Gartner®

Predicts 2021: Artificial Intelligence Core Technologies

Published 22 December 2020 - ID G00735439 - 14 min read

By Analyst(s): Farhan Choudhary, Afraz Jaffri, Svetlana Sicular, Arun Chandrasekaran, Anthony Mullen

Initiatives: Artificial Intelligence

The development process for AI is clear to enterprises today, but the pressing need for AI operationalization will shift the focus to continuous delivery of AI-based systems. Data and analytics leaders must focus on technologies that bridge the gap between development and continuous value delivery.

Additional Perspectives

- Summary Translation: Predicts 2021: Artificial Intelligence Core Technologies (08 January 2021)

More on This Topic
This is part of 2 in-depth collections of research. See the collections:

- Applying AI — Key Trends and Futures
- Over 100 Data and Analytics Predictions Through 2025

Overview

Key Findings

- Artificial intelligence (AI) implementations face obstacles when moving from pilot to production, due to security and ethical concerns, a lack of technology infrastructure and its integration, according to Gartner's 2019 AI in Organizations Survey.
- The technical complexity associated with creating production workflows overwhelms many organizations and exposes them to potential risks and liabilities.
- Gartner's 2019 AI in Organizations Survey highlights that the complexity, scope, quality and accessibility of data remain challenges to the operationalization of AI at many organizations.

Recommendations

Data and analytics leaders looking to accelerate their organization's progress toward operationalizing AI should:

- Invest in creating a platform-centric approach to sustainable AI development by using tools that bridge the gap between development and operations for AI model deployments.
- Catalyze AI implementations and shift the focus toward a scalable approach through strategic use of production-ready cloud services and platforms.
- Ensure the reusability of various AI components by adopting solutions and techniques such as feature stores, model stores or hubs, marketplaces and transfer learning.

Strategic Planning Assumptions

By 2023, organizations that scale graph techniques will deliver five times more AI models, for multiple use cases, into production than those that don't.

By 2024, 70% of enterprises will use cloud and cloud-based AI infrastructure to operationalize AI, thereby significantly alleviating concerns about integration and upscaling.

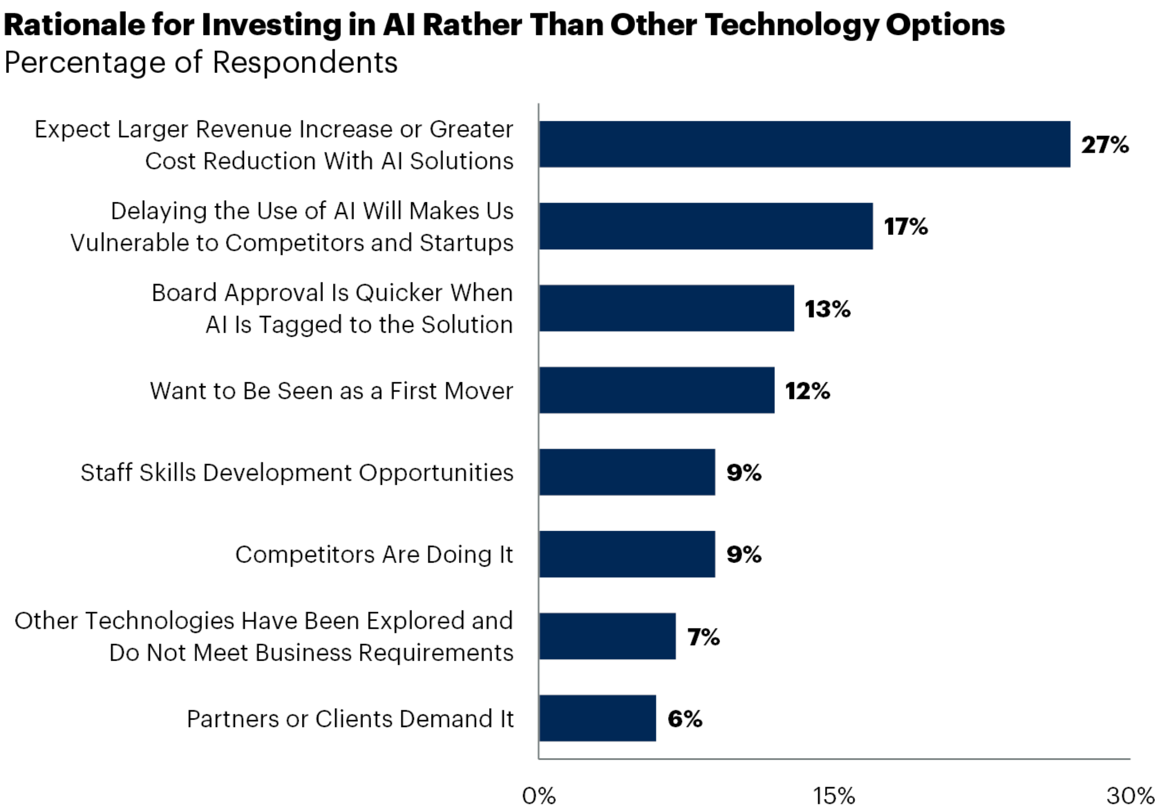
By 2024, use of synthetic data and transfer learning will halve the volume of real data needed for machine learning.

