January 2026

Comparing automated accessibility tools

An independent assessment

An evidence-based look at which automated accessibility tools deliver meaningful WCAG coverage.

adience

Table of contents

01
Executive summary

02
Methodology

03
Research results: WCAG standards coverage

04
Research results: valid issues

05
Summary

01

Executive summary

A controlled comparison revealing major differences in how five leading accessibility tools detect WCAG issues—and what those differences mean for buyers.

As digital accessibility requirements continue to tighten, organizations face growing pressure to ensure their websites and digital services meet WCAG standards.

Automated accessibility tools are often the first line of defense in identifying potential issues. Yet buyers are presented with a crowded market of solutions that, on the surface, appear to offer similar coverage.

In practice, however, differences in what these tools detect—and what they do not—can affect how accessibility risks are identified and prioritized.

This report summarizes a comparative analysis conducted by Adience on behalf of AudioEye to evaluate the relative performance of five leading automated accessibility tools. The aim was to assess how these solutions performed on identical tasks across well-known websites, and how effectively they detected issues aligned with WCAG Levels A, AA, and AAA.

The research reveals that detection results varied sharply across the five tools. Some identified comparatively few issues—or, in some instances, none—at certain conformance levels, whereas others demonstrated selective strengths.

This variation matters: the consequences of non-compliance can be significant, so buyers must be confident that their chosen tool provides robust and reliable detection. Although these tools often promise broad coverage and rapid issue resolution, this research provides an independent comparison using a controlled and repeatable methodology to help potential buyers make an informed choice.

> Adience applied the same rigorous testing approach to every tool, ensuring each was evaluated under identical controlled conditions. What surprised us most was the significant scale of variation in the results.
>
> Chris Wells, Managing Director, Adience

Key findings at a glance

Detection varied widely across the five tools tested. Some tools returned very few findings—or none at all—at certain WCAG levels on specific sites, even though other tools identified multiple valid issues on the same pages.

Tool C (AudioEye) demonstrated the most consistent detection coverage. Within the scope of this experiment, it identified validated issues across all six websites and all WCAG levels (A, AA, and AAA), whereas other tools identified issues on only selected sites and levels. For example, Tool C (AudioEye) detected 509% more WCAG Level A issues than the worst-performing tool.

Tool C (AudioEye) identified more unique WCAG success criteria through automated testing than any other tool, returning findings for 10 unique success criteria, whereas other tools could only automatically detect 7-8.

Variation increased at more stringent WCAG levels. Differences among tools were most pronounced at Levels AA and AAA, where several tools failed to report any findings across multiple sites, highlighting inconsistent application of automated rules at these levels, and indicating meaningful variation in how tools surface advanced accessibility barriers.

Implications for buyers

These findings show that not all automated accessibility tools provide the same visibility into accessibility risk. A tool that detects fewer issues may appear to indicate lower risk, but this research demonstrates that some tools may give a false sense of security with limited detection capabilities. By contrast, broader and more consistent detection may enable earlier identification of gaps and more informed remediation decisions.

Within the scope of this experiment, one tool—Tool C (AudioEye)—demonstrated broader and more consistent automated detection across sites and WCAG levels. We hope buyers will use these findings as part of their due diligence when selecting an accessibility testing solution.

02

Methodology

Adience conducted an independent evaluation of five leading automated accessibility tools to compare how each product performed when assessing identical website pages against WCAG criteria.

The purpose was to examine differences in automated detection behavior under controlled conditions, using a structured, repeatable testing approach to provide buyers with an evidence-based comparison of their performance.

The tools

Five widely used accessibility tools were included. To avoid commercial sensitivities, the report refers to them as Tool A, Tool B, Tool C (AudioEye), Tool D, and Tool E throughout. The tools included in the analysis are referred to anonymously throughout this report. For transparency, the set of products assessed included widely used tools from these vendors: EqualWeb, Deque, UserWay, and accessiBe.

The websites

Adience used each tool to analyze the same set of URLs using its automated testing functionality. The analysis reviewed the issues identified by each tool's automated detection processes.

This approach ensured that the comparison focused solely on automated detection capability, replicating how typical buyers might use these products in their accessibility testing.

The websites were:

BEST BUY
Starling Bank
Yale
U.S. Citizenship and Immigration Services
Cleveland Clinic
cognism

WCAG criteria assessed

The analysis measured how each tool performed against the WCAG framework at Levels A, AA, and AAA. Detection was evaluated at both the standards level (i.e., the number of WCAG success criteria for which each tool returned any findings) and the issue level (i.e., the number of true positives, or valid issues reported by the tool).

