

IBM AI Applications

AI maturity framework for enterprise applications

March 2021

IBM

Authors

Rishi Vaish is the Chief Technology Officer (CTO) for IBM AI Applications, a market leading portfolio of enterprise applications spanning Asset Management, Facilities Management, Supply Chain Management, Engineering Lifecycle Management and Weather Business solutions. He is responsible for driving innovation through Machine Learning, Data Science, Site Reliability Engineering and Hybrid Cloud architecture for the portfolio. Through his career Rishi has over 15 years of executive leadership experience in technology, product strategy, product development, product operations and product management from startups to large scale organizations across a variety of industries and technologies. He has deep expertise in AI-based applications, hybrid cloud technologies, cloud computing, software-as-a-service and application middleware. He is passionate about driving innovation, modernization and transforming product and technology organizations to meet scale and growth demands.

vaish@us.ibm.com | www.linkedin.com/in/rishivaish/



Ashish Agrawal leads the Strategy and Portfolio Planning team for IBM AI Applications. He is responsible for collaborating with business and technology leaders to establish long-term growth strategy for the business unit as well as is greatly involved in the execution of the strategy. He works closely with the CTO office as well as the product teams in identifying new technologies like Artificial Intelligence (AI) that should be leveraged to innovate the product portfolio. His 20+ years of Strategy and Operations experience at IBM, BCG, Deloitte, and Shell spans across multiple industries like High Tech, Energy, Oil & Gas, Retail, Healthcare and Life Sciences with a focus on helping clients grow and overcome their complex business challenges.

aagrawa@us.ibm.com | www.linkedin.com/in/ashish-agrawal-972283



Shalini Kapoor is IBM Fellow for Artificial Intelligence and drives research, strategy, and development to maximise the infusion of AI. She is leading IBM's AI market-leading capabilities, driving consistent, reusable implementation patterns, business transformation and value, and making AI accessible and useable to business users, and not just data scientists. She is a torchbearer of Good tech projects in India focusing on STEM and AI for women and underprivileged while working with schools, universities, government and industry leaders. She has 14 patents and is a prolific speaker in academia and various industry forums.

kshalini@in.ibm.com | www.linkedin.com/in/kshalini



Rob Parkin is the Principal Scientist for IBM's AI Applications business unit. He has worked for over 20 years in applying large scale machine learning and optimization technologies to solve core business problems. In his current role, he leads the development of new AI technologies to support IBM's Supply Chain, Weather, IoT, Facilities Management and Engineering Lifecycle product portfolios. Prior to his work at IBM, he held research positions at Stanford University and at the National Center for Atmospheric Research.

rparkin@us.ibm.com | www.linkedin.com/in/rob-parkin-b751b61/



Introduction

Artificial Intelligence (AI) is transforming how we live, work and do business. From smartphones to smart factories, AI is ubiquitous – and whether we realize it or not – it is influencing our everyday interactions with businesses. AI has disrupted our lives by entering the entire gamut of areas like recommendations during ecommerce purchases, personalized video streaming options, predictive texts while typing, and prioritizing social media feeds.

As organizations move on their AI journey, they have to infuse AI in their enterprise applications and sometimes build new AI offerings from scratch. The use cases for AI infusion span industries and product categories, and cover a wide spectrum ranging from providing insights, recommending the next best actions, predicting and forecasting business outcomes to fully automating business processes. Most organizations start their AI journey and soon want to scale AI but get stuck during the process. Identification of all AI capabilities, technologies, client adoption levels, and value derived from AI are some salient metrics that govern the journey's success.

At IBM AI Applications, we are infusing AI into enterprise products ranging from asset management, facilities management, supply chain management, engineering lifecycle management to weather business solutions. Our experience of infusing AI and delivering value to clients has been collated in this white paper and has led us to create the **AI Maturity Framework**. This framework, now a standardized methodology in our organization, is a measure of how mature AI is in any enterprise application. We deliberate on the factors that contribute to AI's maturity and how this is a journey and not an end state. This framework prescribes a methodology for assessing business and technical aspects of AI.

