Empowering Your Organization with Responsible AI

Sponsored by: Microsoft

Author: Ritu Jyoti
June 2020

Introduction

Artificial intelligence (AI) offers transformational possibilities for customers, businesses, and society. Originally a discipline limited to academic circles, machine learning (ML) is now increasingly mainstream. From helping radiologists detect lung cancer to supporting decisions on whose bank loan is approved, from preventing fraud and real-time anomaly detection to predicting customer churn, and from providing customers with 24 x 7 self-service to empowering customer-services teams with AI-powered virtual agents, it is being used in more visible and impactful ways. But even as AI systems deliver value, their application could give rise to a host of unwanted outcomes and, sometimes, serious concerns. Artificial intelligence is often perceived as a black-box technology with the potential for unintended negative consequences, including some that are not yet known or experienced. Disastrous repercussions — including unintended consequences on health and medication if an AI medical algorithm goes wrong, exposure of personal and private information, and litigation exposures from potential racial bias — are possible. Businesses realize the potential of AI but are struggling in their implementations and in adapting their businesses to the multifaceted issues (e.g., legal, ethical, and sociological) of this transformational technology. Building trustworthy, responsible AI systems is fast becoming a business imperative.

Nearly every stage of the machine learning pipeline is vulnerable to biases as well as safety, security, and privacy issues that can cause a system to underserve users, disadvantage already-disadvantaged subpopulations, or even cause physical harm. Responsible AI needs to be approached as an end-to-end process — from new technology research and development (R&D) to production deployments to business practices on applying AI including data acquisition and understanding, modeling, deployment, and operations.

According to IDC's AI Strategies BuyerView Global Survey of 2,056 organizations conducted in March 2020, almost half of the respondents reported the lack of implementation of trustworthy AI to be a significant challenge for deploying AI technologies into production. In addition, the number of firms reporting AI as a risk factor in their annual shareholder reports (filed to the Securities and Exchange Commission in 2018) has more than doubled in 2018 according to a *Wall Street Journal* article.

Naturally, researchers and technology vendors want to help customers mitigate these risks. While rapid advancements are being made on all fronts, there are still gaps and challenges ahead.

In This White Paper

IDC defines Trustworthy AI as a framework to build trust in your AI solutions. It focuses on the development and use of AI solutions in a manner consistent with user expectations, organizational values and societal laws and norms. It also incorporates the right planning, oversight, and governance. Microsoft's equivalent framework is Responsible AI. Responsible ML capabilities in Azure Machine Learning support responsible development and use of machine learning models.

For both the practitioners involved in the machine learning life cycle and the governing bodies involved in improving their organizations' AI practices, this white paper shows how your organization should approach responsible ML and how you can invest early and appropriately to empower your organization with responsible AI. Lastly, IDC provides an analysis of Microsoft's principles for implementing responsible and trustworthy AI and ML and outlines the breadth of Microsoft's capabilities.

AI is a true competitive differentiator; it improves business agility and accelerates time to market with newer products and services.

AI Adoption Trends and Challenges

AI is a true competitive differentiator; it improves business agility and accelerates time to market with newer products and services. AI adoption is at a tipping point. According to IDC's *AI Strategies BuyerView Global Survey* conducted in March 2020, only 20% of all AI initiatives are currently in production. Cost of the AI solution, machine learning operations (MLOps), and lack of trustworthy AI are the top three challenges for deploying AI technologies into production (see Figure 1A).

Figure 1A

Top 3 Challenges for Deploying AI Technologies into Production

n = 2,056, Source: IDC's AI Strategies BuyerView Global Survey, March 2

As part of the scope of this document, let us now review IDC's perspective on the foundational elements of trustworthy AI (refer to figure 1B), understand how you as a practitioner involved in the ML life cycle or as a member of your organization's AI governing body can support the build and deployment of responsible AI, and how Microsoft can help you in this journey.

