Cool Vendors in AI Governance and Responsible AI

Published 10 June 2021 - ID G00746831 - 18 min read

By Analyst(s): Svetlana Sicular, Moutusi Sau, Sumit Agarwal, Arun Chandrasekaran, Farhan Choudhary, Shubhangi Vashisth, Avivah Litan

Initiatives: Artificial Intelligence

Adoption of artificial intelligence is pushing organizations from declarations of their responsible and ethical AI principles to actions. Data and analytics leaders can already find the tools to assure AI fairness, bias mitigation, explainability, privacy and compliance.

Additional Perspectives

- Summary Translation + Localization: Cool Vendors in AI Governance and Responsible AI
  (09 July 2021)

Overview

Key Findings

- Responsible AI and AI governance are the disciplines that require supporting tools to assure trust, fairness and explainability. The tools market is fragmented, but it is evolving very quickly. Many capabilities are already available.
- Synthesized and Hazy offer synthetic data as the means of fairness, privacy, confidentiality and bias mitigation.
- TruEra provides explainable AI to address bias, model quality and assess risk exposure.
- DreamQuark delivers a fairness engine and explainability for the financial services industry.
- Arthur automates AI monitoring to ensure responsible AI outcomes.
- Tools, no matter how good they are, are never sufficient to drive trust, ensure fairness or mitigate bias. The discipline of AI governance is necessary to address the problems that tools cannot: Contextualize the responsible AI definitions, assess organizational risks and put humans in the loop to add common sense.

Recommendations

Data and analytics leaders should:

- Promote responsible development of AI. It should be spearheaded by an AI governance committee that provides standards, guidelines and interventions for responsible AI testing and implementations.
- Create a function of a neutral AI model validator, a member of the data science team who independently validates bias mitigation and assures model fairness, compliance and explainability for the models he or she did not develop. This is the role that requires responsible AI tools.
- Examine how synthetically generated data could be used in your AI development to mitigate bias, improve fairness and assure regulatory compliance by providing the necessary information within legal and ethical constraints.
- Implement disciplined approaches and tools for transparency and explainability of AI-enabled decision making to minimize misinterpretations of AI results.

Strategic Planning Assumptions

By 2023, all personnel hired for AI development and training work will have to demonstrate expertise in responsible development of AI.

By 2024, 60% of the data used for the development of AI and analytics solutions will be synthetically generated.

By 2025, the concentration of pretrained AI models among 1% of AI vendors will make responsible AI a societal concern.

Through 2025, 80% of organizations seeking to scale digital business will fail because they do not take a modern approach to data and analytics governance.