This white paper is a ready reckoner for enterprises across industries to measure the readiness and completeness of AI applications that they have implemented. These enterprises are the ones who have either infused their existing legacy applications with AI or have bought new bespoke AI applications or even those who are evaluating independent software vendors (ISV) selling specific AI applications. In the same light, this whitepaper would be used as a tool by ISVs and system integrators (SI) to measure the maturity of the AI applications that they build and implement for clients. The factors contributing to the maturity are both business- and technology-centric, hence the measurement will be a combination of both. The value of the maturity framework can be derived by a spectrum of business-oriented roles like CXOs, product managers and business leaders who are correlating the extent of the impact of AI to their revenues. The maturity framework is also relevant for technical roles like enterprise architects, application developers and data scientists who can measure the extent of advance AI features that they have included to advance the solution.

The journey of AI infusion in enterprise applications

Maturity curve is always a journey and maximizing AI's value for any organization starts with assessing the current state of AI adoption, gaining consensus on short- and long-term goals, and understanding the execution capabilities and gaps. There are many maturity frameworks available – from Gartner¹, IDC², Ovum³, and others – that can help assess an organization's AI maturity and related skills. We recommend The AI Ladder: Accelerate Your Journey to AI⁴ as a starting point for leaders interested in understanding enterprise AI's value and the steps necessary to implement it at scale. The AI ladder explains how a data strategy is an essential foundation before the AI infusion journey commences. After understanding the AI ladder, assess the maturity of your organization by using IBM AI Maturity Assessment⁵ tool. If you are keen to assess the maturity of the machine learning process, then we recommend the paper on Characterizing Machine Learning Process.⁶

This white paper assumes that you already have an AI strategy in place and have identified the applications where AI needs to be infused and understand the impact that AI will have for the end-users. Once this is achieved, organizations wanting to scale AI in applications ask, "How good is the AI included in my applications?", "What is the level of AI sophistication that I need to include"? "Is there a trade-off between the cost of AI functionalities to be included and the revenues the AI application is impacting?"

This white paper covers these critical questions for business and technology leaders in-depth and explains how to move from one stage to the other while advancing our AI capabilities at each level.



How good is the AI in my enterprise applications?

What constitutes “good” AI is a challenging question to answer. The field of AI is rapidly expanding and there are always better models and techniques available to improve our usage of AI. As a gauge, there are tens of thousands of research papers and articles published each year – at an increasing rate. Thanks to a vast number of programming language libraries, new research makes its way into products very quickly. To be able to answer the question of goodness effectively, we studied various parameters. These range from how it impacts the business users; the direct value clients derive from AI; how sophisticated the chosen tools and frameworks are; the inclusion of bias detection and removal algorithms; the quality of data and cleansing mechanisms; and how explainable the AI is to end-users.

Applications differ in their maturity curve and we observed that the inclusion of the above parameters impacted the application’s success. For example, an application that is low in improving the quality of the input data and has no bias detection would give lower value to end clients. Where you run your AI in enterprise applications adds another dimension to the maturity model. The next generation of enterprise applications will be based on a hybrid cloud software model. This means that the application containers could be running in a public cloud or in some private cloud behind a firewall. There might be a need to monitor the performance of AI models and this feature would need to be factored in the AI design itself.

In this paper, we share the IBM AI Maturity Framework for Enterprise Applications.

The maturity framework outline

The framework is comprised of seven dimensions which are a combination of business and technical capabilities that determine the success of an AI application. The seven dimensions of the maturity framework are shown in Figure 1 below:

A. Impact on your business

B. Value to the end client

C. Technology sophistication

D. Trustworthiness

E. Ease of use

F. AI operating model

G. Data

For better readability, factors mentioned in A and B are purely intended for business leaders measuring the maturity of AI applications. If you are part of an enterprise product company such as an independent software vendor or even if you are a systems integrator, then you are likely to find all of A and B applicable. If you are adding AI to enterprise applications as part of an IT organization, you can modify the business impact dimensions to suit your specific business goals. Dimensions C, D, E, F and G are intended for the technical community who are building technical features in AI applications. For business leaders, understanding dimensions D and E is also essential as the factors described here have a direct impact on usage and trust in AI.

For each of the dimensions, a given AI capability in the enterprise application is ranked on a scale of 0 to 3 and then aggregated into a single score. Based on the scoring of a given AI capability it will fall into one of the three phases also shown in the ladder in Figure 1 below:

– **Silver**

– **Gold**

– **Platinum**

For a given AI capability, the ranking by the seven dimensions is also plotted on a six radii radar plot for easier visualization. For the radar plot, the “impact on your business” dimension (dimension A) is pulled out of the radar since the other six dimensions are specific to the AI capability vs. the business area. As a sample, one of the AI capabilities in our IBM AI Applications is shown in Figure 2 below. As you can see, this capability is on its way from gold to platinum.