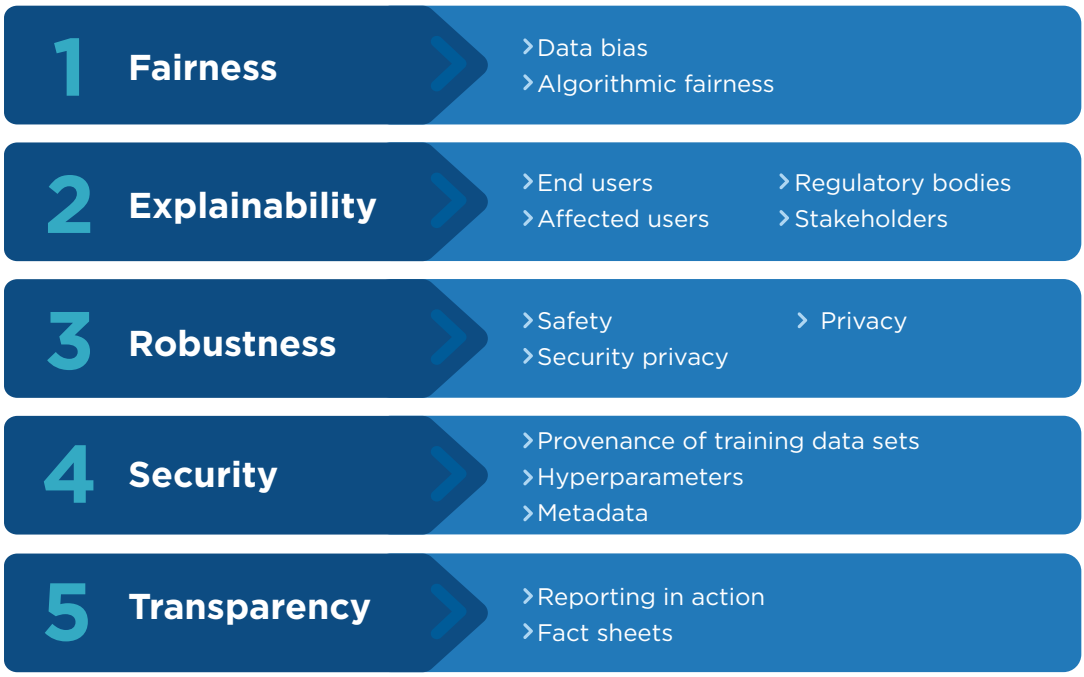
Trustworthy AI

With the creation of a technology that augments human beings in many high-stake decisions, the need for responsible AI cannot be underestimated. Such technology must be rooted in a common vision of human values and deliver results that positively benefit society, all while protecting individual privacy. Providers offering these powerful technologies must ensure that they are used for good and do not cause harm or have unintended effects (see Figure 1B).

With the creation of a technology that augments human beings in many high-stake decisions, the need for responsible AI cannot be underestimated.

Figure 1B

Foundational Elements of Trust



Source: IDC, 2020

Fairness

Recent advances in AI have made it smarter, faster, and, in many cases, more human-like. But despite these advances, AI can inherit our flaws. AI algorithms are not neutral. They are built by humans and used by humans and can therefore reflect societal biases. They can also behave unfairly for other reasons. Examples of AI systems behaving unfairly are found in image searches, hiring software, financial search, and so forth. Studies have shown, for example, that recidivism prediction tools can amplify racial bias in the criminal justice system. Analysis of an AI-based recruiting tool revealed that it was amplifying gender bias in the tech industry by withholding employment opportunities from women.

Explainability

If AI systems are obscure and unable to explain how or why certain results are presented, the lack of transparency will undermine trust in the system and in any results. Understanding how AI models arrive at specific decisions is a key principle of trusted AI. Different stakeholders require explanations for different purposes and with different objectives, and explanations will have to be tailored to their needs. In other words, one explanation does not fit all. For example, an end-user (physician) might need insights and justification on why a treatment was recommended, an affected user (loan applicant) would need understanding of the factors that led to his/her loan denial and what he/she could do to get it approved, and a regulatory agency (the Federal Trade Commission (FTC) or Equal Employment Opportunity Commission (EEOC), might need proof that the system did not discriminate to ensure fairness for its constituents.

Understanding how AI models arrive at specific decisions is a key principle of trusted AI.

Robustness

AI systems should be safe and secure, not vulnerable to tampering or to compromising the data they are trained on. AI robustness is determined by three underlying factors: safety, security, and privacy. An AI system can be fair and explainable but still unsafe to use. AI safety is typically associated with the ability of an AI model to build knowledge that incorporates societal norms, policies, or regulations that correspond to well-established safe behaviors. A model must work well in all scenarios that it is confronted with — many models work well on average but have unsafe or unpredictable behavior at the edges or in corner cases. AI system designs, like AI models, must also be robust to withstand model errors. Increasing the safety of AI models is a key element of a trusted AI system.

Deploying autonomous AI systems (e.g., self-driving cars or robotics) across public or industrial arenas could pose a risk or harm. Trust in systems that both support and offload current work tasks poses the risk of ignoring our knowledge of those skills. This will make it more difficult to judge the correctness and outcome of these systems and, in the end, make it impossible for human interception. We need to ensure a human is in the loop in the running of these AI systems. We also need to ensure robust testing for all potential scenarios and system designs that are robust to withstand model errors.

