Warden AI

State of AI Bias in Talent Acquisition 2025

A data-driven review of AI bias, compliance, and responsible AI practices in TA.

AI TRUST ASSURANCE
Warden AI
INDEPENDENTLY EVALUATED

Foreword by Kyle Lagunas

Kyle Lagunas
Founder & Principal,
Kyle & Co

After a decade advising HR and Talent leaders on how to adopt technology responsibly, I've seen excitement around AI quickly give way to concern, especially around bias and fairness.

As AI tools embed deeper into hiring and workforce decisions, we face a defining moment: will we carry forward old inequalities, or use AI to build something better?

New and forthcoming legislation make this choice all the more urgent. Pending court cases signal that AI risk is no longer theoretical. For HR and TA leaders, and technology vendors, bias is now a legal and reputational issue.

But we must resist the temptation to simply run away from the "bias boogeyman." Instead, we need to lean in, and find real answers to the real risks we face.

This report brings a number of interesting points together to crystallize this critical conversation. It offers a grounded view of where AI is impacting TA decisions today and moves us past assumptions and toward evidence.

While bias is real, it is also measurable, manageable—and, thankfully, mitigatable.

Executive summary

Bias in AI has become one of the most urgent and emotive issues in hiring, with high-profile lawsuits like Mobley vs. Workday pushing the risks into the legal and media spotlight. This report cuts through the noise with a data-rich view of how bias is playing out in real-world AI systems.

75% of HR/TA leaders cite bias as one of the top concerns with adopting AI. At the same time, they are under pressure to solve business problems using AI, creating a sharp tension.

The reality is more nuanced than the headlines suggest. While some AI systems do show bias, **85% of AI systems met accepted fairness thresholds**.

In fact, AI can be fairer than the human processes they replace. The data suggests that **AI systems can deliver up to 45% fairer outcomes** for women and racial minorities than human-led decisions.

That said, bias does exist in some systems, **15% of tools failed basic fairness tests**, and the gap between AI systems and different vendors varies by up to 40%.

Vendors are investing in responsible AI practices and preparing to meet AI regulations, but many have a long way to go. There's a promising level of transparency with the majority of vendors sharing publicly available documentation. Yet transparency to end-users is limited with **only 15% made aware that AI is used**.

The question is no longer whether AI can be biased, but how we manage its use in a way that improves outcomes. Done right, AI offers a real opportunity to reduce longstanding inequalities, but only if it's developed, chosen, and deployed with care.

1. **Concerns over bias are slowing AI adoption**
   75% of HR leaders cite bias as a top concern when evaluating AI tools.
2. **AI can be demonstrably biased**
   15% of AI systems fail standard fairness benchmarks and many academic studies identify bias in lab settings.
3. **Yet AI is typically more fair than humans**
   In some cases, AI is up to 45% fairer for women and racial minorities.
4. **Vendors are investing in responsible AI practices**
   Practices are evolving but transparency to end-users is lacking.

About this report

This report analyzes high risk AI systems, such as those used to influence decisions in talent acquisition and intelligence, utilizing multiple data sources:

- **Bias audits from real systems**

  150+ audits of AI/automated tools using over 1 million test samples, tested for fairness using standard methods.

- **Survey data from vendors and practitioners**

  50+ responses on Responsible AI practices, transparency, and regulatory readiness.

- **Human bias benchmark**

  A blend of academic and industry studies on human bias, converted into comparable fairness metrics.

- **Public audits and explainability reports**

  100+ public documents from vendors sharing audits and transparency information.

150+
AI audits

50+
Survey contributions

100+
HR tech vendors

1 million +
Test samples

Warden AI

Special thanks to our contributors

Beamery
greenhouse
Popp
Classet
Endorsed
SquarePeg
sense
Juicebox
CONVERZAI
kula
Purplefish
icims
HeyMilo
inploi
SteppingStone
CodeSignal
CIELO
BLACK STAR AI
TECHWOLF
fuel 50
SPARK HIRE
Colleva
spotted zebra
apriora
Tezi
Censia
JOBMA
hackajob
findem
sapia
XOPA AI
MAXIMIZING OBJECTIVITY
Humanly

Table of contents

Perception

How big of a concern is AI bias?

"Buyers are asking more questions about AI bias and it's becoming a key factor in procurement decisions."