Figure 1: AI Maturity assessment framework

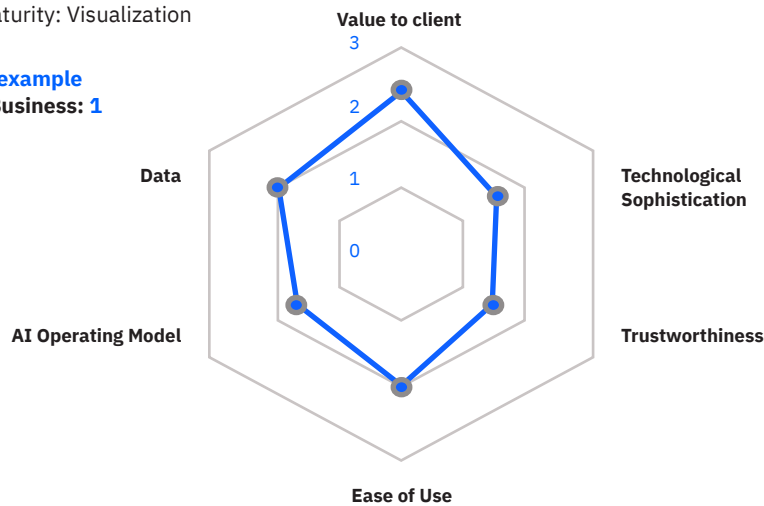
We created a model to assess the AI Maturity of our Offerings to measure progress towards AI Infusion.

Dimensions	Criteria
Impact on our Business	Business Impact, Portfolio Impact
Value to Client	Business Process Outcome, Differentiators
Technology Sophistication	Appropriateness of the technology to the business problem, Learning Techniques, Reuse of Models, Use of Inner Source or Open Source
Trustworthiness	Integrity, Quality, Bias (Fairness), Explainability, Security
Ease of Use	Intuitive for use by the intended user
AI Operating Model	Deployment (Manual, Automated), Update Frequency, Infrastructure/Architecture Scale, Monitoring
Data	Data Acquisition & Instrumentation, Data Management



Figure 2: AI Feature/Offering Maturity: Visualization

Illustrative example
Impact on Business: 1



Maturity framework: What the phases mean

1. Silver: AI capability that scores at this level will include factors that have been more recently introduced into the product to become AI ready. This is the first stage where you discover what AI is, how it impacts business, the tools and technologies required to implement AI, how to prepare data for usage in AI, etc. This is a capability level that enhances the experience for users but is not mission-critical to the business outcome that enterprise users are seeking.

2. Gold: AI capability that scores at this level delivers a meaningful business outcome to the users. It will deliver a competitive edge for the AI offering or application in the market. The capability provides recommendations based on optimization or offline training on data and provides basic explanations for why a particular decision or recommendation was made. The AI features are usable by line-of-business users without having to involve data scientists. In this capability level, a good data hygiene and good automation of the engineering processes producing the capability is demonstrated.

3. Platinum: AI capability that scores at this level is a sustainable differentiator. It is part of a mission-critical workflow for the enterprise users and they rely on it for automated decision making and only focus on exceptions. The AI capability is sophisticated and has mechanisms to adapt to incoming data and learn by feedback provided by end-users. The decisions made by the AI are clearly explained and understood by business users. These users can adjust business level dials and levers to tune the outcome they desire. There are very strong and automated data management and data governance measures in place.

Not every AI capability needs to graduate from one phase to the next – organizations will need to assess the cost and benefit of investing in the capability based on users’ business outcomes and the cost to graduate them from silver to gold to platinum.