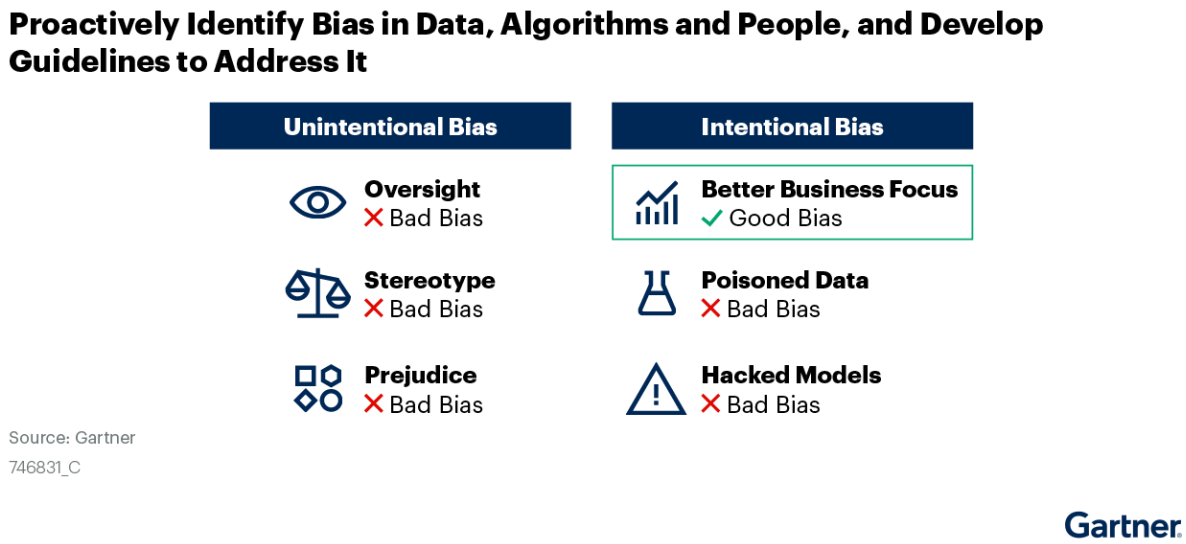
Analysis

This research does not constitute an exhaustive list of vendors in any given technology area, but rather is designed to highlight interesting, new and innovative vendors, products and services. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

What You Need to Know

Responsible artificial intelligence (AI) is in the spotlight because of the increased organizational and societal exposure to AI bias, distrust and lack of fairness that stand in the way of AI adoption. The more ubiquitous AI becomes, the more AI practitioners recognize in their everyday work that data and algorithms amplify and perpetuate human biases. Some biases are accidentally lurking, and others are intentionally introduced in the outputs (see Figure 1). Not all intentional biases are bad: Chatbot politeness is a good bias.

Figure 1: Proactively Identify Bias in Data, Algorithms and People, and Develop Guidelines to Address It



Obvious biases, like racism and sexism, are straightforward to think of. But, they are hard to mitigate because algorithms can deduce race and gender from proxy parameters, such as typical female names or postal codes with the dominant racial demographics. More implicit bias is difficult to spot: It includes a long history of cultural norms, beliefs and rules established as people interacted with each other. For example, a data scientist might overlook that a number of clicks on the website can be discriminatory against age. AI can perfectly classify a stereotypical western wedding but be blind to the weddings in India and Africa.

The questions about responsible AI and AI governance are among the top five topics of Gartner clients' inquiries about artificial intelligence in 2021. Gartner clients ask, if bias is introduced, how can they intervene and create transparency? How can they assess if the data science output is fair, and, if so, how can they prove it? Is there a tool to help automate the ability to intervene when the model introduces bias in its lifetime? What tools are available to explain a model to increase adoption?

This Cool Vendors report showcases cool vendors that give you tools to deal with the topics of responsible AI and AI governance:

- Bias mitigation and fairness
- Explainability
- Trust and transparency

- Privacy and regulatory compliance

Many things in AI are done for the first time ever. Mistakes will happen. The cool vendors in this report help address how to fix mistakes with minimal losses.

Arthur

New York City, New York, U.S. (www.arthur.ai/)

Analysis by Arun Chandrasekaran

Why Cool: Arthur's integrated machine learning monitoring platform provides a single pane of glass that enables bias detection and explainability. It also offers model performance monitoring by detecting data drift early in the cycle. Arthur's bias detection capabilities are differentiated through a built-in fairness threshold for protected classes, unique heat map representations for computer vision use cases and detailed visualizations for fairness metrics, enabling its usage by both business and technical users. Another key differentiation built into its platform is counterfactual explanations for AI, which are critical to understanding how to lower risk effectively.

Counterfactual explanations analyze the model and the data to find the easiest path to a different, more desirable outcome. For instance, if a model identifies you as being high risk for a health condition such as diabetes or heart disease, a counterfactual explanation identifies the easiest route to lowering the risk identified by the model. While there are existing approaches that work across all algorithms such as LIME and SHAP, they are constrained by limitations across specific use cases. Arthur enables IT leaders to build a common explainbility framework across use cases and data types — tabular data, computer vision and NLP — and ensures the outputs of these techniques are actionable.

In summary, Arthur provides an AI-model-monitoring solution that delivers three key capabilities:

- Helps identify unjust bias in your models and provides recommendations for rectifying them to ensure that customer-facing models are accurate and you are treating your customers, partners and employees fairly.
- Provides real-time visibility into model performance and outcomes. It alerts the AI engineering and data science teams to data drift in models or simply models not delivering against performance service-level agreements (SLAs).

- Engenders trust in your AI systems with prediction-level explainability, which is critical for both decision analysis and regulatory compliance.