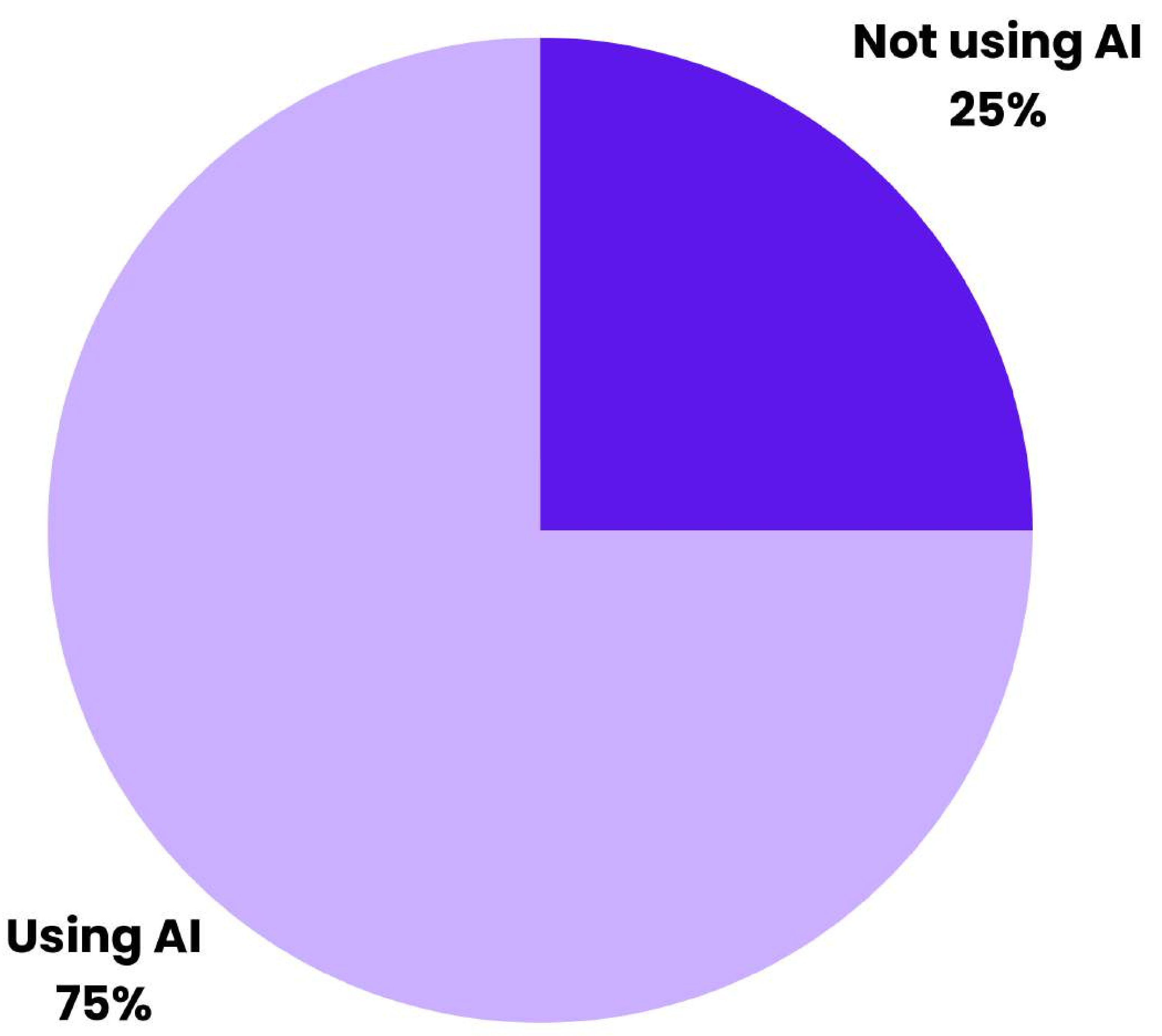
Jung-Kyu McCann
Chief Legal Officer,
Greenhouse Software

Warden AI

Perception

TA is embracing AI more than ever before

Three out of four TA teams say they're already using AI or automated systems.



What this means

- Pressure from executives to cut the bottom line.
- Fear of missing out from all the hype about impact of AI solutions.
- Increased tangible value in AI solutions.