Figure 3: Maturity Framework Output – the Phases



Maturity framework: Detailed criteria

A. Impact on business

Sub Criteria	Definition	Silver	Gold	Platinum
A1. Business impact	<p>% of Revenue and/or % of Deal Wins influenced by this AI capability</p> <p>If Revenue or Deal Win info is not available, then:</p> <ul style="list-style-type: none"> – Key contributor in an analyst ranking (e.g., IDC, Gartner) – Key contributor in increased brand impression 	<p>Minimal Contribution</p> <ul style="list-style-type: none"> – Revenue Influenced: < 5% – Deal Wins Influenced: < 5% <p>Analyst Ranking, Brand impression: Not a factor</p>	<p>Significant Contribution</p> <ul style="list-style-type: none"> – Revenue Influenced: 5%-20% – Deal Wins Influenced: 5%-20% <p>Analyst Ranking, Brand Impression: A minor factor</p>	<p>Exceptional Contribution</p> <ul style="list-style-type: none"> – Revenue Influenced: > 20% – Deal Wins Influenced: > 20% <p>Analyst Ranking, Brand Impression: A major factor</p>
A2. Portfolio impact	<p>Drives adoption of the offering and broader portfolio.</p> <p>For example:</p> <ul style="list-style-type: none"> – Usage of the offering that could be measured in # of time users log in, # of licensed users or # of API hits – Increase in customer satisfaction measured by NPS scores 	<p>Low</p> <ul style="list-style-type: none"> – Usage: Increase of 1-5% – NPS: Increase of 1-3 points 	<p>Medium</p> <ul style="list-style-type: none"> – Usage: Increase of 5-20% – NPS: Increase of 4-6 points 	<p>High</p> <ul style="list-style-type: none"> – Usage: Increase of 20%+ – NPS: Increase of 7+ points

B. Value to client

Sub Criteria	Definition	Silver	Gold	Platinum
B1. Business process outcome	Defined and documented business outcome delivered by the AI capability. This is the measurable “business” outcome that would not have been delivered if AI was not infused. Every AI project must have clear metrics (short term, long term) by which to judge whether the project was a success (Metrics examples: NPS, Wins)	Business value or ROI for the AI project is unclear OR still being developed	Business value/ROI is documented but the metrics do not always tie-back to overall BU KPIs	Business value/ROI is clearly documented, and metrics are tied to overall BU KPIs. Both short term and long-term metrics are documented and tracked
B2. Differentiators	Clearly defined differentiator as a result of the AI capability. These differentiators allow clients to select our offerings over the competitors’	Clear and unique differentiators are still being assessed and documented	Differentiators are documented and give the offering competitive edge at present. However, the competition is catching up and working on a comparable capability	Differentiators are documented and sustainable. They give the offering a clear competitive edge today and near future. Enabling client to select our offering over the competition

C. Technology sophistication

Sub Criteria	Definition	Silver	Gold	Platinum
C1. Appropriateness of technology to the business problem	Are the tools and technologies used appropriate based on the business challenge to be solved (e.g., ML, NLP, NLU, RPU, ChatBot)?	Simple analytics with predictive capability generated from ML	Prescriptive analytics, making recommendations based on optimization or rules	Adaptive Learning and decision-making: Learning and/or decision-making processes that dynamically learn through feedback and adapt their strategies to new conditions
C2. Learning Techniques	Maturity of learning technique embedded and demonstrated by the AI model	Offline training for a specific task or a statistical technique to generate an insight	Supervised, offline or online training for multiple tasks utilizing regular feedback for learning	Reinforcement learning, Adversarial learning, techniques that learn how to learn with experience
C3. Re-use of Data Science models	How easy is it to re-use the DS model to allow for scaling and to gain operational efficiencies	Models are packaged into containers that must be manually modified and deployed for each new use case	Models are deployed as stand-alone, RESTful services that applications can easily call	Models deployed as services, composed from re-usable building blocks that can be re-arranged to form different workflows for different use cases
C4. Use of Inner Source or Open Source	How effectively does the technology make use of Inner Source and Open Source code to gain development efficiencies, build communities, mitigate ethical/risk concerns and encourage reproducible experiments?	Inner/Open Source are used for some components of the AI capability. Common examples would be data validation and transformation, feature engineering and visualization and reporting	Most AI capabilities were built on Inner/Open Source frameworks that have been packaged and shipped with the capability. Common examples would include model training, serving, monitoring and explainability. Experiments are internally reproducible and open to communities within IBM	All AI capabilities were built on and shipped with Inner/Open Source frameworks. Inner/Open Source libraries are regularly updated to the latest versions. Key functionalities and tests of the developed AI have been contributed back to the Inner/Open Source project