Measures

The study focused on two core outputs:

Standards coverage – the breadth of WCAG guidelines for which each tool returned at least one automated detection.

Valid issues (true positives) – verified accessibility issues identified by each tool. Totals were aggregated by WCAG level (A/AA/AAA) and by site to understand both volume and consistency in detection.

Scope of the report and exclusions

The analysis focuses on WCAG standards coverage and validated issues identified through automated testing. Other outputs, including false positive rates, issue severity, legal coverage, and assessments of the quality of automated fixes, were excluded either because they require subjective interpretation or because tools do not present this information in a directly comparable way.

While false positives can have a material impact on the effort needed to check and resolve reported issues, thereby wasting valuable time and resources, they were excluded here to maintain a strictly objective, like-for-like comparison across tools. These exclusions ensure the analysis remains focused on objective, like-for-like comparisons across tools.

The data

Adience combined all the outputs from the automated scans into a single dataset. The research team reviewed detection patterns by site, WCAG level, unique standard, and number of valid issues to identify differences in breadth, depth, and consistency of coverage across the five products.

03

Research results: WCAG standards coverage

Let us begin by examining how each of the five tools performed with respect to WCAG standard coverage in the research.

Here, Adience recorded the number and range of WCAG success criteria for which a tool returned at least one finding during automated testing.

Across the dataset, the tools varied considerably in the range of WCAG success criteria they reported, with differences more noticeable at the higher WCAG levels. Some tools identified findings across a wide set of Level A, AA, and AAA success criteria, whereas others spotted fewer or none at certain levels.

Level A – Baseline standards coverage

At WCAG Level A, most tools returned findings against multiple success criteria on most sites. However, coverage was not uniform. Some tools failed to return any Level A findings for individual sites, even though other tools identified issues on the same pages.

Tool C (AudioEye) was able to identify findings against more Level A criteria than other tools.

Tool C (AudioEye), Tool A, and Tool E all returned Level A findings on all six websites tested.

Other tools demonstrated less consistent application of Level A automated rules, with gaps appearing on selected sites. These inconsistencies highlight that even baseline accessibility checks are not applied equally across the tools.

	# Level A criteria spotted across all sites	# websites Level A findings were spotted on
Tool A	5	6
Tool B	5	5
Tool C (AudioEye)	7	6
Tool D	5	5
Tool E	6	6

Level AA – Regulatory compliance standards

Variation in standards coverage increased at Level AA. Several tools returned findings against only a small number of AA success criteria on specific sites, and in some cases, returned no AA-level findings at all.

By contrast, Tool C (AudioEye) consistently returned Level AA findings across all the tested websites, indicating more comprehensive automated coverage at the compliance level.

Across the six sites tested:

- One tool returned zero Level AA findings on all six sites in this dataset.
- One tool returned zero Level AA findings on three of the six sites.
- Two returned zero Level AA findings on two of the six sites.
- Tool C (AudioEye) returned Level AA findings on all six sites.

Level AAA – Advanced standards coverage

Differences in detection were most pronounced at Level AAA. Coverage at this level was uneven and inconsistent across the tools, with some returning findings against a limited set of AAA success criteria and others failing to spot any AAA-level findings.

- One tool returned no Level AAA findings across all six sites.
- Four other tools returned Level AAA findings on all six sites.

Within the scope of this experiment, Tool C (AudioEye) consistently returned findings across a wider range of WCAG success criteria than the other tools. It was the only tool to return findings at all WCAG levels (A, AA, and AAA) across all six websites tested, whereas other tools returned no findings at certain levels on multiple sites.

	# Level AA criteria spotted across all sites	# websites Level AA findings were spotted on
Tool A	2	3
Tool B	1	4
Tool C (AudioEye)	2	6
Tool D	0	0
Tool E	2	4

	# Level AAA criteria spotted across all sites	# websites Level AAA findings were spotted on
Tool A	1	6
Tool B	1	6
Tool C (AudioEye)	1	6
Tool D	2	6
Tool E	0	0

Unique WCAG success criteria detected

In addition to differences in overall standards coverage, the analysis examined whether any of the tools identified WCAG success criteria that others did not detect within the scope of this experiment.

Based on the results shown below, Tool C (AudioEye) identified more unique WCAG success criteria through automated testing than any other tool, returning findings for 10 unique success criteria. By comparison, the other tools identified between seven and eight unique criteria each.