Analysis

What You Need to Know

More than one-quarter (27%) of the respondents to Gartner's 2019 AI in Organizations Survey identified revenue increase or cost reduction as a primary reason for investing in AI rather than other technology options (see Figure 1). Organizations are now interested not just in a "pathway" to production but in a "speedway" to production. However, most enterprises become dismayed when they find out how exceptionally challenging it is to operationalize AI, as opposed to how (deceptively) easy it is to perform proof of concepts (POCs) or small-scale projects.

Figure 1: Rationale for Investing in AI Rather Than Other Technology Options



n = 607 (all respondents)

Q. What is most often your organization's rationale for investing in AI versus investing in other technology options?

Source: Gartner 2019, AI in Organizations Survey

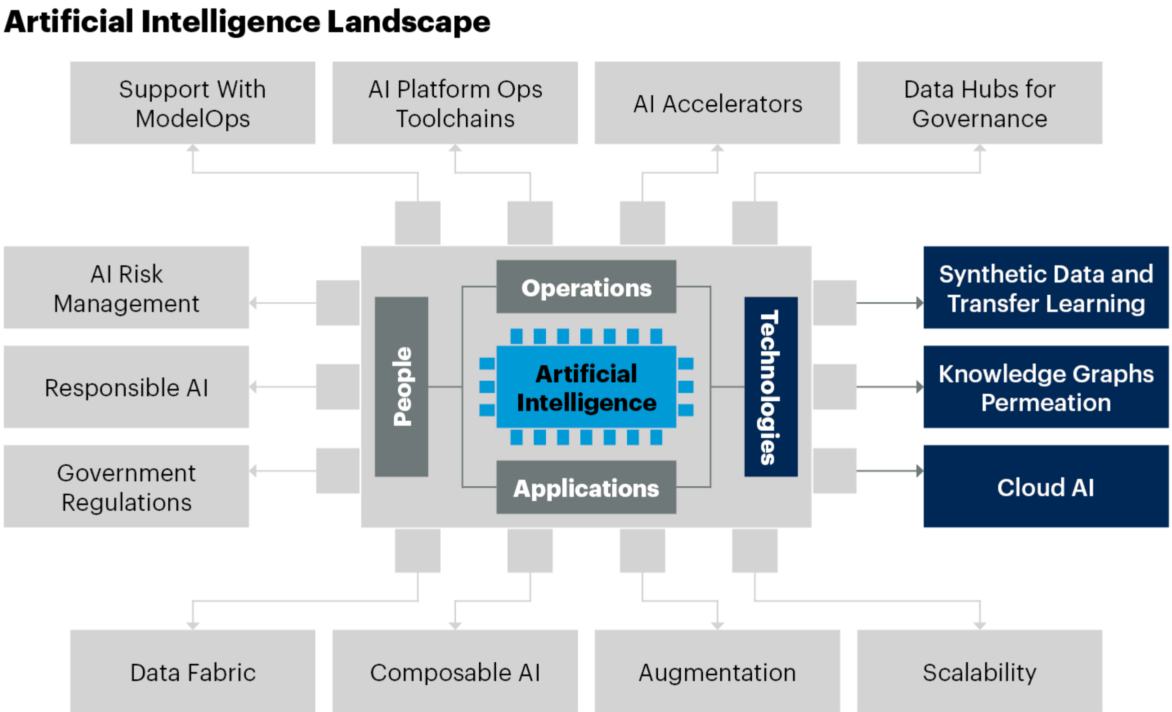
735439_C

Gartner

Data and analytics leaders are struggling to navigate the maze-like AI ecosystem. They need to craft strategies meticulously that account for all of the following (see Figure 2):

- Core technologies
- Applications
- People
- Operations

Figure 2: Artificial Intelligence Landscape



Source: Gartner
735439_C

Strategic Planning Assumptions

Strategic Planning Assumption: By 2023, organizations that scale graph techniques will deliver five times more AI models, for multiple use cases, into production than those that don't.