In addition to traditional software security and privacy concerns, AI adds new types of attacks that need to be considered such as data poisoning or adversarial AI methods. The accuracy of AI models is directly correlated to the vulnerability of the input data set. That relationship may be exploited by malicious actors that can try to alter specific data sets and/or influence the behavior of an AI model. Testing and benchmarking AI models

Testing and benchmarking AI models against adversarial attacks are key to establishing trust in AI systems.

against adversarial attacks are key to establishing trust in AI systems. Access to vast amounts of data will enable AI systems to identify patterns beyond human capabilities. In this there is a risk that the privacy of individuals could be breached. AI practitioners need to secure and govern the use of data derived from human activities online or in real life. Recent advances require that we pay even closer attention to these issues to create the levels of trust needed to realize the full benefits of AI. Simply put, people will not share data about themselves — data that is essential for AI to help inform decisions about people — unless they are confident that their privacy is protected, and their data secured.

Lineage

AI systems should include details of their development, deployment, and maintenance so they can be audited throughout their life cycle. AI models are constantly evolving, making it challenging to trace their history. Establishing and tracking the provenance of training data sets, hyperparameter configurations, and other metadata artifacts over time are important to establish the lineage of an AI model. Understanding the lineage of AI models helps establish trust from a historical perspective that is difficult to achieve by just factoring fairness, explainability, and robustness alone.

Transparency

The subject of disclosures and transparency in AI systems is a nascent area of research, but one that is key to the mainstream adoption of AI. Just as people use information sheets for hardware appliances or nutrition labels on foods, AI practitioners should consider establishing a fact sheet for AI models, which should answer basic questions about AI models such as whether a data sheet is available for the data set used to train the service, if the data set was checked for biases, if there was any bias mitigation performed on the data set, and if algorithmic outputs are explainable/interpretable.

How Organizations Can Build and Deploy AI Responsibly

Incorporating Governance

An AI platform that is built on a foundation of trusted and responsible components must be coupled with governing business processes that ensure the application of

core ethics and regulations. Clear policies with respect to data privacy, decision rights, and transparency should be established that ensure compliance with relevant laws and regulations. Accountability policies need to be put in place, not only to determine who is held responsible when AI system outputs go wrong but also to reinforce the magnitude of the responsibility in implementing such powerful technologies. Governance processes must oversee the validation, monitoring, and analysis of AI models and results. In the initial stage of deployment, human knowledge and expertise should complement machine-based decisions to enable trust and improve the sophistication of algorithms.

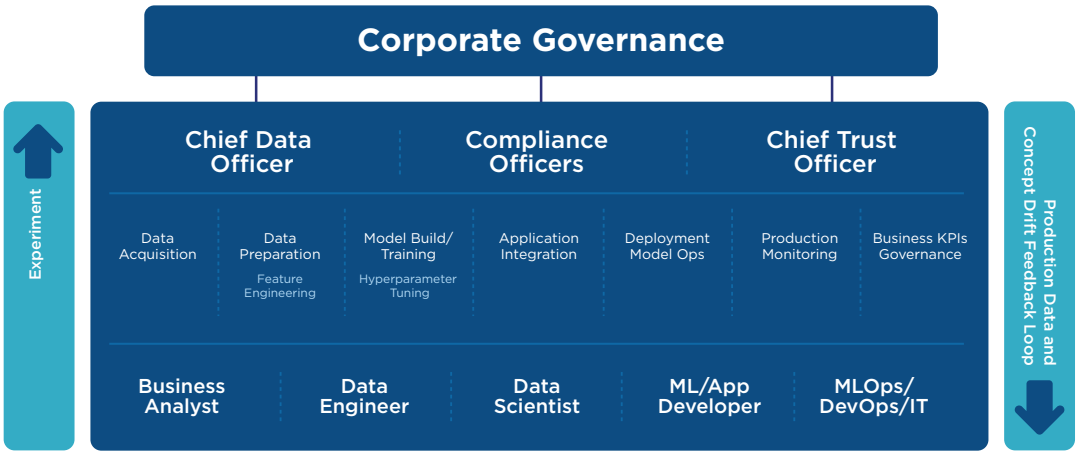
Every persona involved in the machine learning life cycle, from experimentation to production, and the corporate governance team has a crucial role in supporting the build and deployment of responsible AI solutions.

Organization-Level Responsibility

Every persona involved in the machine learning life cycle, from experimentation to production, and the corporate governance team has a crucial role in supporting the build and deployment of responsible AI solutions (see Figure 2). Company leaders, and ultimately the entire workforce, should be educated on the societal, legal, and ethical impacts of working alongside AI. Organizations should be prepared to make algorithms, attributes, and correlations open to inspection so that participants can understand how their data is being used and how decisions are made.

Figure 2

User Personas in Machine Learning Life Cycle and Corporate Governance



Source: IDC, 2020

Data teams should strive for diversity: to have a representation of the population where the algorithms would be deployed to ensure that diverse training data and more thoughtful feature sets are employed, leading to less bias in the data.