D. Trustworthiness

Sub Criteria	Definition	Silver	Gold	Platinum
D1. Integrity	How do we ensure data integrity throughout its lifecycle? Data Provenance and Data Lineage are known and documented. Understanding the history and origins of a data set as well as what happened to data after it was collected and prior to its use	Data dictionary and lineage are known at time of ingest	Data definitions and lineage for multiple projects are documented and used consistently. Developing a canonical data model	Data lineage is documented, tracked across transformations and published through the application
D2. Quality	How do we ensure data quality throughout its lifecycle? Data quality is measured to understand common issues and information content, what kinds of corrections are made and how improvements are measured.	Basic type validation and density-based outlier detection are reported	More advanced outlier detection, normalization, interpolation and standardized quality reporting	Automated detection and correction for a wide range of possible issues; Cross table and multivariate validations.
D3. Bias (Fairness) <small>For complete list on measuring Bias in an application/enterprise, please see the link given in the conclusion section.</small>	How do we reject bias towards groups, sets of individuals, or data attributes? Ability to prove that the outcomes are fair and not skewed either due to the model or the data (like incomplete, limited/insufficient, missing, corrupt, biased, ambiguous)	Basic bias reporting relying on manual intervention for remediation	Bias assessment and remediation are done for key features of data.	Bias assessment and remediation are done as a standard practice with proactive improvements in the approach and implementation of the techniques.
D4. Explainability	How can we explain, in business terms, the logic behind how the AI came to its conclusion? Outcomes can be explained and backtracked to the data and model that generated them. Explanations provided are tailored to stakeholders' needs.	Explainability is not available and AI solution is largely a "Black box". However, efforts are underway to make it available	Explanations for Inferencing and Decision Management are available on a limited scale.	Explainability is built into the solution and can be provided based on stakeholders' needs
D5. Security	How can we shield AI and AI infused services against cyber threats or adversarial attacks? Applications and algorithms are resistant to attacks from either data manipulation or direct security flaws	Risk and security management are minimal and reactive, mainly relying on key actors.	Policies, processes, and standards are defined and institutionalized for security and risk management at a consistent level across the enterprise and partners.	Risk and security management are comprehensive across the enterprise and among partners and customers, allowing for continuous feedback and improvement.

E. Ease of use

Sub Criteria	Definition	Silver	Gold	Platinum
E1. Intuitive for use by the intended user	How well can the AI capability be used by the intended end-user? Does it require experts (like Data Scientists) to be able to use and interpret the outcomes?	Basic AI features or capabilities that increase usability and deliver value and insights to end users. AI provides little or no Explainability, which can lead to low confidence levels. Users may have to rely on external assistance to interpret outcomes, provide additional insights and/or create custom solutions.	More mature AI capabilities that provide significant value and insights to users and helps them accomplish their goals more effectively. AI clearly conveys its value and reasoning, enabling users to build confidence in the insights delivered. AI delivers insights that increase user effectiveness, efficiency, and satisfaction.	Advanced AI capabilities that are tightly integrated into the experience and leverages Watson Moments that not only significantly improve the ease of use and efficiency of the product, but also enable users to be more productive and make smarter decisions faster. Users have a high degree of confidence in the AI and the unique insights delivered through the offering.

F. AI operating model

Sub Criteria	Definition	Silver	Gold	Platinum
F1. Deployment (manual, automated)	Uses tools that automate model building, data cleaning, and other key processes. Use of CI/CD pipeline	Manual build, integration, containerization and testing tools and processes are used for deployment	Use of continuous integration tools to build and test artifacts and containers for deployment with widely available status reporting	Deployment is automated using CI/CD and operational process is structured and enables dark launch in production/test environments
F2. Update frequency	How frequently can models be updated or retrained? Is the process automated?	No established plan for ongoing training of the AI model	AI model is trained at a regular basis based on human understanding of model quality	AI model update is automatic and is based on the automated assessment of model and data quality
F3. Infrastructure/ architecture scaling	Infrastructure is scoped, designed, and deployed to allow it to adequately scale with the solution	Consistent usage of containerization. Ad hoc usage of siloed technologies (chips, NVMe, composable computing, cloud computing etc.) critical to AI workloads; no defined architecture. No use of accelerated computing.	AI workload deployments are based on an architecture that is reviewed and modified occasionally. Architecture supports centralized containerization and workflow management. Limited use of accelerated computing.	AI workload deployments are based on flexible, centrally governed architecture with containerization and auto-scaling to adapt to ongoing user needs. Dynamically balancing use of accelerated computing and general-purpose computing for AI.
F4. Monitoring	Analyzing and evaluating the ongoing efficacy of the model (e.g. model drift analysis)	Basic logging of predictive accuracy and results	Reporting of how the model accuracy changes over time and the typical lifecycle	Alerting and automatic triggering remodel based upon continuous monitoring