Source: State of responsible AI survey, 2025

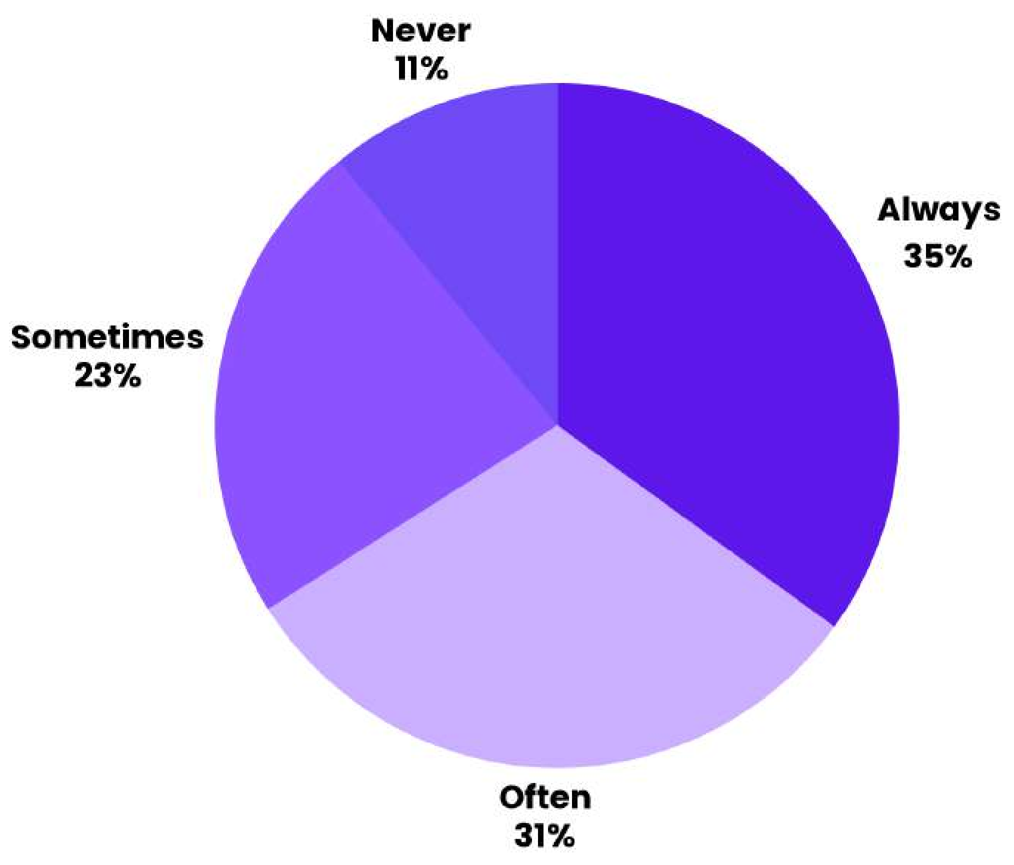
Warden AI

Perception

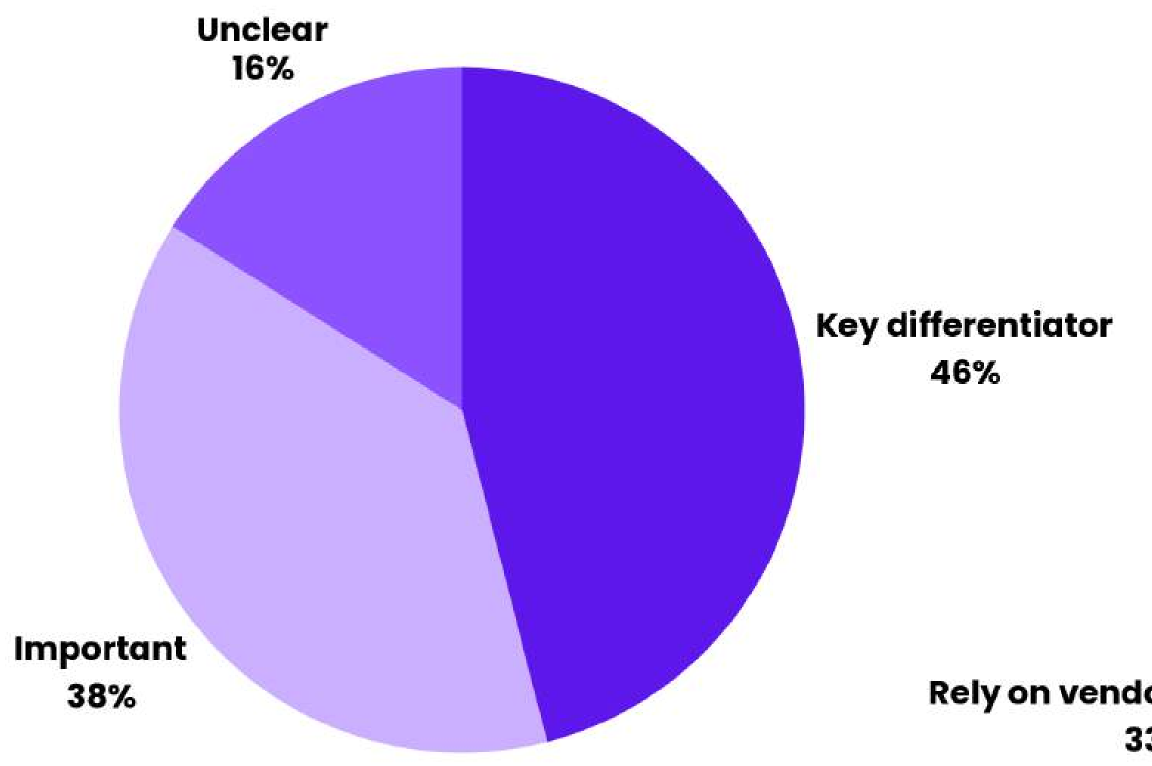
But concerns over AI risks are catching-up

AI risk is now a commercial issue, not just a technical one.

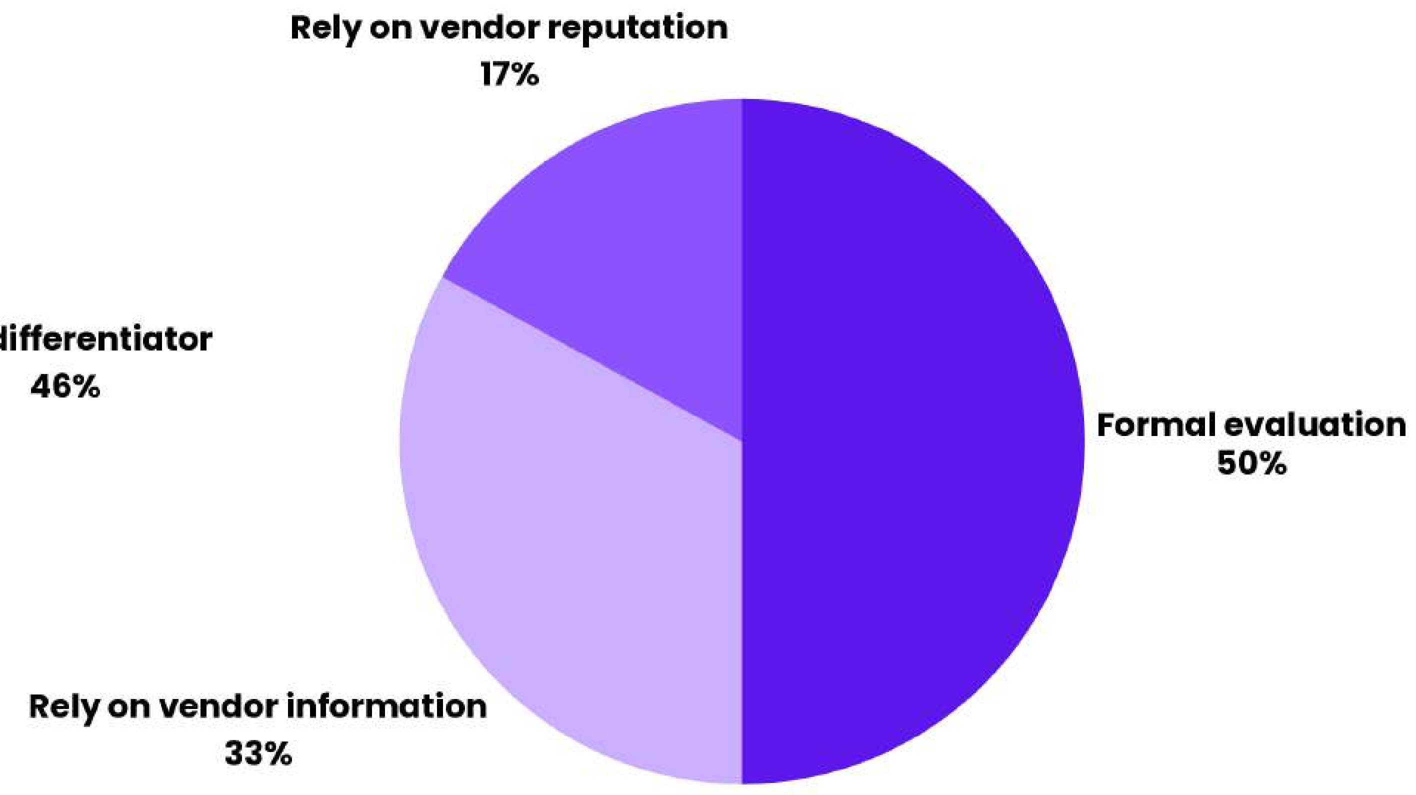
How often are AI risks a key buying factor?



