



451 Research Market Insight Report Reprint

Observability is made for AI use cases

May 24, 2024

by **Mike Fratto**

The idea of artificial intelligence taking over the world is not far-fetched. In consumer and enterprise products and services, research institutions, and most fields of employment, AI is being developed and used. The path to its benefits will take some time, however, as software vendors navigate the complexities of adding AI to customer applications.

S&P Global
Market Intelligence

This report, licensed to Kloudfuse, developed and as provided by S&P Global Market Intelligence (S&P), was published as part of S&P's syndicated market insight subscription service. It shall be owned in its entirety by S&P. This report is solely intended for use by the recipient and may not be reproduced or re-posted, in whole or in part, by the recipient without express permission from S&P.

Introduction

The idea of artificial intelligence taking over the world is not far-fetched. In consumer and enterprise products and services, research institutions and most fields of employment, AI is being developed and used. This is particularly true in IT operations, since machine learning and AI can have a significant impact on identifying security issues and configuration errors, as well as find ways to optimize IT systems, predict issues before they occur, and help IT administrators understand increasingly complex systems. However, the path to those benefits will take time as software vendors navigate the complexities of adding AI to customer applications.

THE TAKE

AI and observability hold great promise for enterprises streamlining and speeding up IT operations, and for IT staff struggling to keep up with their users' demand for more agile applications that are adaptive to change and resilient to outages. Observability vendors are adding AI capabilities to their products and services to help customers make better use of their offerings and discover features easily. AI expands the applicability of observability beyond IT-only to nontechnical users who can use natural language queries to create visualizations and dashboards, which would normally require expert product development.

Context

For decades, IT management vendors have been promising automated intelligent management, but ultimately lacked the tooling to dynamically learn and adapt to changes in applications and their infrastructure. In large part, the data was collected, but making sense of it was extremely difficult; therefore, taking actions and making recommendations relied on a growing number of static rules that never worked as well as promised.

Advances in ML and AI are changing that dynamic, since both are well suited to analyzing vast amounts of data and using it along with other data sources like stored institutional knowledge in service desks, product documentation and wikis. This can help with troubleshooting and root-cause analysis, recommend remediation steps and configuration changes, and predict the likelihood of failures before they occur, which shortens downtime and improves the reliability of IT applications.

Early successes with enterprises are paving the way for greater adoption, with 50% of respondents to 451 Research's AI & Machine Learning, Infrastructure 2023 survey saying they are currently investing in AI for IT services; while 47% of those not yet using AI for IT services management plan to do so in the next two years.

AI use cases

We see a few use cases from vendors on a regular basis, like GenAI for feature documentation and self-service support, co-pilots for code and query generation, and data analysis. This is not an exhaustive list of use cases, and there are many approaches to achieving similar goals, but these seem to solve significant pain points for customers, while giving vendors time to develop more advanced capabilities in analysis and automation.

GenAI

This is one of the first places vendors start with implementing AI in products, and will often begin with loading curated content — from documentation, service tickets and public sources like forum posts, for example — into an existing large language model. The resulting language model helps answer questions about product features and uses, which can be a powerful tool to onboard new technical and nontechnical users by taking them directly to content for the task at hand.

GenAI is often used in single-ask questions without contextual follow up. More advanced uses of GenAI offer an interactive experience where follow-up questions take the previous questions and answers into account. Both approaches are useful, but the latter's contextual and interactive features, employed by vendors like BigPanda and Observ, are more in line with user expectations.

The inputs into the language model should be curated, ensuring the GenAI does not generate erroneous answers (aka hallucinate). Sensitive data such as company and user names should be scrubbed from the public sources, leaving an information base that users can ask questions of. GenAI may or may not use retrieval augmented generation, a technique where live, or more current, data is accessed by the AI to augment its trained models.

GenAI is also fairly low risk for the vendor, provided its data sources are properly curated, it scrubs sensitive data and does not include queries in the model. Vendors do not need to train or host their own models. They can use one of the many LLM services to host their data, which would be especially attractive for cloud-based observability vendors.

Co-pilots

Co-pilots are another use case that enhances the useability of the observability platform, helping technical and nontechnical users with creating queries and dashboards. This opens up the platforms to a much wider audience than technical IT, and makes the observability software more accessible, and so more valuable overall.

