apart from core machine learning and data science. Model validators build trust in AI and ML solutions by playing the role of a QA engineer. They will test the ML model for fairness and trust and ensure that it is free from bias and its outcomes are explainable to business. Subject matter experts from business domains can play the role of model owners and provide ownership on the ML models as well as provide continuous feedback on day-to-day operations from the business perspective. Model owner will not be a specific title within the advanced analytics team, but rather a dual-functioning role. Model owners will usually be SMEs from the respective business units for which the AI solution is being developed. Technical and business professionals should look to explore these two emerging roles as part of their training and upskilling in ML and AI. The addition of these roles to a core AI team brings balance in the form of providing quality assurance, testing, monitoring and explainability for the AI solution.

Advanced analytics professionals should not work in isolation, but instead, should work together as part of a team. Having a well-rounded hybrid AI team taking advantage of both domain and analytics experts can go a long way toward ensuring the AI solution development is use-case driven. The inclusion of SMEs within the advanced analytics team can ensure the AI/ML solutions are curated to the business' needs. More skilled, experienced hires should be added as the organization matures. This ensures that advanced analytics proliferation is constant and that data literacy spreads in a uniform manner.

## Conclusion

AI usage and adoption in companies is increasing at a fast pace, and technical professionals looking to work in challenging roles within the AI space should look to upskill themselves accordingly. Emerging roles of the model validator and model owner help in validating and ensuring the AI solution meets the customer's expectations. Data scientists and ML engineers will be asked to play increasingly hybrid roles where the functionalities of both will overlap as more and more companies move toward productionizing their ML models. Citizen data scientists can play a crucial role in spreading data literacy within their functional units as well as help lay the foundations for future AI development. The ML architect will assume more responsibility, especially toward governance, ethics and security, as AI solutions take center stage and proliferate into the mainstream enterprise application architecture. It is important for these roles to work together as part of a core AI team to evaluate AI propositions, define best practices and achieve greater strategic success by leveraging the skills of each individual role accordingly.

## Evidence

**2021 Gartner AI in Organizations Survey:** This survey was conducted to understand the keys to successful AI implementations and the barriers to the operationalization of AI. The research was conducted online from October through December 2021 among 699 respondents from organizations in the U.S., Germany and the U.K. Quotas were established for company size and for industries to ensure a good representation across the sample. Organizations were required to have developed AI or intended to deploy AI within the next three years. Respondents were required 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. Respondents were also required 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 survey was developed collaboratively by a team of Gartner analysts and Gartner's Research Data, Analytics and Tools team. Disclaimer: Results of this survey do not represent global findings or the market as a whole, but reflect the sentiments of the respondents and companies surveyed.

## Recommended by the Author

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

Machine Learning Playbook for Data and Analytics Professionals

What Are Must-Have Roles for Data and Analytics?

Incorporate Explainability and Fairness Within the AI Platform

Magic Quadrant for Data Science and Machine Learning Platforms

# Actionable, objective insight

**Explore these additional complimentary resources and tools for Data & Analytics Leaders:**

**Webinar**
### 100 Data & Analytics Predictions Through 2026
Elevate D&A strategies to advance business problem solving.

**Watch Now**

**Webinar**
### Gartner 2022 Analytics & BI Platforms Magic Quadrant Highlights
Explore the latest insights on market trends and vendor positions.

**Watch Now**

**Roadmap**
### Migrating Data and Analytics Architectures to the Cloud: Roadmap
Build an effective strategy to embrace multicloud complexity.

**Download Roadmap**

**eBook**
### Essential Guide to Augmenting Decisions With Artificial Intelligence
Decision automation can drive competitive advantage. Know when and how to use it.

**Download eBook**

Already a client?
Get access to even more resources in your client portal. Log In

# Connect With Us

Get actionable, objective insight to deliver on your mission-critical priorities. Our expert guidance and tools enable faster, smarter decisions and stronger performance. Contact us to become a client:

**U.S.:** 1 844 466 7915

**International:** +44 (0) 3330 603 501

Become a Client

**Learn more about Gartner for Data & Analytics Leaders**
gartner.com/en/information-technology

**Stay connected to the latest insights**



Gartner®