Is Responsible AI a driver of success?



How do HR buyers assess AI systems?



What this means

- There's a strong risk/reward trade-off with AI solutions. Typically, the more value the AI delivers, the more it is likely to be automating of a previously person-powered process, and the more risk there is from reduced human control.
- Organizations procuring AI solutions are waking up to the AI risks and adapting their procurement processes.
- Lawsuits about AI discrimination are gaining traction in court (e.g. Mobley vs. Workday).

Source: State of responsible AI survey, 2025

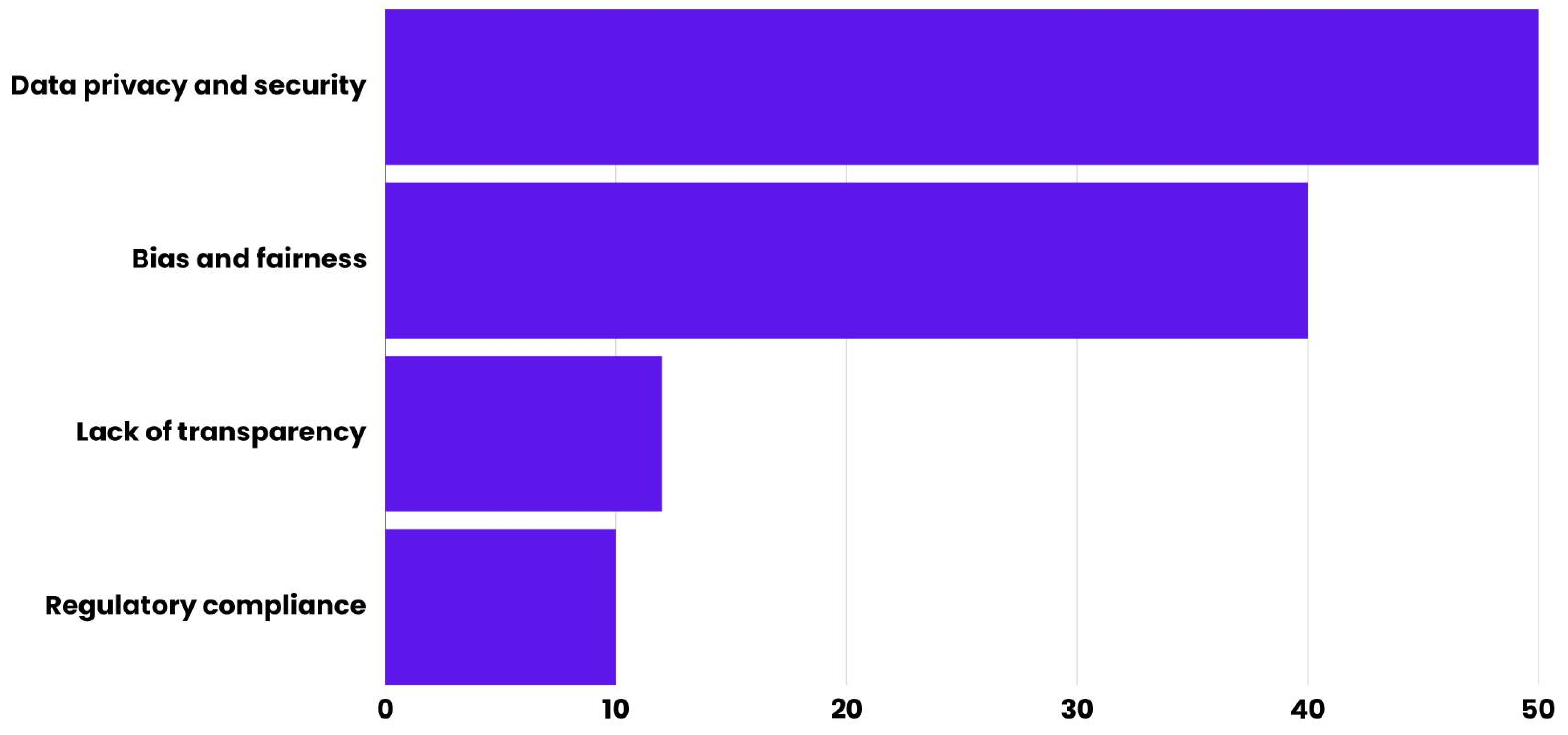
Warden AI

Perception

Bias is one of the top concerns slowing AI adoption

75% of HR leaders name bias as a top concern when evaluating AI, second only to data privacy.

Top concerns for HR leaders considering AI adoption



What this means

- Data security and privacy are long-standing software risks that most organizations have become mature in managing.
- On the other hand, bias is a largely new phenomenon compared to traditional software.
- The TA processes AI is automating inevitably surround sensitive people-oriented situations like progression, so people recognize that the potential for discriminatory bias is high.

