

Increase speed and accuracy with AI-driven static analysis auditing

Audit Assistant leverages past audit decisions to power machine learning-assisted auditing, validating results immediately and dramatically reducing manual auditing efforts and excess noise in scan results



The Audit Assistant service uses machine learning algorithms from hundreds of millions of anonymous audit decisions. These decision models can be automatically applied to OpenText Static Application Security Testing results by using Audit Assistant.

Machine learning for auditing

OpenText Application Security as a Service offering OpenText™ Core Application Security runs thousands of static, dynamic, and mobile scans per week, scanning billions of lines of code. OpenText Core Application Security takes customer application source code, runs the scan, then (as a value added service) passes these raw scan results to a team of expert auditors who are subject matter experts. These auditors identify and prioritize the noteworthy findings while removing the noise from the results. Consequently, OpenText Core Application Security customers receive actionable results and can primarily focus on fixing the most critical issues.

The Audit Assistant service uses machine learning algorithms to feed off the hundreds of millions of anonymous audit decisions from OpenText Core Application Security experts. These decision models are actively used and developed for OpenText Core Application Security, but are also technologies that can be automatically applied on premises to OpenText™ Static Application Security Testing results by using Audit Assistant.

Leveling up the AI

Account for model drift

Model drift is a statistical phenomenon where, as the world changes, initially effective models may have new opportunities to improve. Cybersecurity does not stand still: novel attacks and vulnerabilities appear all the time, and new countermeasures and detections respond in kind.

The new Audit Assistant models provide a way to refresh the approach and address any model drift that may have happened since the technology was first pioneered, along with designing for a learning pipeline that will be more resilient to drift going forward.

Incorporate additional data and context

Key to accurate machine learning is always data: the more available data, the more contextual clues are afforded to the AI to learn from. The more granular the target to predict, the more effective and precise the trained models can become.

The Audit Assistant models take in additional anonymized vulnerability metadata about the scans that allow them to be more predictive than ever before. In addition, while the original models predicted a binary value (e.g., false positive or not), the new models support more granular category labels than before (e.g., False Positive, Exploitable, Suspicious, Mitigated).

Audit Assistant gets better and better over time. The more you audit your vulnerabilities, the more the models learn about what's appropriate for your project.

More flexibility to learn from your unique environment

Audit Assistant models learn as a prerequisite from all of the expert human auditors and language experts on OpenText Core Application Security audit team—an incredible resource that is unparalleled in this industry. However, we recognize that every customer is different, every project is different.

We designed the new Audit Assistant model pipeline to also learn the unique behaviors in your projects. This learning gets better and better over time; the more you audit your vulnerabilities, the more the models learn about what's appropriate for your project—all while remaining sensitive to your IP, using anonymized vulnerability metadata, and without operating on the source code itself.

Design and testing methodology

Like software engineering, effective data science demands a disciplined approach to model development. The OpenText data science team started by testing the existing models to measure their strengths and weaknesses.

We also interviewed and worked with the OpenText Core Application Security audit teams to better understand how they applied their language and security expertise to perform their audits. We worked with the OpenText Core Application Security engineering teams to use real-world, anonymized production data for our training and testing. Finally, we performed exploratory data analyses and feature engineering to develop the new models on the latest data.

We performed model testing and validation on actual data from OpenText Core Application Security, over the course of 20 weeks. Because Audit Assistant models forecast into the future, we performed all model validation using proper back-testing approaches (i.e., training on past data, comparing results against future data).

As we deploy these new models into production, the mechanisms to validate and measure their performance happens continuously and automatically

Results of our testing

Accuracy

The new models show strong accuracy performance. In our testing, using 20 weeks of real-world data from our production OpenText Core Application Security systems, our new models were 60% more accurate, on average, for predictions compared to the previous generation's models. In particular, the JavaScript, Java, and HTML models demonstrated great improvements.

More accurate AI than ever before

Next-generation Audit Assistant models are more accurate than ever before. In our testing of real data over a 20 week period, there was an overall 60% accuracy improvement compared to the previous models. And it gets even better over time!