Examples of WCAG success criteria identified only by Tool C (AudioEye) in this experiment include:

- 1.4.11 Non-text Contrast
- 2.5.3 Label in Name

Key: ✓ Yes ✗ No

WCAG success criteria identified at least once across the six sites

Criteria	Tool A	Tool B	Tool C (AudioEye)	Tool D	Tool E
1.1.1	✓	✓	✓	✓	✓
1.3.1	✓	✓	✓	✓	✓
1.4.1	✗	✗	✗	✗	✓
1.4.3	✓	✗	✗	✗	✓
1.4.6	✗	✗	✗	✓	✗
1.4.11	✗	✗	✓	✗	✗
2.1.1	✗	✓	✗	✗	✓
2.4.1	✓	✓	✓	✓	✗
2.4.3	✗	✓	✓	✗	✗
2.4.4	✓	✗	✓	✗	✓
2.4.6	✗	✓	✓	✗	✓
2.4.7	✓	✗	✗	✗	✗
2.4.8	✗	✗	✗	✓	✗
2.5.3	✗	✗	✓	✗	✗
3.2.2	✗	✗	✗	✓	✗
3.2.5	✓	✓	✓	✗	✗
4.1.2	✓	✗	✓	✓	✓
Total identified	8	7	10	7	8

This highlights how differences in automated rule sets can affect not only the number of identified issues but also the specific accessibility requirements revealed during testing.

In summary:

1. Tools varied significantly in the breadth of WCAG success criteria they detected. Detection at the AA and AAA levels was inconsistent.
2. Some tools reported no issues across certain WCAG levels on multiple sites in the experiment. This risks giving a false read of compliance.
3. Tools differed not only in the number of issues identified, but in the range of WCAG success criteria surfaced. One tool (AudioEye) identified more unique criteria than the others in the experiment, highlighting variation in automated detection rules.
4. This variation illustrates how automated tools differ in their rule sets and in how they interpret or connect issues to WCAG references.
5. Within the scope of this experiment, Tool C (AudioEye) consistently returned findings across a wider range of WCAG success criteria than the other tools, indicating greater automated rule coverage.
6. In several instances, Tool C (AudioEye) identified findings mapped to WCAG success criteria that no other tool detected on the same pages. These are not listed exhaustively here, as the purpose of this section is to highlight patterns of coverage and the variation in automated detection across tools.

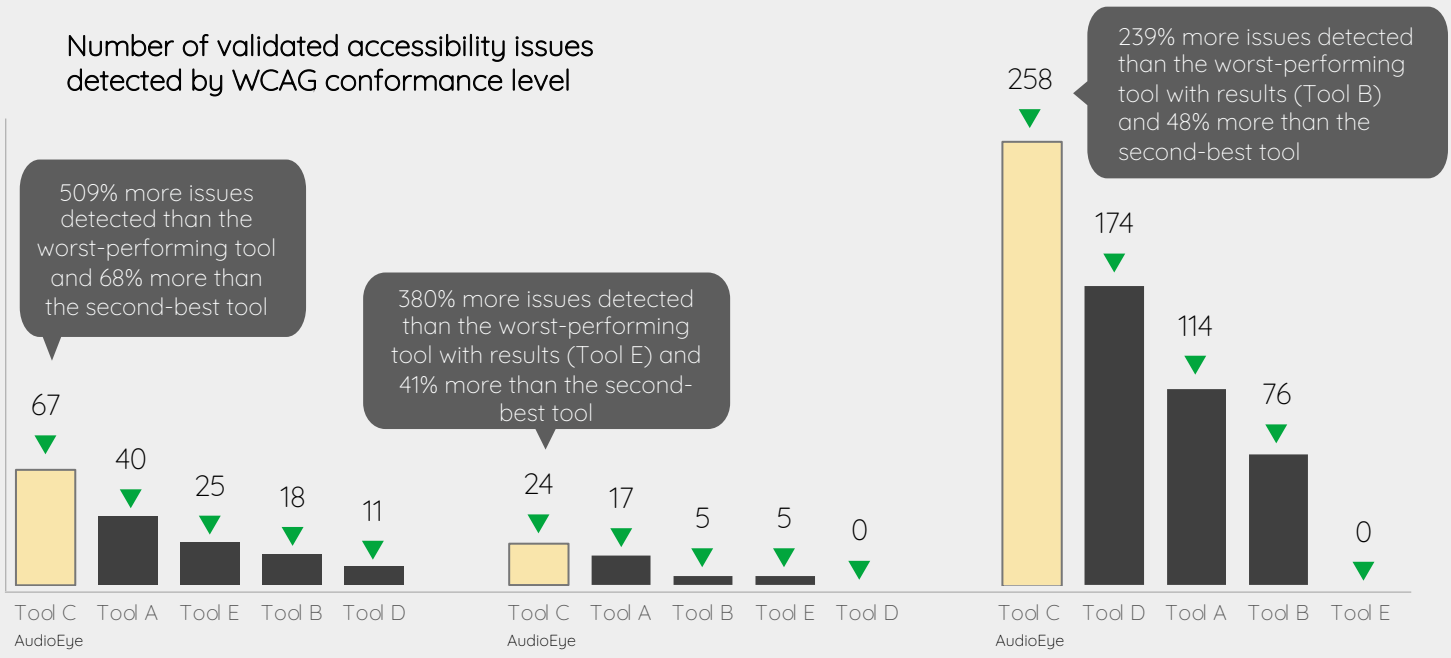
> Not all automated detection tools are equal. A tool that detects fewer issues may give the impression of stronger compliance than actually exists. In contrast, a tool that detects issues more broadly may support earlier identification and remediation of accessibility gaps.
>
> Chris Wells, Managing Director, Adience