Source: State of responsible AI survey, 2025

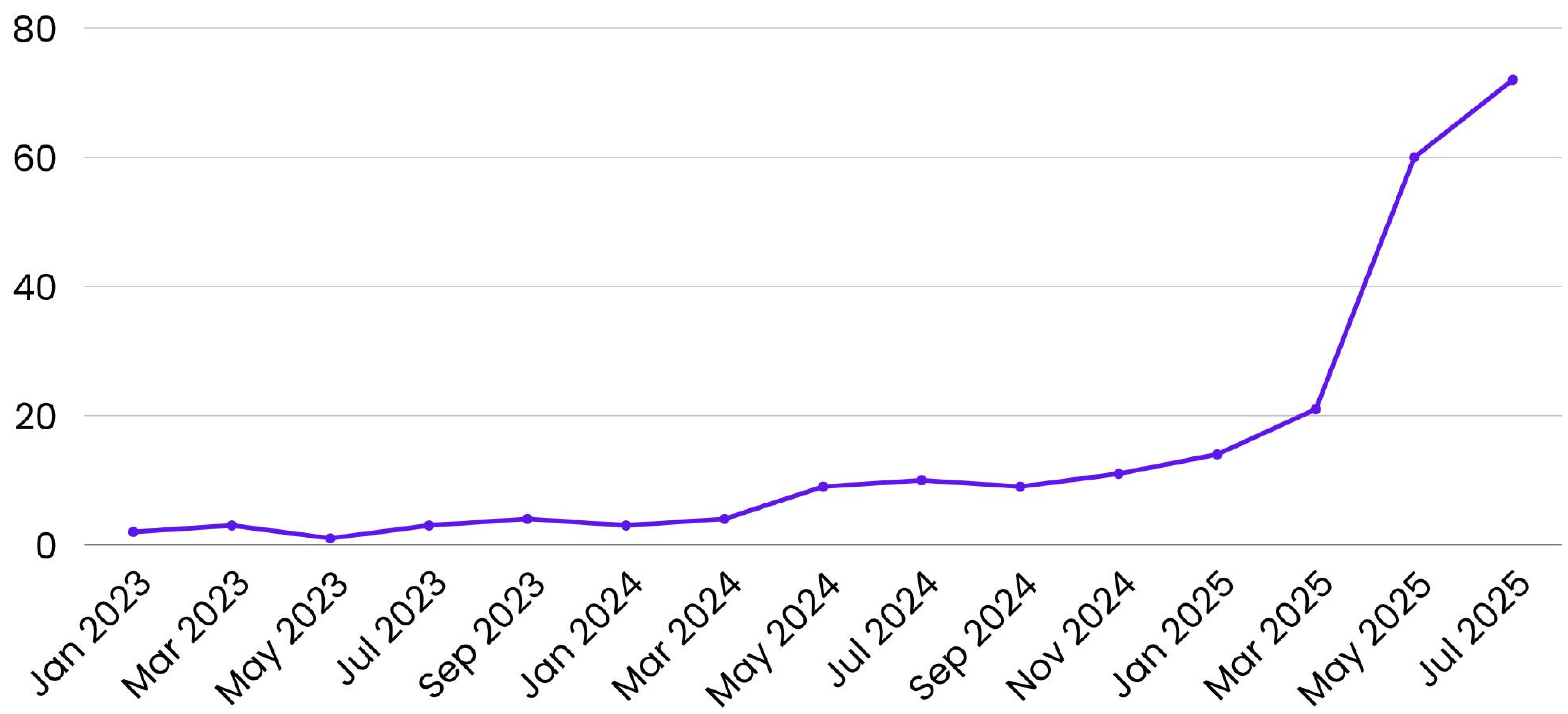
Warden AI

Perception

Mobley vs. Workday is the latest wake-up call

The collective-action ruling of the Workday case is reshaping liability of AI.

Media mentions of Mobley vs. Workday over time



What this means

- Mobley vs. Workday is the first AI lawsuit to progress to collective action status.
- It sets a new precedent of vendors being potentially liable for employment decisions.
- Signals AI bias risks are not theoretical but reputational, legal, and financial.