Corporate governance teams (chief trust or data officer or the compliance professionals) need to provide ethical oversight for AI applications, whether developing or applying AI. Clear policies with respect to data privacy, decision rights, and transparency should be established that ensure compliance with relevant laws and regulations. They or their teams must create centralized templates and governance processes. Governance processes must oversee the validation, monitoring, and analysis of AI models and results. In the initial stage of deployment, human knowledge and expertise should complement machine-based decisions to enable trust and improve the sophistication of algorithms. AI applications must include internal and external checks to ensure equitable application across all participants. Governance teams must develop communications practices to explain AI-related decisions. As with human intelligence, AI models begin with a trigger event, apply reasoning, and produce a decision; oversight should be applied to all of these stages.

Data teams should strive for diversity: to have a representation of the population where the algorithms would be deployed to ensure that diverse training data and more thoughtful feature sets are employed, leading to less bias in the data. Establishing tests for identifying and minimizing bias in training data sets should be a key element in establishing fairness in AI systems, especially for AI apps with a tangible social impact such as credit or legal applications. AI application developers, business analysts (aka citizen data scientists), and data scientists must work together to ensure that they test for possible scenarios and mitigate risks.

There is a natural trade-off between the explainability of AI models and their accuracy. Highly explainable AI models tend to be very simple and therefore not incredibly accurate. From that perspective, establishing the right balance between explainability and accuracy is essential to improving the trust in an AI model. Data scientists and business functions must collaborate on establishing the right balance.

The ultimate responsibility is to ensure transparency to end users and support an organization's auditing requirements for ongoing performance monitoring and compliance. The onus to describe mitigation methods, testing methodologies, how the service was checked for robustness against adversarial attacks, or how and where usage data from service operations is retained, stored, and kept is shared between ML engineers or AI application developers and testers and the business users.

MLOps or DevOps engineers should ensure robustness: the safety and security of the platforms used for building and deploying the AI solutions.

Considering Microsoft Azure Machine Learning for Building, Deploying, and Implementing Responsible AI Solutions

Microsoft is committed to making sure AI systems are developed responsibly and in ways that warrant people's trust.

Responsible AI at Microsoft

Microsoft is at the forefront of AI innovations and is optimistic about what these technologies can do for our lives and businesses, now and in the future. Microsoft is committed to making sure AI systems are developed responsibly and in ways that warrant people's trust. In 2016, Microsoft took an active role in the industry in creating a set of human-centered principles to guide the creation and use of AI. It has evolved these into six principles: fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. These guide Microsoft's end-to-end approach to AI, from its development to its deployment. Today, Microsoft is putting its principles into practice by embracing diverse perspectives, fostering continuous learning, and proactively responding as AI technology evolves.

The company is operationalizing responsible AI at scale across Microsoft, accomplished through Microsoft's AI, Ethics, and Effects in Engineering and Research (AETHER) Committee and the company's Office of Responsible AI:

- AETHER is tasked with advising Microsoft's leadership around rising questions, challenges, and opportunities brought forth in the development and fielding of AI innovations.
- The Office of Responsible AI puts Microsoft principles into practice by setting the company-wide rules for responsible AI through implementing our governance and public policy work.

Together, AETHER and the Office of Responsible AI work closely with Microsoft engineering and sales teams to help them uphold Microsoft's AI principles in their day-to-day work. An important hallmark of Microsoft's approach to responsible AI is an ecosystem that operationalizes responsible AI across the company, rather than having a single organization or individual leading this work. Microsoft looks at responsible AI as an end-to-end process, and its approach to responsible AI leverages the process of building privacy and security into all of its products and services from the start.

Core Capabilities: Technical Overview

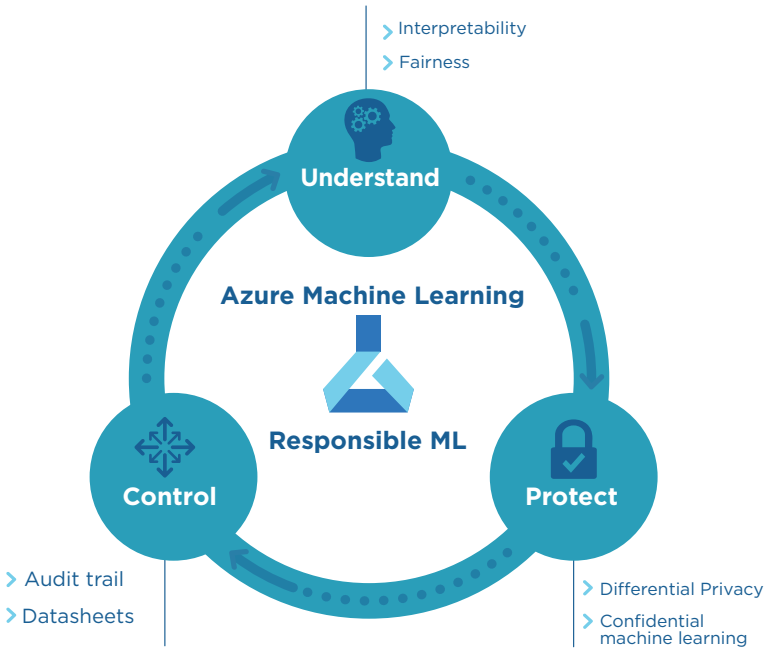
Designing AI to be trustworthy requires creating solutions that reflect principles that are deeply rooted in important and timeless values. Figure 3 provides an overview of such responsible capabilities available in Azure Machine Learning. IDC examined each of these areas and offers example scenarios.