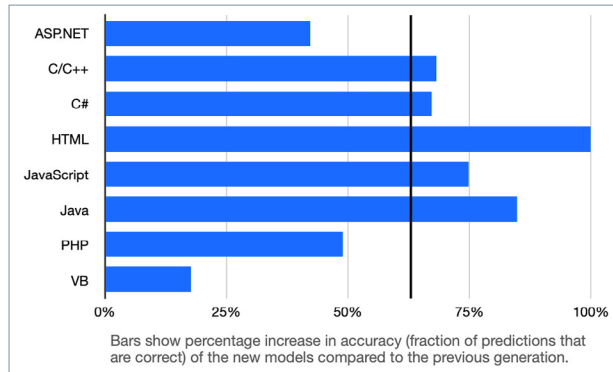


Figure 1. Improved model accuracy

Recall

When measuring the performance of machine learning models, it is important to not only focus on how many predictions the models got right, but also how many items were detected. This is measured through something called “recall,” which is the percentage of issues correctly identified. A high recall score means fewer false negatives and that little slips through the cracks.

The new Audit Assistant models show an average recall of nearly 80 percent in our testing, which is excellent.

Also, while the previous models were binary classifiers (False Positive or not), the new Audit Assistant models are multi-class predictors that learn to predict all the observed category labels. This additional granularity is helpful if your organization uses custom category labels beyond the standard default set.

More granular predictions that don’t miss a beat

The new Audit Assistant models are now able to learn to predict different category labels, such as Suspicious or Exploitable. These multi-class models have excellent recall performance, identifying on average 80% of all issues and nearly 100% in certain cases.

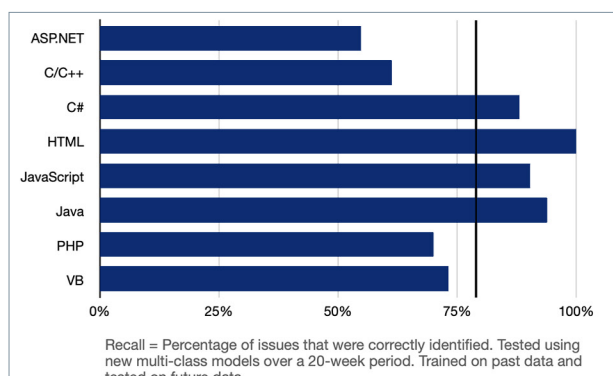


Figure 2. Recall performance

Resources

Audit Assistant Web Page

[Learn more >](#)

Audit Assistant Datasheet

[Learn more >](#)

Accuracy over time

As machine learning models continue to consume more training data over time, they become more accurate. The new Audit Assistant model pipeline balances recent and past data; recent behaviors are weighted more highly than distant past behaviors. Further, there is also a balance between the global model (trained on the OpenText Core Application Security data that represents aggregate learning across many customers of the OpenText Core Application Security audit service) and the local model (behaviors that are unique to your organization and project).

These new features help address model drift while ensuring that models get better, not worse, over time.

AI that keeps up with you

The new Audit Assistant models have new features that allow it to keep up with your unique project, project language, or organizational behaviors... even if it changes over time. The more time the models spend learning from your system, the better they become.

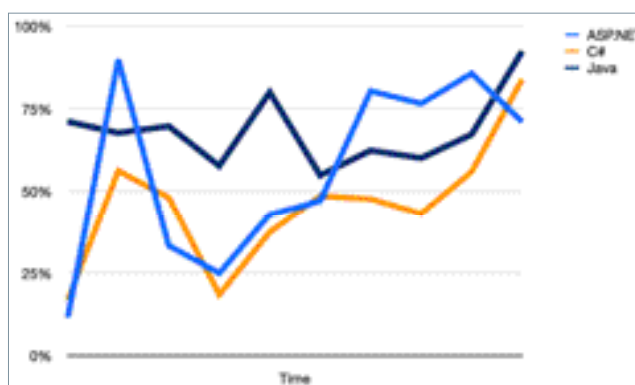


Figure 3. Model accuracy over time

Conclusion

Auditing raw static scan results is the most time-consuming and effort-intensive manual aspect of SAST and requires a skillset that is often difficult to find and keep.

The initial release of Audit Assistant was ground-breaking and an industry first. This capability uses machine learning to democratize the OpenText Core Application Security human audit team's expertise to help close the skills shortage in performing effective audits.

Now, five years later, the next generation of models for Audit Assistant make the AI more effective than ever. We can't wait to get these next-generation models into your hands