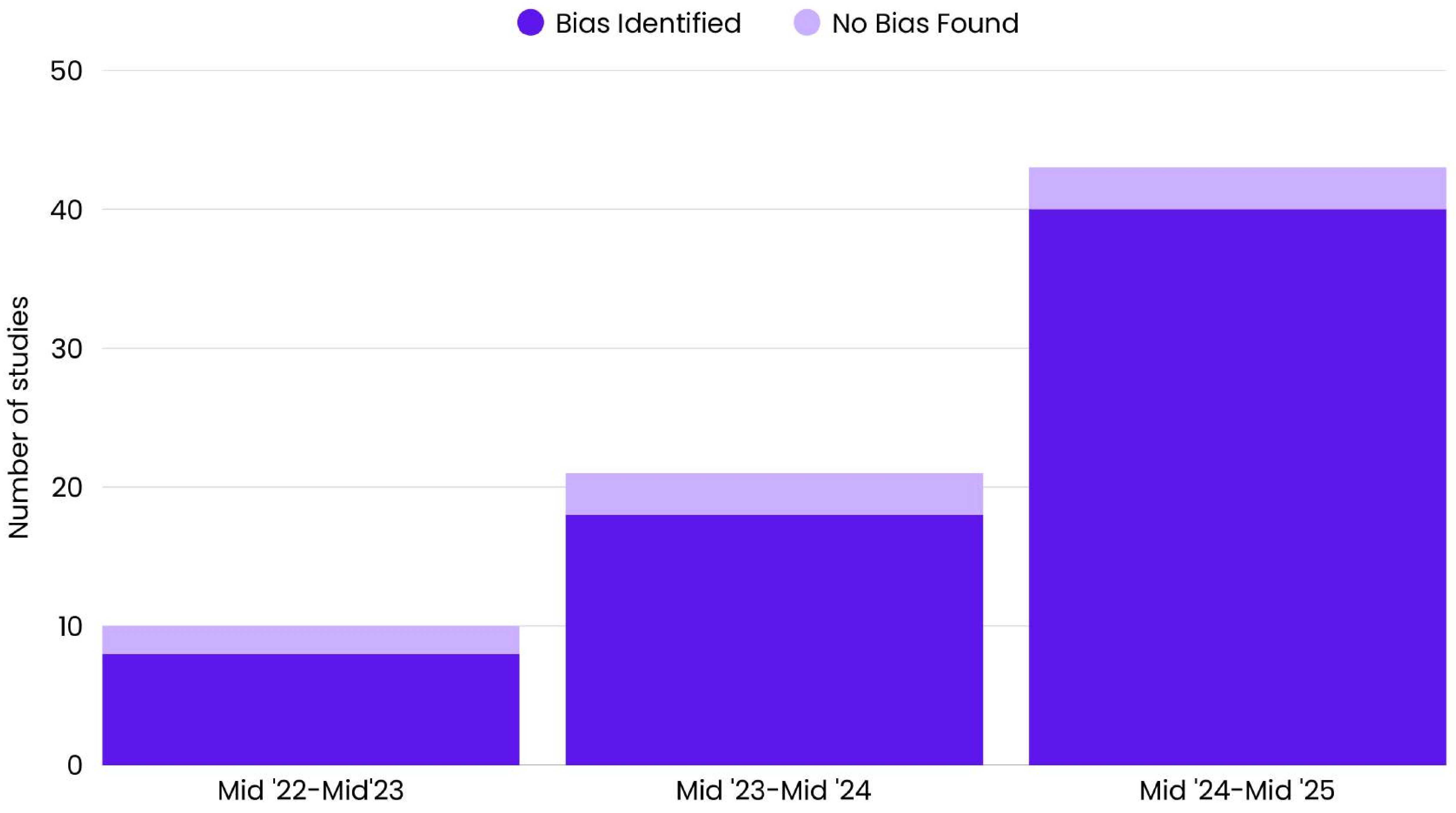
Warden AI

Perception

And a growing body of research finds bias in AI models

New studies are published finding that AI models have inherent biases that discriminate in employment use-cases.

AI bias studies in HR use-cases (past 3 years)



What this means

- Because HR AI models are often trained on historical data, and human decisions are highly biased, such models can often be proven to be biased.
- It indicates the growing awareness of AI bias as an issue in this industry.
- Foundational models gives researchers access in a way they previously wouldn't have when all models were developed in-house.

Warden AI

Perception

But lab results don't tell the full story

Most bias studies don't reflect how AI is used in real-world applications.

Techniques used in academic studies to elicit AI bias

Technique	Description	Why it can be misleading	Example study
Name substitution	Replacing candidate names with ones associated with specific races or genders to signal demographic attributes.	Most real-world AI systems remove names from being shared with AI models.	Bloomberg article on ChatGPT bias
Forced binary choice	Presenting the AI with two nearly identical candidates differing in only one attribute and asking which one should be selected.	Simplifies decision-making to a binary choice, unlike real hiring scenarios where candidates are evaluated individually.	Who Gets the Callback? Generative AI and Gender Bias
Ranking from a pool	Asking the AI to rank a group of candidates together.	Real-world systems typically score candidates individually, so each rating is independent of the overall pool.	First come, first hired? ChatGPT's bias for the first resume it sees
Prompting for reference	Directly asking the AI which candidate is more suitable without a structured evaluation framework.	Real-world systems typically provide defined criteria to evaluate against.	Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval

What this means

- Many studies use simplified or artificial setups to test AI bias, like swapping names or forcing binary choices, that don't reflect how real AI applications operate.
- These studies raise awareness, but they often exaggerate risks or overlook how AI is actually used in structured workflows.
- To meaningfully assess AI fairness, we need testing approaches that mirror real-world systems and decisions.

Bias Benchmarks

How biased are AI systems in practice?

Warden AI

“Just like people, some AIs are fair and some are not, so it’s important that you pick the right ones to work with.”

Sarah Smart
Former Head of TA Product, JPMorgan Chase
Managing Partner, HorizonHuman

How we measure and benchmark AI bias

There are two statistical techniques that capture both group-level and individual-level outcomes.

How we measure bias

	Disparate impact analysis	Counterfactual consistency
What it measures	Whether groups (e.g. male vs female) receive similar outcomes on average	Whether attributes or proxy attributes (e.g. name) impacts the AI's judgement
How it works	Compares selection rates between demographic groups	Swaps variables like name, gendered words, educational institutions or ethnicity in identical profiles and checks for changes
Fairness threshold	≥ 80% selection rate parity between groups (the "four-fifths rule")	≥ 97.5% output consistency after attribute changes
Best at spotting	Systemic gaps in treatment between groups	Individual-level bias in how attributes influence outputs

What this means

- Group-level bias: are outcomes similar across demographic groups like gender or race?
- Individual-level bias: would the same person receive a different result if their name or other traits were changed?
- By combining both methods, we can detect bias that might otherwise go unnoticed from systemic disparities to subtle attribute effects in how decisions are made.