G. Data

Sub Criteria	Definition	Silver	Gold	Platinum
G1. Data acquisition & instrumentation	How easy is it to acquire data and get it into the system for analysis?	Data acquisition and preparation are individuals' responsibility with no offering-wide guidelines.	Data acquisition procedures and platforms are in place; acquisition and preparation are documented for all data.	Data acquisition and data preparation are provided as a service that supports real-time provisioning of all needed resources including data sets, expertise, and tools.
G2. Data management	How robust is governance, provenance, and standardization of data throughout the lifecycle?	Rigid undefined processes with potentially unpredictable outcomes.	Agile processes are well defined, standardized and accepted to support early and continuous delivery of data to data scientists/engineers.	Agile processes are continuously improved by quantitative feedback based on IT and business metrics to support early and continuous delivery of data to data scientists/engineers.

Illustrative example

In thinking about how to apply a maturity model, examples are often very useful. AI Applications uses anomaly detection in a number of different applications in its portfolio. Here we describe a fictitious ecommerce application called Anomaly Finder as an example. We assume that the application was recently launched and has had limited marketing activity. At a high level, this application actively looks at a stream of key performance indicators (KPIs) for an ecommerce business over time, finds any anomalies relative to the history of the KPI, alerts the user to the anomaly, and then suggests a potential cause of the issue and a potential action to take if one exists for this type of issue. In this section, we will walk through how we might have evaluated this application in terms of its AI maturity.

The impact on the business criteria is meant to be an evaluation of an application's impact on both IBM revenue and on the brand of the product portfolio. Anomaly Finder was scored as silver for both business impact and portfolio impact. Since it was just launched it does not yet contribute a significant amount of revenue to the portfolio and while it has been mentioned by analysts in a couple of reports, it was not listed as a significant factor in the analyst's rating for the portfolio. It has also only had a limited impact in pulling through sales or increasing interest in other parts of the product portfolio. Overall, its impact on business area has not been significant enough thus far to rate it above silver.

The value to client criteria tries to measure how the application has created business value in terms of the client's defined business objectives for their AI journey for that specific business process. It also tries to estimate how distinct the value of the application for improving that business process is relative to competitive solutions from the client's perspective. One of the key components of Anomaly Finder is Risk Avoidance Reporting. This module directly estimates the benefit of following the recommended actions relative to what the outcome likely

would have been had no action been taken. The KPIs used are directly tied to the metrics that ecommerce companies use to evaluate their business performance. In its first few implementations, the solution has shown significant benefits to clients based on real-time tracking of these KPIs. Therefore, it is rated platinum for business process outcome. There are several features that clients and analysts have told us are novel in the market, but we believe that those moats could be challenged in the next few years. We rated Anomaly Finder gold in terms of its differentiators.

Technological sophistication tries to understand the maturity of the AI in the application based on the relevance of the techniques used relative to the business problem being solved, the sophistication of the learning methods used, and the ability to easily re-use or extend those methods used to enable continuous improvements in the AI in future releases. The appropriateness of technology criteria tries to understand the level of reasoning that was required for the application to effectively help the human in solving the business problem. A silver level provides the human with insights that they would not otherwise have had but relies on the person to make any decisions. Gold describes methods that enable the machine to make decisions based on learned insights and input from the human as to the goals and constraints of the business problem. Platinum applications apply learning to the decision process itself, modifying their reasoning in solving the problem based on learnings from previous actions they have taken. Anomaly Finder uses a proprietary set of algorithms to understand the behavior of other metrics at the time of the anomaly and uses its understanding of those behaviors to suggest potential causes of the anomaly and to recommend potential actions based on previous scenarios. It was rated gold for appropriateness of technology.