Even for experienced IT professionals, building queries using custom query languages poses a challenge, unless they do it often enough to become proficient. Even so, creating and optimizing queries takes time and expertise that may not be transferrable to other systems, fostering a sense of lock-in and presenting a hurdle to adopting new platforms. Nontechnical staff would likely struggle with writing queries. Co-pilots help users who know what they want to ask, but cannot build the queries themselves. Some observability services, like Chronosphere, will analyze queries that may be long running, and optimize them or even stop them from running.

Similar to the state of GenAI, most vendors have or are working on co-pilot features that are focused on query building and optimization, since those are the types of tasks customers struggle with most. Dynatrace Inc. announced in 2023 its Davis CoPilot which, among other things, brings a natural language capability that allows administrators to create queries and dashboards, automate root-cause analysis, document changes, and create event rules. None of these capabilities is unique to Dynatrace, but the breadth currently is.

With other observability products, natural language queries are usually single-shot interactions, but more advanced interactions like contextual refinements are coming. Dashboard building using co-pilots should simplify the difficult task of building useful, informative dashboards, and will likely open up observability data and analysis to a much wider audience.

Data analysis

Data analysis might be the AI use case that is most commonly thought of, but it is exceedingly difficult to build an AI system trained on enough environments while preserving customers' private data. Great strides have been made with ML in surfacing patterns in data, and have been put to good effect in security alerting. In observability, AI and ML can be applied to the data that is collected as it is streamed in, something observability pipeline vendors like Cribl and Mezmo are doing, and it can be applied to processed data and the observability platform itself.

Observability vendors like BigPanda, Cisco Systems Inc. (AppDynamics), Dynatrace and Splunk Inc. are using AI for tasks like pattern recognition in the collected data, and correlating it to find application topologies and dependencies. AI helps users make sense of the flood of data generated from complex and dynamic applications, with the goal of surfacing events that impact application performance and reliability, overlay those events on an application topology or dependency graph to assess the impact to dependent applications and services, and help IT administrators get to root cause and remediation quickly. This is the area of AI that requires the most research and development, because it is central to nearly all other use cases of AI aside from documentation and product usage.

However, not all AI requires the use of pretrained models to be effective. Grok.ai is shipping an AIOps platform that uses ML and AI trained solely on the customer's data in real time, using several techniques including clustering and reinforced unsupervised learning. These techniques relate events that repeatedly happen together, forming increasingly stronger correlations the more often the pattern manifests. Those events become predictors of failure, which is useful for improving operations.

AI-powered recommendations

Along with the use of AI, enterprises are expanding their use of automation to help streamline all facets of IT management. Combining AI with IT automation enables recommendations to remediate or optimize applications, or suggest impending failure in applications based on events streaming into the platform and the training of the AI.

Software vendors like Pagerduty Inc. and ServiceNow Inc. can automate IT workflows from within their own platforms and integrate with external IT service management systems, including competitive observability platforms, as part of an orchestrated effort. AIOps platforms such as BigPanda and Dell APEX AIOps, in part from Dell Technologies Inc.'s acquisition of Moogsoft in 2023, have been using ML, and now AI, to identify issues and suggest remediation steps and optimizations. The upshot of having so many AI systems in play is that enterprises will have to prioritize one over the others to achieve consistency in decision-making.

The future

Dynatrace and Cisco with its ApplicationDynamics have been shipping observability capabilities specifically aimed at how AI applications are behaving, evaluating the cost or token usage for AI services, and in some cases monitoring the queries and data being sent to the AI, to reduce the exposure of sensitive data. Given the size of the training models, the cost and complexity of training and maintaining models, and (depending on the use case) the amount of infrastructure used for answering queries, many enterprises are looking to cloud-based services for their AI needs.

Applying observability techniques to these services should provide valuable insight into cost and performance. Greater insight can be gained when the AI models and applications are owned by the enterprise and running on its own cloud or on-premises infrastructure. Similar to cloud cost management, we expect to see an increase in AI application observability to help companies gain control over critical performance and spending.

The big question is: When does AI come to your observability platform? It is already starting to arrive, and we expect its use to increase and improve. Nearly all the observability vendors we have spoken with have AI on their road maps, and we expect to see many new product announcements and enhancements over the next 12-24 months.

CONTACTS

Americas: +1 800 447 2273

Japan: +81 3 6262 1887

Asia-Pacific: +60 4 291 3600

Europe, Middle East, Africa: +44 (0) 134 432 8300

www.spglobal.com/marketintelligence

www.spglobal.com/en/enterprise/about/contact-us.html

Copyright © 2024 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global keeps certain activities of its divisions separate from each other to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain nonpublic information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its websites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.