Designing AI to be trustworthy requires creating solutions that reflect principles that are deeply rooted in important and timeless values.

Figure 3

Responsible AI in Azure Machine Learning

Source: Microsoft, 2020



Understand

This first pillar of the responsible capabilities in Azure Machine Learning focuses on helping data scientists and business analysts (aka citizen data scientists) gain deeper understanding of their models through automated analytics and insights, support the explainability requirements of different stakeholders including the governing body, and build models for fairness. Microsoft offers two sets of capabilities (interpretability and fairness) in Azure Machine Learning to support this objective. To understand how these capabilities can help, let us consider this scenario.

Scenario: You are a data scientist or a business analyst at a bank, building a predictive model for whether to provide a small business with a credit line or a loan. You may only have on hand the business name and a history of its bank transactions. In addition to managing the bank's financial risk, model results must also satisfy regulatory oversight.

Interpretability

Why You Need It

For this scenario, based on availability of the existing data, you can have access to features such as transaction slope per month, largest payment per quarter, and standard deviation of revenue per week. Additional data sources from outside the bank would open a world of potential features that may improve the predictive model — for example, foot traffic figures or the company's publicly available online activity such as number of tweets and date of next renewal for the website. All of these could provide meaningful new features that could impact the predictive model. In this case, you can see that the ML model is now getting complex, with lots of parameters, and you may not have visibility into how the models "think" to explain the outcome for regulatory oversight.

During inferencing, interpretability capabilities could be used by the governing body to support the explainability requirements of regulators.

How It Works

Explanations surfaced through Azure Machine Learning can enable better understanding of your model behavior by highlighting the data features that have the most influence on your model predictions. Interpretability techniques like EBM, LIME, SHAP, and global surrogate models are used to provide model understanding and can be accessed through interactive visualizations. They empower you to dive deeper into the model by reviewing the feature importance for the whole data set and smaller subsets, down to individual data points. You can also further explore how the model will behave with the interactive what-if analysis showing how modified data features lead to different predictions. With these generated model explanations, you can validate the rationale, identify errors, and retrain the model. You can easily access this in Azure Machine Learning.

While building and training the model, let us say you want to assess if using foot traffic figures can help you improve the accuracy of this predictive model for loan processing. You could use Azure Machine Learning what-if analysis capability to validate the rationale and retrain if needed.

For straightforward models, you can select from a list of state-of-the-art interpretable yet performant glassbox models (e.g., EBM) to use during the ML training phase. For any pretrained models, Azure Machine Learning provides a rich set of black-box or gray-box explainers that support any classic ML models (e.g., random forest and SVM) or deep learning models (trained via PyTorch, TensorFlow, and Keras) without making changes to the model training life cycle.

Interpretability capabilities are integrated with different stages of Azure Machine Learning (training and deployment stages) and can be both used at training time and deployed alongside the actual model to provide explanations at inferencing in real time. During inferencing, interpretability capabilities could be used by the governing body to support the explainability requirements of regulators.

The interactive visualization via a simple API call in-notebook widget allows you to gain a deeper understanding of the model by focusing on fairness metrics and to slice and dice the data for useful insights.

Let us say that your model denies the approval of the loan for a small business - whose owner is from a protected class. You could use the Azure Machine Learning interpretability capabilities at inferencing time to understand which feature 'transaction slope per month' or 'the largest payment per quarter' or an alternate group of features influenced the decision. You can use this understanding to explain to the regulatory body that your organization did not discriminate against the small business owner protected class or group.

Fairness

Why You Need It

ML models may behave unfairly by negatively impacting groups of people, such as those defined in terms of race, gender, or age. You can use the fairness capabilities inside Azure Machine Learning to assess whether a ML model is behaving unfairly, to compare multiple models in terms of their fairness and their performance, and to mitigate unfairness.