Learning techniques is an evaluation of the sophistication of the machine learning capabilities embedded in the application. Silver

typically implies offline, batch-based learning or a complex analytic that provides a single type of insight. Gold techniques are generally supervised, may utilize online training or train a model for multiple tasks based on feedback from prior actions. Platinum implies methods that can learn about how to improve themselves as a part of the learning process. They effectively learn how to learn as they gain experience. Anomaly Finder utilizes a random forest model that uses feedback from prior actions and thus was evaluated as gold.

The predictive and recommendation models for Anomaly Finder are containerized and are deployed as services. However, those services cannot be easily called by other applications. Therefore, the models have limited re-use and were evaluated as Silver for that criteria. They would need to be made widely available to become gold and would need to be broken up into composable services (e.g., data cleaning, feature engineering, model estimation, hyperparameter tuning, model scoring) to become platinum.

The predictive models in Anomaly Finder utilize both Inner Source and Open Source code. They use a set of time series prediction capabilities from an Inner Source library from IBM Research and several common Open Source packages like pandas, scikit-learn, PyOD and Seaborn in different parts of the AI pipeline. Both the Inner and Open Source packages are integrated into the Anomaly Finder containers along with test data so that experiments used to develop Anomaly Finder can be reproduced. The team has not yet contributed some of the key functionality back to the Inner Source library from Research. Anomaly Finder would be rated Gold for use of Inner and Open Source.

Trust is of paramount importance in AI applications. As maturity grows, users will generally need to cede more control over business processes to the application. They must believe that the application is making reasonable decisions that take into account situational context in order to allow this transfer of responsibility. Integrity is a measure of how trusted the data and transformations of data are in the application. Quality looks at how sophisticated the data and model validations are in ensuring that the application is learning appropriate, relevant behaviors. Bias measures how well the AI handles cases where biased data or algorithms might affect the recommendations that it makes. Explainability evaluates how well the application helps the user to understand why it made a particular decision. Security measures how resistant the application is to being tampered with or attacks from unauthorized parties.

The data used in Anomaly Finder conforms to a canonical data model and the sources of the data are documented. However, the application does not yet consistently track transformations or report on lineage as a part of the user application. We rated Anomaly Finder gold in terms of integrity. The Quality of the data is validated using machine learning techniques, some normalization is done and there is a published data quality report. Anomaly Finder, however, currently relies on manual intervention to correct several potential data and model issues. It was rated gold for Quality. The regression algorithms used in the application are tuned to reduce some types of bias, but they do not exhaustively evaluate bias in the data. Anomaly Finder was thus rated silver for bias. The application has an Anomaly Explainer module that uses the underlying data to help users understand the potential cause of

the anomaly and why a given recommendation was made. Since the explanations are only available for a limited set of solutions near the recommended one, we rated the application as gold. Anomaly Finder doesn't take special measures to secure its AI capabilities but relies on the broader application platform for security. This platform, however, has extensive, well defined and institutionalized security measures. We rated Anomaly Finder gold for Security.

Ease of use criteria focus on how well the AI demonstrates its value to the user in an intuitive manner, how easy it is for the user to understand the reasoning behind why a particular recommendation was made, and the extent to which it enables a user to accomplish their task without involving experts in IT or data science. All of the technical math and engineering have been abstracted away in Anomaly Finder so that a line of business user can use a simple set of sliders to define the relative importance of different business objectives they wish to have considered in the reasoning behind the recommendations. The explainer module helps the user visualize the data points that were critical in making the recommendation. They typically only need access to a data scientist or engineer in very complicated scenarios that may at first seem counterintuitive. Conversations with our users have revealed some areas where we believe that the AI still needs tighter integration with the overall experience for the application. Therefore, we rated Anomaly Finder gold for ease of use.

The AI operations model criteria attempt to evaluate the maturity of the processes required to successfully install and maintain the AI in a production or client environment. Deployment evaluates the maturity of the processes and level of automation for building the AI from source code, testing and deploying it. Update frequency describes the process for updating the training of AI models. Infrastructure scaling evaluates the AI's ability to scale up or down based on workloads to ensure performance and minimize operational costs. Monitoring evaluates an application's ability to give the user visibility into and manage the quality of the AI models over time.

