

Accelerating Artificial Intelligence With InterSystems IRIS Data Platform

Technology Brief





Introduction

There's nothing artificial about artificial intelligence (AI). The business world will soon be separated between those who have embraced it to improve their processes and become more agile — and those who failed to adopt AI and fall by the wayside.

Agility starts with initiative. And the time to change is now.

Recent advances in artificial intelligence technologies and storage and processing power are dramatically transforming today's business landscape. More and more organizations are leveraging AI and big data to deliver innovative new services, create new revenue streams, improve customer experiences, and streamline business processes. Not only are established organizations implementing AI-based initiatives, but start-ups are leveraging AI to disrupt practically every industry, increasing the competitive pressure on established organizations.

McKinsey has documented the benefits of applying AI in more than 400 use cases across 19 industries.¹ PwC predicts that AI will contribute \$15.7 trillion to the global economy by 2030.² And according to Gartner, 77 percent of senior managers report that data science is already delivering significant value to their organizations.³

To respond to this shift in the business landscape, InterSystems has developed a complete data platform which speeds and simplifies the development, deployment, and maintenance of real-time, data intensive analytic applications in a variety of industries.

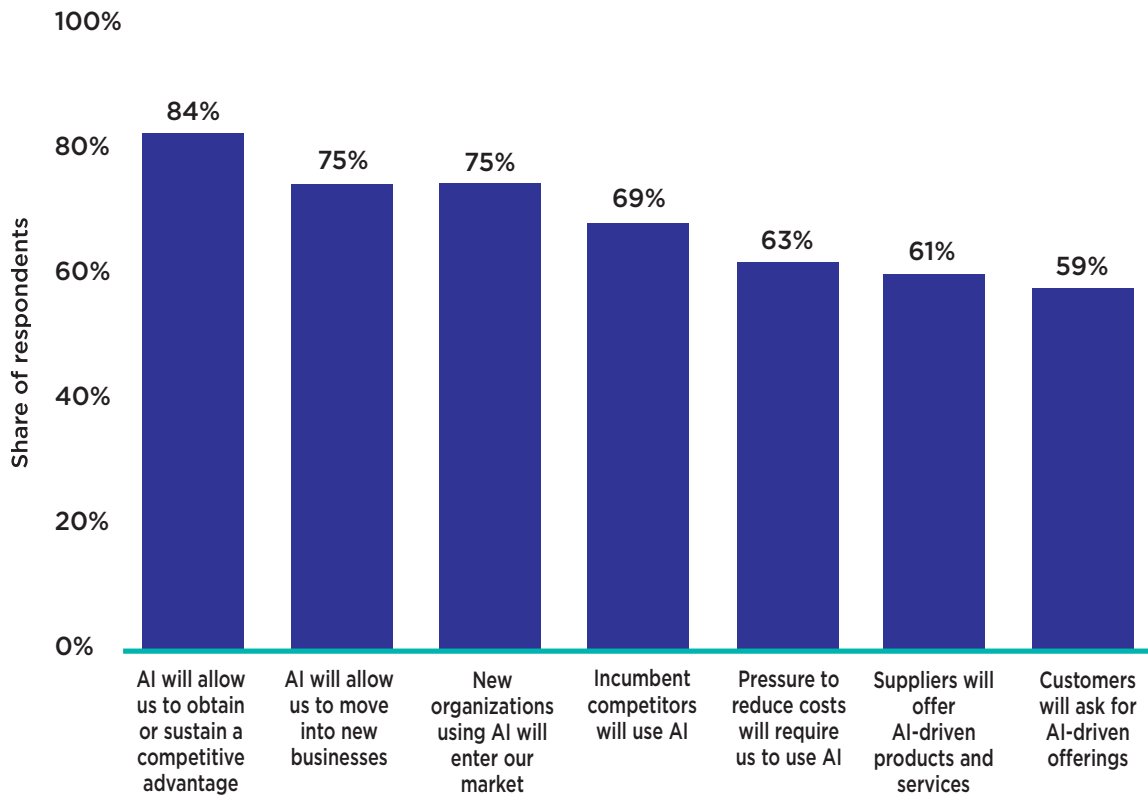
This paper details the various ways in which InterSystems IRIS Data Platform™ accelerates the time to value for enterprise initiatives that leverage artificial intelligence.

¹ <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>

² <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>

³ Gartner, 2018. "Magic Quadrant for Data Science and Machine Learning Platforms."