Arthur provides model monitoring across on-premises and cloud-based data science platforms and supports several popular AI frameworks. The founders of Arthur come from Capital One, an early adopter of AI in financial services, where they held leadership positions within its AI team. Their existing customers include large enterprises such as Humana as well as startups.

Challenges:

- Many data science platform vendors include explainable artificial intelligence (XAI) capabilities as part of a broader platform. There are also several emerging startups such as Fiddler, Imandra and Kyndi in this space.
- The company's product documentation, support and go-to-market strategies are quite nascent today and need to evolve to cater to the exacting requirements of large enterprises. Moreover, understanding of regulations at both a geographic and an industry level is critical for future success.

Who Should Care: Data and analytics and AI governance leaders looking for effective explainability and model monitoring with actionable recommendations in regulated industries such as financial services, healthcare, high technology and government should consider Arthur in their shortlist for XAI solutions.

DreamQuark

Paris, France (www.dreamquark.com/)

Analysis by Moutusi Sau and Farhan Choudhary

Why Cool: DreamQuark provides explainability, bias detection and regulatory compliance in the financial industry through its product Brain. Brain has its own fairness engine inbuilt along with a data leakage detector and model report generator, and a data obfuscator addresses security, privacy and auditability concerns. The product allows organizations to scale their AI initiatives by end-to-end machine learning (ML) orchestration with out-of-the-box monitoring, explainability, governance, security and risk management capabilities.

Explainability is the core strength of the platform: DreamQuark uses a variety of methods and technologies to explain model behavior in ad hoc and post-hoc scenarios. DreamQuark uses open-source explainability tools like SHAP, as well as its own open-source TabNet (attentive Interpretable Tabular Learning).

Brain operates in a SaaS model that allows users a fast and secure way to scale AI initiatives. This is important because Gartner's AI in Organizations Survey also revealed that one of the top barriers to AI implementations was security and privacy concerns. Reference customers reported AI models deployed in production in about four weeks, while the 2019 AI in Organizations Survey shows that it takes, on average, nine months to productionalize AI models. The platform delivers model monitoring and governance: Every model Brain generates comes with a detailed report that can be customized for auditors and regulators. The platform also tracks concept and data drift, and raises individual warnings where necessary. Brain also allows business users to play with data inputs using a model simulator that comes with a host of explainability features that can be used for better collaboration and decision augmentation.

Challenges:

- Currently, the use cases are targeted to a niche constituency, which may hinder adoption and growth for DreamQuark. The solution is focused for wealth managers but is starting to come up with use cases and applications for retail banks.
- DreamQuark is geography-focused — Continental Europe, U.K., Switzerland, Singapore, Japan and Canada.
- The Brain engine works for models that are built on their own platform. Importing and exporting models from other platforms is not yet supported.

Who Should Care:

- Wealth managers, insurers, asset managers and retail banks looking for an easy-to-use off-the-shelf solution with advanced capabilities in XAI, model monitoring and governance should consider DreamQuark.
- Banking CIOs with security and privacy concerns around model deployment and orchestration can leverage the Brain platform.
- Banking CIOs should explore this product in lending practices for their easy-to-use explainability features.

Hazy

London, U.K. (hazy.com/)

Analysis by Sumit Agarwal and Farhan Choudhary

Why Cool: Hazy's approach to generate synthetic data uses a combination of differential privacy, Bayesian networks, generative adversarial networks (GANs) and various machine learning algorithms. Traditional anonymization techniques such as masking or randomization are often not secure and face limitations in maintaining referential integrity across multiple data entities. Traditional test data generation techniques are also not sufficient in capturing the data patterns and require manual analysis to define the patterns. Hazy offers a synthesizer to train models using real customer data. These models are then passed to the Hazy hub that includes the generator, which creates the synthetic data. The synthesizer and the generator can be in two different environments. No data is passed between these components, ensuring data security and privacy as part of this process.

In addition, Hazy applies differential privacy by adding noise to the data. A small amount of noise may not provide sufficient privacy, while too much noise may reduce the utility of data. The amount of noise can be configured by adjusting the value of epsilon (ε) in the Hazy platform. This approach provides a scalable and generic platform that can be extended to different modeling methods.

A majority of Hazy's customers are from financial services, in addition to consulting and government customers.