For deployment, Anomaly Finder utilizes continuous integration tools for building and deployment but does not currently have dark launch or A/B testing capabilities. We rated it as gold for deployment. The predictive models are updated on a regular basis based on a scheduling algorithm rather than on an automated assessment of model quality, so the application was rated as gold for update frequency. The AI workloads for Anomaly Finder are based on an architecture designed specifically for their use. They are containerized and additional compute resources can be added to increase scale. It does not have the capacity to auto-scale based on current needs and thus was rated as gold for infrastructure/architecture scaling. Anomaly Finder has logging in both the prediction and recommendation models that captures both accuracy and results. The application does not currently keep track of these over time or have alerting that could trigger a remodel and thus was rated silver for monitoring.

Access and respect for data are essential for any AI application. The criterion around data are intended to capture the ease with which new data can be utilized by the application and the robustness of the processes and systems in place for managing the data once it has been acquired. The processes for data acquisition and ingest for

Anomaly Finder have been standardized, documented and to some extent automated. We have not yet created a standalone service that supports all of the major aspects of these processes. Therefore, we assessed that the application was gold in terms of data acquisition and instrumentation. Data management assesses the care taken with data in terms of understanding its provenance, governing its access and usage to respect data ownership and standardizing how data is delivered to the application for production usage and to data science for continued research and development. Anomaly Finder has a relatively static, standard set of processes that are used by our DevOps team to ensure that the data is treated respectfully and to enable continuous delivery of the necessary data to both production and research environments. Since the processes are static, we rated data management gold.

Based on this evaluation and in collaboration with Offering Management, we would have defined new development objectives that focus on making further investments in model re-use, monitoring and data quality as the next steps forward in improving the maturity of Anomaly Finder as an AI application.



Conclusion

As you apply AI to enterprise applications, establishing goals around the maturity of the AI is critical to ensuring that you deliver value to your enterprise clients, use the appropriate level of technological sophistication, ensure that the AI is trustworthy and easy to use by your targeted line-of-business users, and that you have an AI operating model in place to manage the AI you've deployed in the field along with strong data management and data governance practices. We hope that this maturity framework will help you measure and progress your enterprise applications on their AI journey.

Below you will find links to enterprise-grade data and AI management stack:

- <https://www.ibm.com/analytics>
- <https://www.ibm.com/blogs/internet-of-things/iot-energy-and-utility/>
- <https://www.ibm.com/blogs/internet-of-things/iot-space-evolution-with-data-and-ai/>
- <https://developer.ibm.com/technologies/artificial-intelligence/patterns/>
- <https://www.ibm.com/blogs/journey-to-ai/2019/10/idc-report-names-ibm-the-1-market-leader-in-ai/>
- <https://www.ibm.com/blogs/journey-to-ai/2020/07/announcing-the-journey-to-ai-blog/>
- <https://aifs360.mybluemix.net/introduction>
- <https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/>
- <https://www.research.ibm.com/artificial-intelligence/trusted-ai/>

Below you can learn more about IBM's Enterprise Applications infused with AI:

Domain	Enterprise Application	AI Capability (example)
Enterprise Asset Management	Maximo	Asset Performance Management
Integrated Workplace Management System	TRIRIGA	Watson Works
Supply Chain Management	Sterling	IBM Sterling Fulfillment Optimizer with Watson
Systems Engineering Software	Engineering Lifecycle Management	Requirements Management
Weather Solutions	Weather Business Solutions	Vegetation Management and Outage Prediction

References



1. Gartner, Artificial Intelligence Maturity Model, Published: 18 March 2020 ID: G00466009
2. IDC MaturityScape: Artificial Intelligence 1.0, #US44119919
3. Ovum, How to Achieve AI Maturity and Why It Matters
4. The AI Ladder: Accelerate Your Journey to AI, <https://www.amazon.com/AI-Ladder-Accelerate-Your-Journey/dp/1492073431/>
5. AI Maturity Assessment, <https://cloudcontent.mybluemix.net/cloud/architecture/assessments/ai-maturity-assessment>
6. Characterizing machine Learning Process: A maturity Framework, <https://arxiv.org/ftp/arxiv/papers/1811/1811.04871.pdf>

© Copyright IBM Corporation 2021
IBM Corporation
New Orchard Road
Armonk, NY 10504
Produced in the
United States of America
2021

IBM, the IBM logo, ibm.com, and Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

All client examples cited or described are presented as illustration of the manner in which some clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions. Contact IBM to see what we can do for you.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.