Reasons Businesses Cite for Adopting AI



Source: Statista 2017 survey of more than 3,000 business executives worldwide⁴

⁴ <https://www.statista.com/statistics/747775/worldwide-reasons-for-adopting-ai/>

AI Adoption by Industry

Sector	Subsector	Potential AI Consumption Impact
Healthcare		3.7
	Providers/Health Services	
	Pharma/Life Sciences	
	Insurance	
	Consumer Health	
Automotive		3.7
	Aftermarket & Repair	
	Component Suppliers	
	Personal Mobility as a Service	
	OEM	
	Financing	
Financial Services		3.3
	Asset Wealth Management	
	Banking and Capital	
	Insurance	
Transportation and Logistics		3.2
	Transportation	
	Logistics	
Technology, Communications and Entertainment		3.1
	Technology	
	Entertainment, Media and Communications	
Retail		3.0
	Consumer Products	
	Portal	
Energy		2.2
	Oil & Gas	
	Power & Utilities	
Manufacturing		2.2
	Industrial Manufacturing	
	Industrial Products/Pure Materials	
Grand Total		3.1

- % Adoption maturity — Near Term (0-3 yr)
- % Adoption maturity — Mid Term (3-7 yr)
- % Adoption maturity — Long Term (7+ yr)

Healthcare 3.7



37%

23%

40%

Automotive 3.7



35%

47%

18%

Financial Services 3.3



41%

59%

0%

Transportation and Logistics 3.2



41%

41%

17%

Technology, Communications and Entertainment 3.1



47%

36%

17%

Retail 3.0



54%

38%

8%

Energy 2.2



39%

44%

17%

Manufacturing 2.2



14%

83%

3%

Sources based on PwC's AI impact index evaluation. Potential scores range from 1-5, with 5 indicating the highest potential impact due to AI, and 1 being the lowest.

InterSystems technology is deployed in thousands of enterprise applications in more than 80 countries. It is used to process the electronic health records of more than 500 million patients and 15 percent of the worldwide financial equity trades made every day.

In addition to our global reach, InterSystems offers top-tier support for the clients we serve through our Worldwide Response Center. Available 24/7 — and staffed by skilled engineers — our goal is to understand your overall business, not just the problem at hand. We are here to support your strategy, planning, performance analysis, migration — and success.

Accelerating AI

While there are plenty of AI tools and frameworks on the market, there is an urgent need to accelerate the application of AI to meet the requirements of the business. InterSystems IRIS Data Platform can deliver many of the capabilities to address this need.

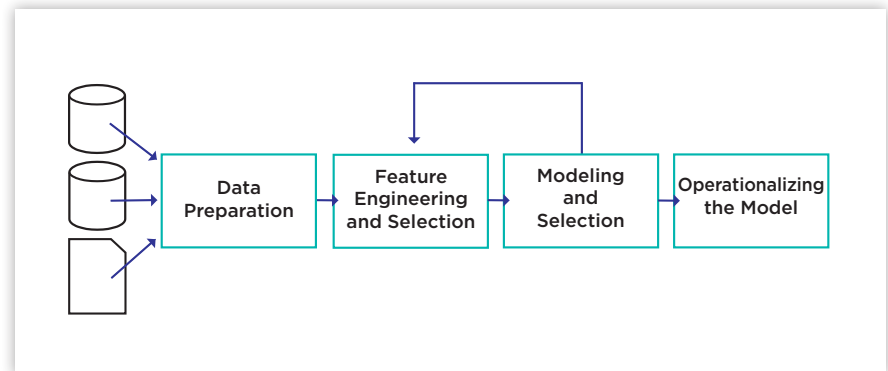
InterSystems IRIS features:

- Horizontally scalable hybrid transaction-analytic processing (HTAP) database management capabilities for executing real-time analytic applications on very large data sets;
- Embedded capabilities for executing structured and unstructured analytics and AI models;
- A flexible multi-model database optimized for AI feature engineering at scale;
- A complete interoperability platform for integrating data and applications, pre-processing data for analytics, and building and executing complex business processes;
- A unified development environment to simplify the creation and maintenance of sophisticated analytic applications.