Analysis by: Afraz Jaffri, Svetlana Sicular

Key Findings:

- Complex optimization problems where constraints are hard to define or difficult to manage can be understood and expressed more easily using graph data structures, after building ontologies that help combine and hierarchize multiple dimensions.
- Preserving metadata, rules and contextual information in a knowledge graph opens up methods of delivering explainable AI solutions, thus enabling use cases that cannot be served by today's "black box" models.
- Traditional machine learning algorithms require optimization to reach a result. By contrast, familiar supervised and unsupervised algorithms such as regression, decision trees, collaborative filtering and clustering, can be run on graph data structures, often with substantial reductions in model size and training time.
- Graph representation learning is emerging as a new field of machine learning that uses the inherent data structure of a graph to build a set of features based on similarity between nodes. It significantly reduces the time spent on feature engineering.

Market Implications:

- Graph databases will increasingly be used for machine learning tasks as the scalability and performance for specific use cases equal or exceed existing processing engines. This will result in the process of data frame wrangling and creation needing to be replaced by models that run natively in graph databases.
- ModelOps for graph databases and knowledge graphs will extend the nascent discipline of ModelOps with tools, processes and standards for working with graph data structures in production environments. Developments will include deployment of graphs on edge devices, stream processing, version control, named graph registries and coreference stores.

- AI solutions will evolve from being based on one type of model, or ensemble of models, to being made of composite models, with graph techniques playing a prominent role. The use of composite AI will require a broader and deeper set of data science and AI skills. Specialist roles such as graph engineer and ontology manager will appear, and existing roles within data science teams will become proficient in graph techniques.
- Existing use cases that were seen as too difficult to solve using existing tools will be opened up for solving by graph algorithms (see Case Study: Answering Critical Business Questions With Graph Analytics (Jaguar Land Rover)). Problems that are expressed using linear programming with constraints and rules can become complex and difficult to interpret. Using shortest path algorithms on a graph can be the quickest way to solve many optimization problems. Another common use of graph algorithms is to answer “find similar” types of questions — For example, to identify a treatment plan for a patient based on other patients with similar characteristics, or to discover sustainable materials that have similar properties to the materials currently in use.
- As graph data structures are used to represent and model data, more attention will focus on developing, enhancing and applying graph neural networks (GNNs). In particular, computer vision applications, natural language processing and recommender systems can be enhanced by GNNs.
- AI systems that have explainability as a standard feature will be expected. Deep learning models that are effectively black boxes due to the inability to decipher what each layer or node in a neural network is contributing can be mapped to knowledge graphs, thus providing semantic meaning for the interpretation of such models. In addition, ontologies and reasoning engines provide the ability to determine what sort of explanation is required, depending on the usage context.

Recommendations:

- Utilize the capabilities of existing data science and machine learning platforms that handle graph data structures. Educate data scientists and machine learning engineers about core graph algorithms and database query languages.
- Identify potential use cases that can be simplified or accelerated with graphs for machine learning by evaluating existing models that require intensive data preparation and feature selection workflows.

- Examine business processes that have a high potential for optimization through the use of graph techniques and algorithms by creating a conceptual graph domain model for the process and testing scenarios that graph algorithms could solve.
- Start to build a knowledge graph for a business domain where applications and data are potentially coupled but siloed from each other by taking a minimum viable approach to schema and graph development.