Challenges:

- Hazy's architecture includes a combination of generic core components — synthesizer to train models, generator to create synthetic data. The solution needs manual steps and configuration specific to each dataset for the model training step and requires execution of a code-based client leveraging the client library included with the product. The Hazy product and professional services team also provides additional support to get clients started. Hazy will need to include more automation and generalization features as its product implementations scale.

- Hazy's challenge is to win the tabular synthetic data market, which is of high interest to the enterprises and to vendors who need representative data to demo and quickly prove their solutions. The synthetic data market has expanded substantially, but Hazy had the early entrant advantage in this market.

Who Should Care: Synthetic data has several use cases, and organizations considering the Hazy solution should evaluate the following:

- Internal constraints with sharing of real customer data with the data scientists due to privacy requirements. This may also be used for provisioning of data science sandboxes.
- Data sharing across internal teams or with an external third-party in a secure and privacy-protected manner. This also supports rapid prototyping by vendors and enterprises limited by data availability or data privacy regulations.
- The decoupling of the synthesizer and generator allows migration of only the synthetic data model to the cloud without having to copy real customer data to the cloud. Organizations may use this functionality to generate synthetic data in the cloud for AI and analytics development and testing.
- Testing models for edge cases is often a challenge. Synthetic data provides a mechanism to generate data for such hard-to-test scenarios.

Synthesized

London, U.K. (www.synthesized.io/)

Analysis by Avivah Litan

Why Cool: Synthesized is cool because it uses AI and GANs to generate synthetic data, and it systematically identifies and mitigates bias in the resulting dataset. Synthesized addresses issues concerning both AI-related privacy and bias, while retaining the same statistical power in the model's original dataset. Bias identification and remediation features support model development and collaboration across organizations without sacrificing data privacy.

Synthesized launched Bias Mitigation in December 2020 in the U.S. Synthesized Bias Identification and Remediation provides users with the ability to test the software for free. The software identifies key variables that can contribute to bias, such as gender, age and income, and which impact decisions. It profiles all possible groups, including those protected under discrimination laws, in the original dataset, and it compares the distribution of predictive scores, such as credit scores, across the groups to make sure there is no statistical bias among them. Synthesized assigns the model a fairness score, based on bias, that is identified and gives users an actionable bias mitigation plan. This includes the ability to equally distribute all attributes within a dataset to remove bias or to manually change singular data attributes (such as legally protected elements like gender) within the dataset, thus enabling granular control over the rebalancing.

The vendor also measures model accuracy and performance with bias removed. Gartner's social media analysis found that customers favor Synthesized when they save time by creating compliance-ready datasets and by providing automated workflows. They engaged publicly with the U.K.'s Financial Conduct Authority (FCA) and the City of London in November 2020 in a "Digital Sandbox Pilot" exercise to develop and utilize synthetic datasets for bank transaction fraud detection. The Alan Turing Institute vetted the Synthesized dataset, reporting it was highly accurate and rendered the same properties as the original dataset.

Challenges: Users can become too comfortable thinking that the Synthesized platform relieves them of responsibility for doing their own manual examinations and analyses of bias in their AI systems. Synthesized must provide its customers with guidelines for human involvement to avoid disappointment as a result of fully delegating bias detection to the software.

Synthesized's model is not explainable since it is based on deep neural networks (DNNs). The firm needs to do a better job interpreting how the model works, to the extent possible based on outcomes, for the different audiences.

The company competes with many firms providing privacy enhancing technology for AI data. Some of these firms have more than one method, such as synthetic data and differential privacy, for accomplishing the goal of protecting and sharing data. Synthesized for now is a one-trick pony, and data and analytics leaders should broaden its offerings to include other privacy-enhancing techniques, whose application will depend on the various use cases.

Who Should Care:

- Data and analytics leaders who need a collaboration on AI models internally and with other organizations but are inhibited to do so because of privacy and data leakage concerns should consider Synthesized. This includes all interested parties in an organization, including AI developers, data scientists, line of business leads, compliance officers, privacy and security staff, legal advisers, and risk managers.

- Organizations who need quality data for model development and cannot obtain it directly should consider this vendor.

- Compliance officers and business managers responsible for AI that identifies and mitigates hidden biases from their AI systems should also consider Synthesized.