AI Pipeline

Developing and deploying AI-based initiatives requires the following steps:

- Data preparation
- Feature engineering and selection
- Modeling and tuning
- Executing and operationalizing AI models



Data Preparation

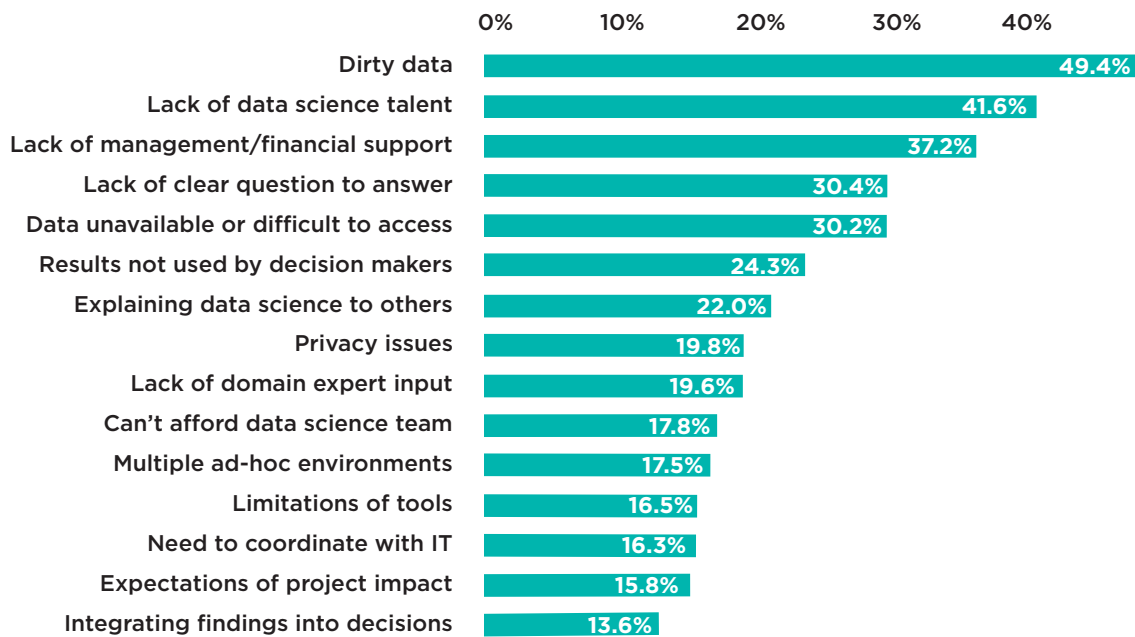
Organizations today are amassing large amounts of data from different sources, in different formats, and with different semantic representations. The more data that can be included in analyses, the greater the opportunity for business benefit, yet raw and dirty data is unlikely to produce accurate results. To make the best use of AI, therefore, large amounts of disparate data must be collected, integrated, and normalized — which is hard work. According to IDG, data scientists typically spend 80 percent of their time on these tasks.⁵

In Kaggle’s 2017 “State of Data Science & Machine Learning” survey, more than 7,000 people involved with data science reported that their most common problem was dealing with dirty data.⁶

⁵ <https://www.infoworld.com/article/3228245/data-science/the-80-20-data-science-dilemma.html>

⁶ <https://www.kaggle.com/surveys/2017>

Barriers to Using Data Science at Work




Source: Kaggle's 2017 "State of Data Science & Machine Learning" survey of 7,276 people.

Although most data scientists would prefer to spend their time analyzing data, the reality is that the vast majority of the work associated with AI deals with "data wrangling," transforming raw data into something useful for AI purposes.

InterSystems IRIS makes it faster and easier for data scientists to collect, integrate, and normalize the data required to build and tune accurate AI models. InterSystems IRIS provides a complete set of embedded integration capabilities, including out-of-the-box connectivity for a wide range of packaged applications, databases, industry standards, protocols, and technologies. Flexible, graphical data transformation capabilities simplify the resolution of differences in semantics and data schemas among data sources, applications, and services. The platform also supports a wide range of industry standards (such as those used in healthcare, financial services, retail, and telecommunications), REST architectures, and web services standards (e.g., JSON, XML, XPATH, XSLT, SOAP, and DTDs).

In addition, InterSystems IRIS provides sophisticated, enterprise-class security, including role-based security with single-row granularity for providing secure access to sensitive data. This feature makes it possible to associate each individual row with a list of authorized users, ensuring that only approved users are able to access specific data items.



“COMING UP WITH FEATURES IS DIFFICULT, TIME-CONSUMING, AND REQUIRES EXPERT KNOWLEDGE. ‘APPLIED MACHINE LEARNING’ IS BASICALLY FEATURE ENGINEERING.”

*Professor Andrew Ng,
Stanford University*



Feature Engineering and Selection

Identifying and creating the best set of features for analysis is critical for generating accurate results with good performance. Feature engineering is the process of transforming raw data into a set of individual attributes that are most relevant for the desired analyses, in formats that can be used by the AI algorithms.