How It Works

Fairlearn supports a wide range of fairness metrics for assessing a model's impacts on different groups of people, covering both classification and regression tasks. The fairness metrics and interactive visualization dashboard can help with assessing which groups of people might be negatively impacted by a model, while the unfairness mitigation algorithms can help with mitigating unfairness in classification and regression models. The interactive visualization dashboard, which is an in-notebook widget, generates visualizations of a model's impacts on groups defined in terms of a sensitive feature (e.g., "sex" or "age"). The dashboard also allows you to compare the fairness and performance of multiple models, enabling you to navigate trade-offs and find a model that fits your needs. The unfairness mitigation algorithms can help you to improve the fairness of your ML models. There are two types of mitigation algorithms, drawing on work by Microsoft Research and third parties: post-processing algorithms and reduction algorithms. Both operate as "wrappers" around any standard classification or regression algorithm.

Let us build on the same example, where your model denies the approval of the loan for a small business — whose owner is over 60 years of age. You could use the Azure Machine Learning fairness capabilities to assess the age of the owner is not impacting the decision. You can also use the fairness capabilities to compare use of various models and associated impact, along with an interactive visualization dashboard to simplify the effort.

Protect

This second pillar of the responsible capabilities in Azure Machine Learning is focused on helping the data teams (data scientists, data architects, data engineers, and citizen

data teams) along with the operations teams (MLOps, DevOps, and IT personas) protect user privacy and guarantee confidentiality in AI applications. Microsoft offers two sets of capabilities (differential privacy and confidential machine learning) in Azure Machine Learning to support this objective. To understand how these capabilities can help, let us consider this scenario.

Scenario: Your organization is building a speech recognition solution that will be used in a regulated industry such as healthcare or financial services. For this scenario, you need to ensure that the platform that is used to develop and deploy the models is protected for different parties to share data privately and also maintains extra levels of confidentiality for personally identifiable information (PII) needs.

Differential Privacy

The differential privacy capability in Azure Machine Learning is integrated with Azure data sources and enables data teams to protect sensitive data against reidentification and record recovery attacks.

Why You Need It

Differential privacy will enable data teams to produce reports and notebooks that extract key information from the data and produce dashboards and analyses that extract key insights. Privacy exposure tracking tools, for either individuals or composition of queries, will help business analysts understand the amount of privacy exposure that correlates with the types and numbers of queries and how much noise has been injected into the insights. It will reduce the risk of privacy intrusions of AI systems so applications can use personal data without accessing or knowing the identities of individuals.

How It Works

The differential privacy capability in Azure Machine Learning is integrated with Azure data sources and enables data teams to protect sensitive data against reidentification and record recovery attacks.

Differential privacy has a wide range of support for different data sources (e.g., csv files, SQL server, and PostgreSQL), programming languages (e.g., SQL, R, and Python), and most popular noise-adding algorithms (e.g., Laplace, Gaussian, Geometric, and Vector). Business analysts can use the most popular statistical queries (e.g., sum, count, and mean) to derive insights from private data. While other techniques such as anonymization, aggregation, and encryption do not protect published reports against reidentification and record recovery attacks, differential privacy provides strong assurances against these privacy attacks against published reports (dashboard, Jupyter notebook, and ML models). Privacy is provided to every individual by injecting a precise level of noise, calibrated on the types and the number of queries accessing private data.

So, in this scenario, let us say you are using a SQL server table as one of your data sources. It consists of voice samples along with PII information such as name, birthday,

and credit card information, which needs to be protected for privacy. Using WhiteNoise differential privacy capabilities with Microsoft Azure, you could easily access the rows in this table without knowing the identities of the individuals.

Confidential Machine Learning

Azure Machine Learning already provides a strong set of data and networking protection capabilities such as Virtual Networks for compute and storage isolation and private link to connect to machine learning workspaces without sending data over the Internet.

Why You Need It

Confidential machine learning will encrypt data and models to help maintain confidentiality in a secure environment.

How It Works

Note: Microsoft will be bringing these confidential machine learning capabilities to developers and data scientists later this year.

Azure Machine Learning already provides a strong set of data and networking protection capabilities such as Virtual Networks for compute and storage isolation and private link to connect to machine learning workspaces without sending data over the Internet. It also supports dedicated compute hosts, and customer managed keys for encryption in transit and rest using Azure Key Vault, TLS/SSL and more.

Building on this foundation, Azure Machine Learning will enable internal Microsoft teams to build models over confidential data, without requiring access to or being able to see the data itself. The secure machine learning environments to build these models will not be accessible to the data scientists. All the ML assets, including trained models, logs, and metrics resulting from the confidential data, will be kept confidential. This approach is fully compatible with open source machine learning frameworks like TensorFlow, PyTorch and can support a wide variety of hardware options, including the latest NVIDIA GPUs, for distributed deep learning.