04

Research results:
Valid issues

This section summarizes the number of valid issues (also described as 'true positives') identified by each tool across the tested websites. True positives represent accessibility issues verified as legitimate within the dataset. False positives were excluded from this study.

Once again, there was substantial variation in the number of valid issues detected by the five tools in our experiment. Some reported relatively few confirmed issues across all WCAG levels, while others surfaced a significantly higher number of findings on the same pages.



Level A - Baseline accessibility

Level AA - The legal compliance level in many jurisdictions.

Level AAA – The highest bar, and the least consistently detected.

Across all sites and conformance levels, one pattern emerged: AudioEye consistently identified more valid issues than the other tools. It returned confirmed findings for every site and every WCAG level, indicating broader automated detection coverage within the boundaries of this experiment.

The remaining tools varied in both volume and distribution of valid issues identified. In some cases, tools returned relatively few confirmed findings, even on pages where others identified substantially more issues, highlighting differences in the depth and consistency of automated detection.

> These results highlight the wide variability of automated accessibility tools. Buyers, of course, will want to carry out their own due diligence before selecting their tool of choice.
>
> Chris Wells, Managing Director, Adience

05

Summary

This study provides an independent comparison of five leading automated accessibility tools, using identical pages and the same testing conditions to assess how each product performs against the WCAG Level A, AA, and AAA success criteria.

The results show that automated detection capabilities vary significantly across tools. Some identified only a small number of issues at certain WCAG levels and, in some cases, reported no findings, even when other tools detected issues on the same pages. Others demonstrated more consistent performance, identifying issues across a broader range of standards and websites.

For buyers, these differences matter. Automated testing is often the first stage of an organization's accessibility management processes, setting expectations about where accessibility risks may exist. A tool that detects fewer issues may provide incomplete visibility into potential barriers, whereas one that surfaces issues more consistently may support earlier, more informed remediation.

Within this experiment, one tool—Tool C (AudioEye)—demonstrated broader detection coverage across all WCAG levels and all tested websites. Other tools showed selective strengths but did not match this level of consistency. These findings do not assess false positives, issue severity, legal exposure, or fix quality. Instead, they are intended to offer a clear, comparable view of how the five tools performed under controlled conditions.

Buyers can use this comparison to understand how automated detection capabilities differ between products and to consider how these differences may influence the identification and management of accessibility issues within their own organization.

Buyers can also use the learnings from this analysis to inform some of the questions they should ask vendors of accessibility tools during the purchasing process.

- Which WCAG success criteria does this tool actually detect through automation, and at which levels (A, AA, AAA)?
- How broad is the tool's automated rule set, and how often is it updated to reflect evolving WCAG interpretation?
- Can the tool demonstrate detection of higher-order issues, or does coverage drop off sharply at WCAG AA and AAA?

> The results highlight the importance for buyers to understand how tools differ in automated detection capability, as this can influence how effectively accessibility risks are identified and managed.
>
> Chris Wells, Managing Director, Adience

Thank you

To get in touch please contact:
chris.wells@adience.com