Feature engineering can involve the creation of new or aggregate variables from raw data, and transforming the structure of variables. This can provide information that is more useful than the raw data alone. For example, a new feature could relate the ratio of the past month's revenue to the past 12 months' revenue for each customer.

Creating new features can make it necessary to transform or encode some variables, such as by changing the structure from categorical to numeric. For example, the frequency of each U.S. state (e.g., New York or California) may be relevant to an analysis, but the algorithm may not accept the data represented as one property with multiple string values. Therefore, an example of a transformation or encoding task would be to transform the property (U.S. state) containing 50 different string values (one for each state) into 50 different features (one for each state), each with a numeric value corresponding to the number of occurrences of the state.

Another common requirement associated with feature engineering is to extract meaning and sentiment from free text. For example, the number of times a concept is contained in a document may indicate the relative importance of the concept, and sentiment analysis can determine whether the tone of the text is positive or negative.

InterSystems IRIS speeds and simplifies various tasks associated with feature engineering. It allows users to create new calculated variables, provides embedded business intelligence (multi-dimensional OLAP) functionality, and distributed SQL processing to support complex calculations on very large data sets with high performance. Data is stored once in the database, and the platform allows both relational and non-relational access to the data for flexibility. If it is easier to work with data in rows and columns, developers can interact with the data using SQL commands. Alternatively, the data can be accessed and manipulated as documents, objects, or key-value data.

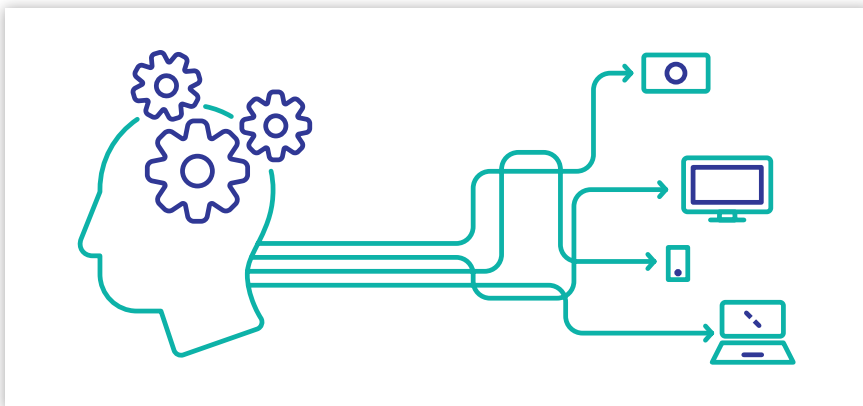
InterSystems IRIS provides built-in natural language processing (NLP) capabilities that infer meaning and sentiment from text. It can automatically identify concepts and relationships in text without upfront work or domain knowledge. These capabilities allow data scientists to extract information from documents, notes fields, social media feeds, and other sources to create new features for analysis.

Feature engineering can produce tens of thousands (or more) of large, sparsely populated matrices. One InterSystems healthcare customer has created 113,000 distinct features to support the development of its AI patient risk models. To handle such tasks with maximum efficiency, InterSystems IRIS stores data in the database as multidimensional arrays and uses horizontal scaling techniques. InterSystems IRIS thus provides high performance with fewer hardware resources for feature engineering tasks.

Modeling and Tuning

Although InterSystems IRIS provides high-performance SQL capabilities, many modeling and tuning tasks are performed using specialized tools. In its 2017 survey, Kaggle lists 49 commonly used modeling tools and identifies Python as the most commonly used data analysis language, followed by R and SQL.⁷

In addition to providing support for xDBC protocols and other standards, InterSystems IRIS integrates directly with Apache Spark via a shard-aware native connector. InterSystems IRIS applications can therefore incorporate Spark processing, and Spark applications can incorporate distributed data from InterSystems IRIS. The connector leverages the partitioned nature of the InterSystems IRIS database so that the Apache Spark worker nodes can automatically connect directly to the data partitions and work in parallel on disparate pieces of data. These parallel, direct connections enable high data throughput and support high-speed data ingestion to horizontally distributed clusters. This architecture also makes it possible for models to be trained incrementally in real time as data is encountered and streamed from production applications, enabling data scientists to keep models current as business conditions and environment variables change.



⁷ <https://www.kaggle.com/surveys/2017>



Operationalizing AI Models

Some applications only require executing AI models independently in batch mode. However, many others require incorporating AI models into business processes that execute in real time in response to events, transactions, or user actions. For example, fraud detection models are often incorporated into real-time credit card transaction processing systems to identify and interrupt potentially fraudulent transactions.