TruEra

Redwood City, California, U.S. (truera.com/)

Analysis by Arun Chandrasekaran and Shubhangi Vashisth

Why Cool: TruEra offers AI quality management solutions for XAI, fairness and model quality management. It makes the model development and validation process structured, analytical and consistent, with explainability at its core. The company's solutions are cool compared to other XAI solutions in the following ways:

- TruEra enables data science teams and AI governance officers to apply a best-practice set of model quality and comparison analytics to each model version across all development platforms. TruEra's approach is model-agnostic.

- TruEra offers a fairness workflow that includes a comprehensive set of bias and fairness metrics, identification of disparities among groups, and mitigation of unfairness guided by the root cause analysis.

- TruEra can accurately answer a broad set of questions around causal drivers of model predictions for individuals as well as aggregate trends relevant to measuring bias and data drift.

- TruEra delivers capabilities for review and governance workflow. It can help data scientists understand, for example, sample selection bias, overfitting and incorrect labeling of data, and can go deeper to assess risk exposure. The vendor uses sensitivity analysis under the hood.

- TruEra has a regulatory engagement program in the U.K., Singapore and the U.S. It can also give customers a self-assessment against regulatory AI guidelines.

TruEra supports a number of popular data science and ML platforms, as well as open-source libraries and frameworks. The company's approach is based on the founders' prior research at Carnegie Mellon University. Standard Chartered Bank is one of TruEra's customers.

Challenges:

- TruEra's solutions are new. The first solution, TruEra diagnostics, became generally available in late 2020. Although the vendor's direction has been well-received, the second solution, TruEra monitoring, released in 2021, has no confirmed results. However, all XAI solutions as a category are quite new, so customers should assess risks against the benefits of such solutions.
- Most data science and ML platforms offer a breadth of XAI capabilities and continue to improve upon those. At the same time, there is a growing set of specialized vendors, including those whom we showcased in our prior Cool Vendors reports, adding to the competition.
- TruEra would need to continue its innovation journey and expand into the area of unstructured data explainability for use cases related with NLP or computer vision. Adoption of XAI solutions is often predicated on a high degree of customer maturity for AI, which limits the pool of initial target customers.

Who Should Care:

- Data and analytics leaders looking for explainability in regulated industries such as financial services, pharmaceuticals, government and healthcare should consider TruEra to address their needs for XAI.
- MLOps teams that look for explainability to assure their production readiness and operations should consider this vendor.
- Data science teams that have attrition concerns may benefit from explainability and transparency to help ensure that, if data scientists quit, they would not leave a black box behind.

Evidence

Gartner's 2019 AI in Organizations Survey was conducted online from November through December 2019 among 607 respondents from organizations in the U.S., Germany and the U.K. Quotas were established for company size and for industries to ensure the sample was a good representation across industries and company sizes. Organizations were required to have developed AI or to intend to deploy AI within the next three years.

Respondents were screened to be part of the organization's corporate leadership or report into corporate leadership roles, have a high level of involvement with at least one AI initiative, and have one of the following roles when related to AI in their organizations: determine AI business objectives, measure the value derived from AI initiatives or manage AI initiatives development and implementation.

Results of this study do not represent global findings or the market as a whole but reflect sentiment of the respondents and companies surveyed.

Social media analytics methodology: Gartner conducts social listening analysis leveraging third-party data tools to complement or supplement the other fact bases presented in this document. Due to its qualitative and organic nature, the results should not be used separately from the rest of this research. No conclusions should be drawn from this data alone as it may not be entirely market representative. Social media data in reference is from 22 February 2019 through 21 February 2021 in all geographies (except China) and recognized languages.

[1] Additional research contributions were provided by the Gartner Social Media Analytics team.

Recommended by the Authors

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

Top Trends in Data and Analytics for 2021: Smarter, More Responsible and Scalable AI

Build AI-Specific Governance on Three Cornerstones: Trust, Transparency and Diversity

Financial Services CIOs Must Focus AI Investments on 'Responsible AI' in 2021

What Non-Technology Executives Should Do in Support of Responsible AI Initiatives

Build Trust in AI Through Explainability

Toolkit: RFP for Data Science and Machine Learning Platforms