Related Research:

How to Build Knowledge Graphs that Enable AI-Driven Enterprise Applications

COVID-19 Demands Urgent Use of Graph Data Management and Analytics

Demystifying the Data Fabric

Cool Vendors in Graph Technologies

Analysis by: Farhan Choudhary

Key Findings:

- Cloud compute costs remain a huge concern when operationalizing AI on the cloud. However, this concern should be weighed against the advantages offered, such as faster times to operationalization especially for early-stage AI adopters (through “precanned” solutions), better security, flexibility and ease of scaling. Whether or not operationalizing AI in the cloud is prioritized, having a cloud migration or cloud-only strategy necessitates serious consideration of such operationalization.
- Deploying AI in production environments requires robust infrastructure, workflow management, role management, flexible provisioning, management and synchronization with other IT systems and services, and the agility to adapt to rapidly changing parameters and processes.
- AI infrastructure that caters for on-premises software cannot scale up with ease and flexibility as the cloud can, especially as new, specialized AI-based compute hardware becomes more accessible in the cloud

- Currently, the hyperscale cloud vendors provide a more accessible and flexible hybrid approach to scaling AI, with both hybrid cloud and on-premises capabilities. Since these hyperscalers are aiming to bridge the hybrid cloud and cloud/on-premises gaps, vendors in the data science and machine learning market will have to significantly develop their all-hybrid (multicloud/cloud and on-premises) product strategies. This will make cloud AI more ready for the future.

Market Implications:

Cloud AI makes sense for organizations looking to expand their AI project portfolio by hedging the risks associated with either on-premises or purely open-source AI. It can benefit them in terms of:

- **Flexible and elastic provisioning of various managed services and portable solutions:** The proliferation of precanned AI solutions, robust parallel compute abilities, and diverse data processing/management capabilities offered by cloud-based providers promises faster times to operationalization and business value generation. Cloud AI also gives a significant edge to organizations in terms of provisioning time, elastic computing and provision of metered access to high-performing infrastructure that, if it were installed on-premises would entail a huge upfront cost.
- **Data gravity:** With incoming data volumes growing and becoming more diverse by the day, cloud-based AI enables better interaction and insight generation in a unified environment, in contrast to upgrading data centers and hardware infrastructure to match growing demand, which would not be cost-efficient or sustainable. Cloud AI also ensures good DataOps capabilities (see Innovation Insight for DataOps) and ModelOps capabilities (see Innovation Insight for ModelOps), which are key to operationalizing AI.
- **Accelerated computation:** To drive rapid innovation, organizations can use the cloud computation capabilities offered by cloud providers to fulfill the requirements that AI projects may have on a real-time basis.
- **Multidisciplinary team collaboration:** For most enterprises, curating capabilities (of both technology and people) on-premises creates friction, competition and unresolvable challenges during multiple hand-offs amid projects. Data and analytics leaders can reduce the complexity by using cloud-hosted AI capabilities for application deployment, such as those of APIs, platform as a service (PaaS) offerings, program functions and publish-subscribe infrastructures.

- **Addressing skills gaps:** Cloud-based AI enables easy-to-use solutions for nonexperts in some use cases. This makes experimentation and the scaling up of AI projects easier for organizations that lack, or have limited access to, AI experts.

Recommendations:

- Use the precanned AI solutions, robust parallel compute abilities and diverse data processing/management capabilities offered by cloud-based providers to improve agility in AI implementations.
- Define a production-grade AI strategy by selecting a cloud ecosystem that enables communication between DataOps, ModelOps and DevOps ecosystems.
- Prioritize projects and use cases that work best with cloud-based solutions and do not require heavy investment in infrastructure or skills — for example, natural language processing, computer vision and augmented machine learning.

Related Research:

Innovation Insight for ModelOps

Magic Quadrant for Cloud AI Developer Services

Critical Capabilities for Cloud AI Developer Services

Cool Vendors in AI Core Technologies

Hype Cycle for Artificial Intelligence 2020

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

Analysis by: Arun Chandrasekaran and Anthony Mullen

Key Findings:

- Poor data quality, inaccessibility of data, lack of adequate training datasets, lack of annotated datasets, and bias in training data are among the top data-related challenges to AI operationalization.

- Advances in generative adversarial networks (GANs) and the potential to apply transfer learning techniques will enable more organizations to train their AI models using artificially generated data (also known as synthetic data).
- Synthetic data can be used to augment real-world data in use cases where there is dearth of data and in scenarios where privacy and cost-effectiveness are priorities.
- Synthetic data can reduce delays in operationalizing AI models, where models can be pretrained with subsets of real-world data and synthetic data and complemented by more real-world data later in the AI life cycle.