Microsoft's confidential machine learning capability will help protect data across the entire data management life cycle by using techniques like encryption and the ability to train models without seeing the confidential data, in a secure environment

So, in this scenario, let us build on the same example of using a SQL server table as one of your data sources. It consists of voice samples along with PII information such as name, birthday, and credit card information, which needs to be protected for privacy. Using Microsoft Azure confidential machine learning along with differential privacy, you can build the models in a secure environment and store all the ML assets throughout the life cycle without compromising the confidentiality requirements.

Control

This third pillar of responsible capabilities in Azure Machine Learning is focused on helping organizations, specifically the governing body, ensure accountability by supporting control and governance throughout the ML life cycle. Artificial intelligence is largely seen as a commercial tool, but it is quickly becoming an dilemma in certain scenarios. In the field of healthcare, there are serious legal issues that need attention. Who is to blame if a smart algorithm makes a mistake and does not spot a cancerous nodule on a lung x-ray? To whom could someone turn when AI comes up with a false prediction? There are serious repercussions and associated legal risks in the absence of governance.

Microsoft is subscribing to the adoption of best practices for the establishment of accountability norms, periodic checks, and internal review boards/ governing bodies to provide oversight and guidance on which practices should be adopted to help address the concerns discussed previously and on particularly important questions regarding development and deployment of AI systems.

Microsoft is subscribing to the adoption of best practices for the establishment of accountability norms, periodic checks, and internal review boards/governing bodies to provide oversight and guidance on which practices should be adopted to help address the concerns discussed previously and on particularly important questions regarding development and deployment of AI systems. Data teams should be the moral conscience officers of the organization; Microsoft offers two sets of capabilities (datasheets and audit trail) in Azure Machine Learning to support this objective.

To understand how these capabilities can help, let us consider this scenario.

Scenario: You (data scientists/business analysts) are developing a machine learning model that your organization could use to identify the best-qualified candidates from among thousands of job applicants for a job opening. In certain situations, you may be required to present the details that drove that decision to an external party.

Audit Trail

Why You Need It

This will help track model version history and lineage for auditability and regulatory requirements and achieve governance, compliance, and control across machine learning assets. IT departments normally provide audit logs and trails to analyze operations and validate and monitor activity. In machine learning, ML engineers or the Ops teams often perform this function. This will help collect system information and maintain integrity of operations.

How It Works

The Azure Machine Learning service model registry automatically tracks datasets, models and their version history, model explanations along with the lineage and artifacts of the model. Azure Machine Learning gives you the capability to track the end-to-end audit trail of all your ML assets by using metadata. You can build audit trails as you tag machine learning assets and automatically track experiments.

You can access real-time data to diagnose errors and warnings on training models, compute targets, images and more. Security and IT teams can also access this data to increase the level of trust with the process.

Azure Machine Learning built-in integration with Azure Monitor and Azure Event Grid helps create alerts when certain events occur such as completion of training runs, model deployment, data drift detection, and more. This provides you (data scientists, developers, and ML engineers) with visibility into the ML lifecycle and the ability to react in an auditable manner.

You can access real-time data to diagnose errors and warnings on training models, compute targets, images and more. Security and IT teams can also access this data to increase the level of trust with the process.

So, in this scenario, let us say for example the employer is a government agency and for a candidate who was not hired you need to provide details that led to that decision. Using Microsoft Azure Machine Learning Audit trail capability, you could easily use the audit trails to satisfy the ask including model explanations.

Datasheets

Why You Need It

Documentation of datasets, models, and other machine learning assets is necessary for responsible AI, increasing transparency and accountability and improving collaboration. Datasheets provide a standardized way to document ML assets, capturing things like motivations, creation processes, intended uses, and ongoing maintenance plans. Microsoft initiated research on datasheets for datasets, including a set of questions and workflow to elicit the information that such datasheets should contain. They are also working with the 'Partnership on AI' and leaders across industry, academia, and government to develop recommended documentation practices for ML assets as part of a project called ABOUT ML.

Improved documentation increases transparency, collaboration, and accountability by helping practitioners uncover implicit assumptions and select appropriate ML assets for their chosen tasks and allowing organization policy makers to enforce responsible creation and use of ML assets.

Artificial intelligence is becoming more ubiquitous and necessary these days. It is poised to transform our lives, businesses, and society.

How it Works

Azure Machine Learning provides custom tags to help implement datasheets for ML models today, and over time will be releasing additional features to make it easier to adopt this practice and to create datasheets for other ML assets. Datasheets are stored alongside ML models to increase searchability, traceability, and accountability. They can help practitioners to create and select appropriate ML models for their chosen tasks.

So, in this scenario, let us build on the same example where the employer is a government agency and for a candidate who was not hired, you need to provide details that led to that decision. Using Microsoft Azure Machine Learning datasheets that support a standard set of documentation along with the model in use, you can easily search through them and provide details on the datasets that were used to train the model.