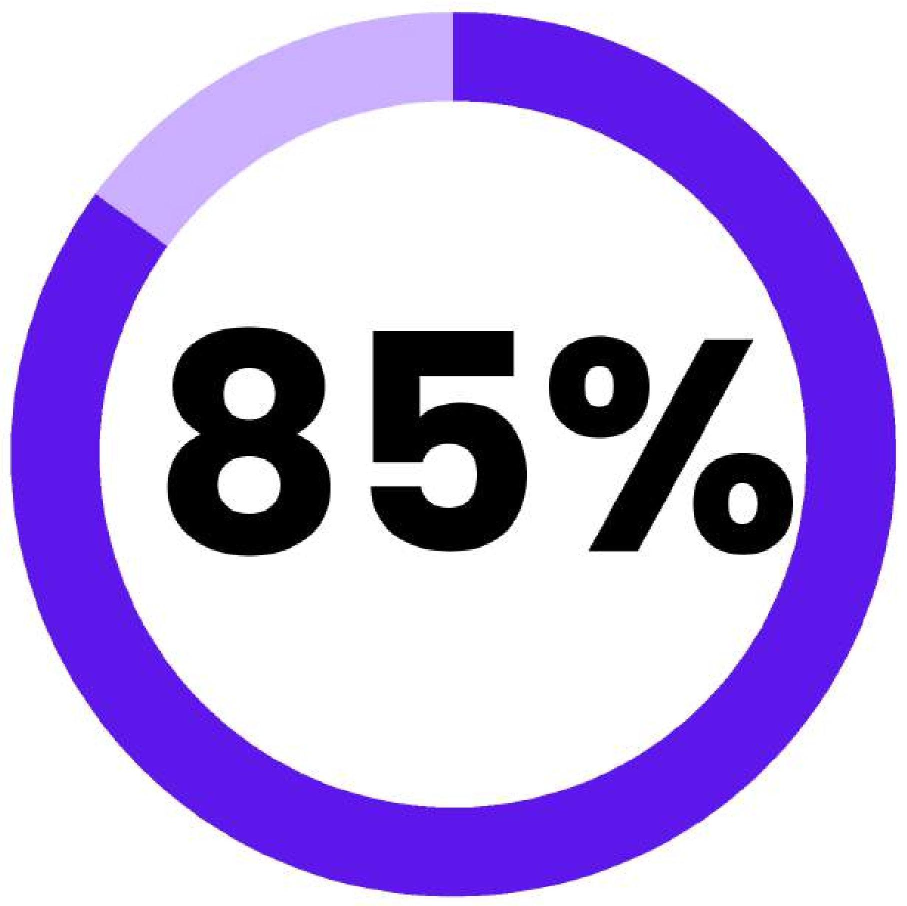
Source: Warden AI methodology

Bias Benchmarks

Most AI systems meet fairness thresholds

85% of AI bias audits showed fairness scores within the “four-fifths rule” threshold across all tested demographic groups.

Disparate Impact



Scored above 0.8 impact ratio for all tested groups

Counterfactual Consistency

Scored above 97.5% consistency for all tested groups

What this means

- Most AI systems tested met accepted fairness thresholds, suggesting bias can be identified and reduced through proper development and oversight.
- High counterfactual consistency scores indicate that many systems treat similar candidates the same, regardless of protected characteristics.
- While not perfect, these results show that fairer AI is not just possible, it’s already happening in many real-world systems.

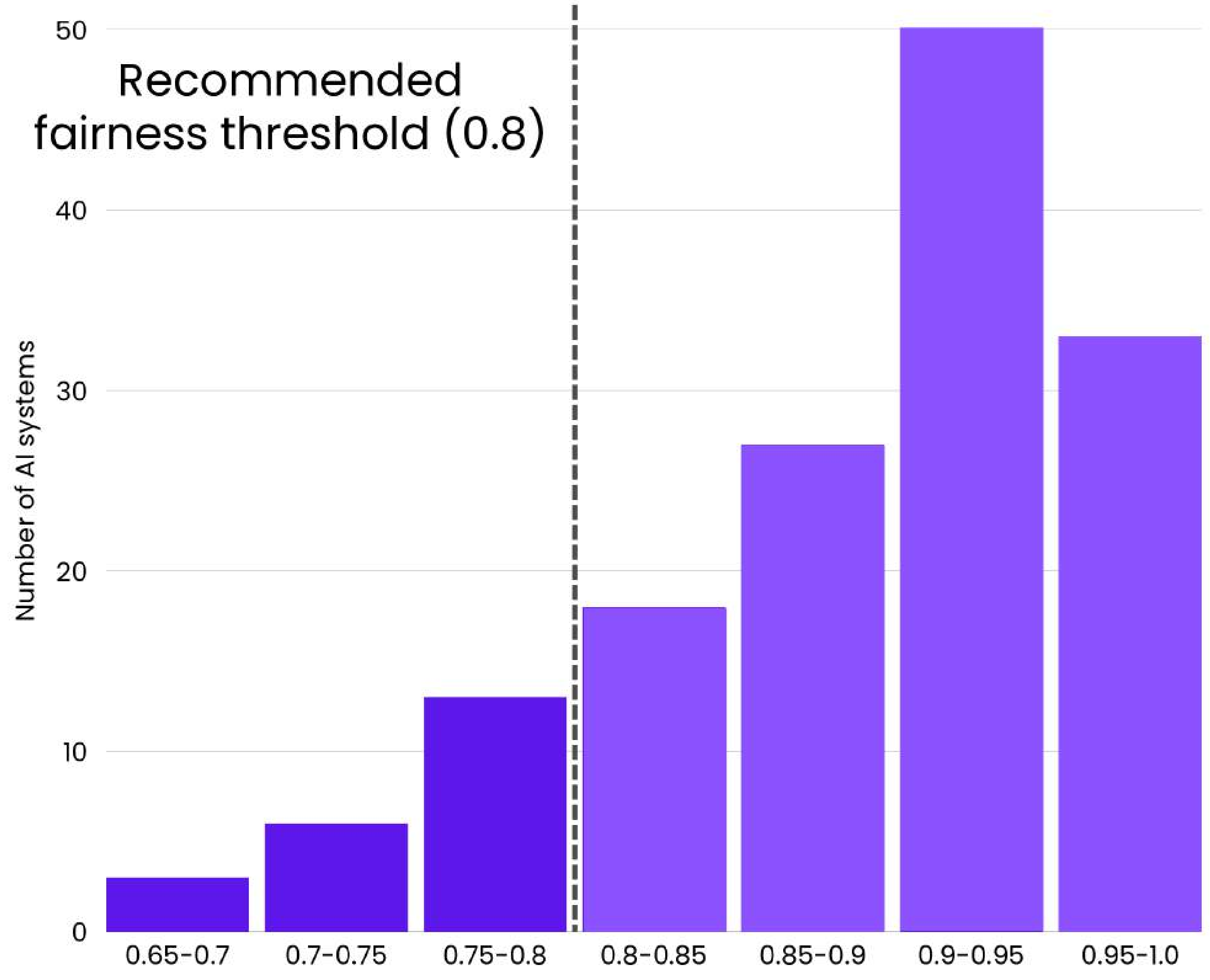
Source: Warden AI customer audit data and published online audits