Market Implications:

Success with AI initiatives requires quick access to high-quality annotated data in order to train AI models. However, getting access to real-world data that is of high quality (that is, well labeled and audited for bias) is often challenging, particularly for deep-learning use cases. The data challenges are exacerbated for novel AI use cases, due to the dearth of real-world data or the prohibitive cost of making such data production-ready. Use of synthetic data can be effective in such scenarios.

Synthetic data is a class of data that is artificially generated, based on real-world situations. It can be used to train AI models for scenarios such as:

- Emerging AI use cases for which little or no real data is available.
- Use cases for which data must be guaranteed to be anonymous or for which privacy must be preserved (such as with uses of medical data).
- Augmentation of real data, especially where the cost of data generation is high or there is a need to balance class distribution within existing training data (such as with population data).

Synthetic data can be generated and enriched by several methods, such as the use of simulation and gaming engines or generative AI techniques involving GANs and transformers. Today, use of GANs is a popular approach for synthetic data, in cases where a variety of GAN architectures (such as CycleGAN, StyleGAN and PixelRNN) is used. Data quality issues associated with synthetic data from GANs can be improved by using transfer learning techniques. Transfer learning improves models trained on synthetic data by using real data to fine-tune them. It takes the knowledge learned from tasks for which a lot of labeled data is available and applies it to settings for which only a limited amount of labeled data is available.

Synthetic data is already used for computer vision use cases in industries such as agriculture, automotive, defense, finance, industrial and retail. In addition, it is starting to be used for conversational AI and other natural language processing use cases, with sectors like healthcare and finance focusing on the use of synthetic data as a privacy-preserving technique.

While synthetic data techniques can score quite highly for cost-effectiveness and privacy, they do have limitations. For example:

- The quality of synthetic data often depends on the quality of the model that created it and the input “seed” dataset.
- If seed datasets will change, it is necessary to regenerate synthetic data using the new characteristics for it to enable meaningful model accuracy.
- Using synthetic data requires additional verification steps, such as comparison of model results with human-annotated, real-world data, to ensure the fidelity of results.

Recommendations:

- Identify use cases within your organization for which your real-world training data could be supplemented with synthetic data. Also, identify use cases that previously could not be addressed due to lack of data.
- Work with security and legal teams to understand the potential for synthetic data usage, in cases where large volumes of personal data is being utilized and privacy needs are high.
- Be cautious about the possibility of technological abuse of GANs. Ensure optimal risk mitigation, model validation and auditing.

Related Research:

Hype Cycle for Data Science and Machine Learning, 2020

Data Science and Machine Learning Trends You Can't Ignore

Cool Vendors in AI Core Technologies

Tech Providers 2025: Why Small Data Is the Future of AI

Preserving Privacy While Using Personal Data for AI Training

Boost Your Training Data for Better Machine Learning

A Look Back

In response to your requests, we are taking a look back at some key predictions from previous years. We have intentionally selected predictions from opposite ends of the scale — one where we were wholly or largely on target, as well as one we missed.

This topic area is too new to have on-target or missed predictions.

Evidence

Gartner's 2019 AI in Organizations Survey was conducted online in November and December 2019. There were 607 respondents from organizations in the U.S., Germany and the U.K. Quotas were applied for company sizes and industries to ensure the sample was representative. Respondents were from organizations that had already developed AI or that intended to deploy AI within the next three years.

Respondents were either members of their organization's corporate leadership team or people who reported to corporate leadership roles, had a high level of involvement with at least one AI initiative, and were responsible for at least one of the following tasks:

- Determining business objectives for AI
- Measuring the value derived from AI initiatives
- Managing AI initiatives' development and implementation

The results of this survey represent neither global findings nor the market as a whole. Rather they reflect the views of the respondents and companies surveyed.

Document Revision History

Predicts 2020: Artificial Intelligence Core Technologies - 11 December 2019

Recommended by the Authors

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

Innovation Insight for ModelOps

Predicts 2021: Artificial Intelligence in Enterprise Applications

Predicts 2021: Operational AI Infrastructure and Enabling AI Orchestration Platforms

Predicts 2021: Artificial Intelligence and Its Impact on People and Society

Data Science and Machine Learning Trends You Can't Ignore

Product Roadmap Priorities: Data Science and Machine Learning Platforms