The industry has coined the term “operationalizing” for incorporating AI models into business processes. In addition, Gartner recently coined a new term, “continuous intelligence” in its 2018 “Hype Cycle for Data Science and Machine Learning.” Gartner defines “continuous intelligence” as “real-time analytics ... integrated within a business operation to process current and historical data and then prescribe actions. This innovation provides decision automation or decision support. It uses multiple technologies, such as augmented analytics, event stream processing, optimization, business rule management, and machine learning.” Continuous intelligence is one of a handful of AI subdomains that Gartner rates as “transformational.”⁸

InterSystems IRIS provides a full suite of tools for developing and executing real-time data- and analytics-intensive business processes that connect with internal and external data sources, applications, and services. It also provides graphical tooling that makes it easy to diagram processes, rules, and workflows to support intelligent business processes.

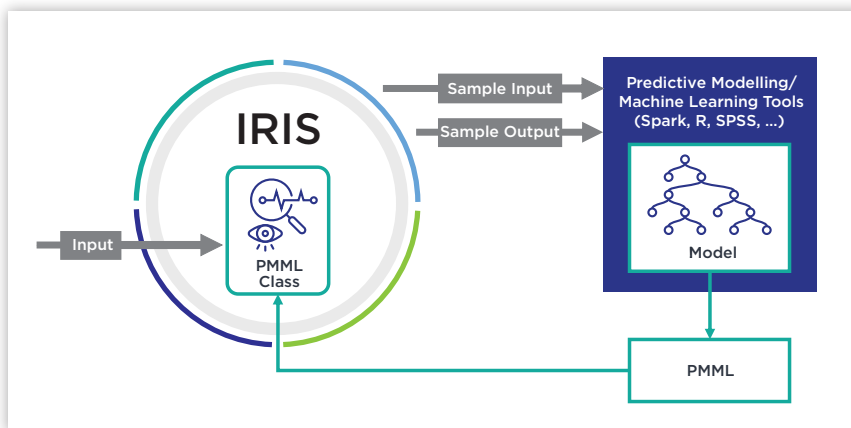
In addition, built-in HTAP capabilities enable AI models to be triggered by events and transactions in real time and executed with a response time measured in milliseconds. This is indispensable to mission-critical applications, such as detecting consumer fraud at the time of purchase, or the risk of customer churn while a call center employee is handling a customer.

AI models created using external tools and applications can be incorporated into InterSystems IRIS applications via the Predictive Model Markup Language (PMML). PMML is an XML standard that fully defines all the parameters of a model developed using an external AI application or framework. Most leading AI tools and frameworks support exporting models via the PMML standard. InterSystems IRIS includes a built-in PMML runtime engine to enable the execution of models developed with external tools without employing any additional tools or passing data across systems.

This architectural approach enables AI models to execute seamlessly within complex, real-time business processes.

⁸ Gartner, 2018. “Hype Cycle for Data Science and Machine Learning.”

Operationalizing AI Models to Execute Within InterSystems IRIS Applications



Built to Scale

InterSystems IRIS database management capabilities have been proven to deliver very high performance at scale in mission-critical enterprise applications, for both transactional (OLTP) and analytic workloads. It supports data ingestion at very high rates and can process analytic workloads simultaneously on both real-time data and very large sets of non-real time data (e.g., historical and reference data) using commodity hardware. InterSystems IRIS is thus ideal for applications that must make real-time decisions by identifying patterns and anomalies at scale. This architecture also performs reliably at scale, even when workloads unexpectedly spike.

Conclusion

InterSystems IRIS is a comprehensive data platform that speeds and simplifies the development, deployment, and maintenance of real-time, data intensive analytic applications. It provides critical capabilities for pre-processing data to support AI modeling and tuning tasks and for incorporating AI models into real-time business processes. In this way, InterSystems IRIS is helping organizations leverage AI and big data to deliver innovative new services, create new revenue streams, improve customer experiences, and streamline business processes to gain competitive advantage — and realize value quickly.

Accelerate your artificial intelligence initiatives and launch quickly with InterSystems IRIS Data Platform.

InterSystems is the engine behind the world's most important applications. In healthcare, finance, government, and other sectors where lives and livelihoods are at stake, InterSystems is the power behind what matters. Founded in 1978, InterSystems is a privately held company headquartered in Cambridge, Massachusetts (USA), with offices worldwide, and its software products are used daily by millions of people in more than 80 countries.

For more information, visit InterSystems.com/IRIS

The power behind what matters.