Challenges and Opportunities For Microsoft

Combining the enthusiasm for AI, the early disappointments, and the growing awareness of the importance of a responsible AI platform, businesses are seeking practical guidance, platforms, and tools for building and deploying effective and responsible AI. Microsoft could offer support such as:

- Continuing to build out industry-specific courses for responsible AI through Microsoft AI Business School.
- Continuing to invest in research and work with governments, academia, and others in the industry to develop effective and efficient tools and technologies to support the build and deployment of responsible AI.
- Accelerating the support for better accountability services.
- Exploring and expanding support of responsible AI tools and technologies to third-party offerings.
- Continuing to simplify the ease of use and deployment of responsible AI tools and technologies.

Conclusion

Artificial intelligence (AI) software platforms provide the functionality to analyze, organize, access, and provide advisory services based on a range of structured and unstructured information.

Artificial intelligence is becoming more ubiquitous and necessary these days. It is poised to transform our lives, businesses, and society. While AI is everywhere, businesses are still awaiting the realization of AI/ML solutions at scale and associated superior business outcomes ranging from accelerated innovation to improved customer experience and improved employee productivity.

As per IDC research, enabling and ensuring build and deployment of trustworthy AI solutions is one of the top three challenges for deploying AI solutions into production. While rapid advancements are being made in terms of tools and technologies to support build and deployment of fair, explainable, safe, secure, private, transparent, and accountable AI solutions, there are still gaps and challenges ahead. IDC advises end users to exercise due diligence and adhere to a corporate AI governance framework to safeguard against unintended or negative consequences.

By turning to responsible ML capabilities in Azure Machine Learning like fairness, interpretability, differential privacy, confidential computing for ML, datasheets, and audit trail, lineage, and quota management, practitioners involved in the ML life cycle and corporate governance body can support the build and deployment of responsible ML solutions.

Message from the Sponsor

To learn more about Microsoft's approach and offerings, select one of the options:

- Get started with Responsible ML, visit Azure Machine Learning Overview.
- Visit Microsoft Responsible AI to understand Microsoft's approach and ethical principles.
- Visit Microsoft Azure AI Innovation to learn more about Microsoft's AI and ML offerings.
- Visit Microsoft Responsible AI resources that brings together all practices and tools.

Appendix

Definitions

Artificial intelligence (AI) software platforms provide the functionality to analyze, organize, access, and provide advisory services based on a range of structured and unstructured information. These platforms facilitate the development of intelligent, advisory, and AI

These platforms facilitate the development of intelligent, advisory, and AI applications, including intelligent assistants that may mimic human cognitive abilities.

applications, including intelligent assistants that may mimic human cognitive abilities. The technology components of AI software platforms include text analytics, rich media analytics (such as audio, video, and image), tagging, searching, machine learning (ML), categorization, clustering, hypothesis generation, question answering, visualization, filtering, alerting, and navigation. These platforms typically include knowledge representation tools such as knowledge graphs, triple stores, or other types of NoSQL data stores.

These platforms also provide for knowledge curation and continuous automatic learning based on tracking past experiences. When these individual technology components are sold standalone, they are accounted for in other software functional markets such as content analytics and search, advanced and predictive analytics, and nonrelational database management systems (NDBMSs).

AI software technologies are a set of technologies that use natural language processing (NLP), image/video analytics, machine learning, knowledge graphs, and other technologies to answer questions, discover insights, and provide recommendations. These systems hypothesize and formulate possible answers based on available evidence, can be trained through the ingestion of vast amounts of content, and adapt and learn from their mistakes and failures through retraining or human supervision.

Learn More

Related Research

- IDC's Worldwide Artificial Intelligence Spending Guide Taxonomy, 1H19 (IDC #US46107920, March 2020)
- Mitigate AI/ML Risks with Trustworthy, Ethical, and Governed AI (IDC #US45792719, January 2020)
- IDC MarketScape: Worldwide General-Purpose Artificial Intelligence Software Platforms 2019 Vendor Assessment (IDC #US43065418, December 2019)
- Worldwide Artificial Intelligence Forecast, 2019–2023 (IDC #US45332319, July 2019)
- Artificial Intelligence Global Adoption Trends and Strategies (IDC #US45120919, June 2019)
- Worldwide Artificial Intelligence Taxonomy, 2019 (IDC #US45013419, April 2019)
- Ethics Considerations in Artificial Intelligence (IDC #US44587118, January 2019)
- Explainable Artificial Intelligence: Feasible, Plausible, or Just a Pipe Dream (IDC #US44587318, January 2019)
- Making PaaS Smarter: The Role of Cognitive and AI Services (IDC #US44512118, December 2018)

IDC Global Headquarters

5 Speen Street,
Framingham, MA 01701,
USA

508.872.8200

Twitter: @IDC

idc-community.com

www.idc.com

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.