Bias Benchmarks

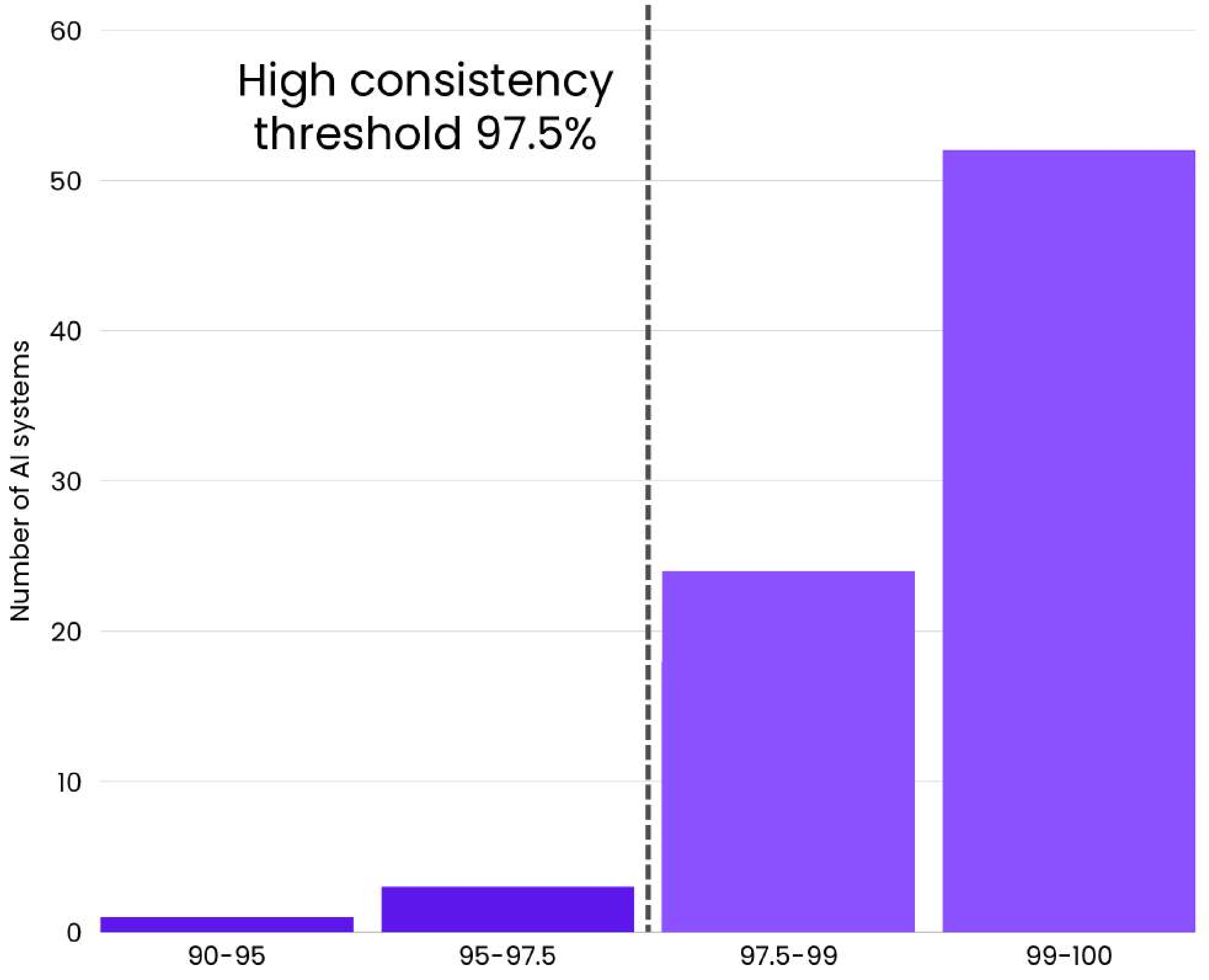
But bias varies widely between systems

Bias metrics across vendors varied by as much as **40%**.

Distribution of scores across AI Systems



Impact ratio range



Counterfactual consistency score (%)

What this means

- While most audited systems meet fairness thresholds, results vary significantly from system to system.
- Even if the industry average looks strong, individual systems may still expose organizations to risk.
- This highlights the importance of choosing AI solutions carefully and monitoring them in deployment.

Source: Warden AI customer audit data and published online audits

Warden AI

Bias Benchmarks

And bias still exists in some systems

15% of AI systems did not meet fairness thresholds for all demographic groups.

Disparate Impact

Scored below 0.8 impact ratio for at least one demographic group

What this means

- A minority of AI systems still show statistically significant bias, failing to meet industry-standard fairness thresholds.
- These cases highlight risks where AI may disproportionately disadvantage certain demographic groups.
- Ongoing testing and oversight are critical to identify and reduce bias before deployment at scale.

Source: Warden AI customer audit data and published online audits

AI vs. Humans

How fair is AI compared to humans?

“We are right to worry about AI bias, but we should not forget that the baseline, human only judgment, is far from bias-free.”

Hung Lee
Recruiting Brainfood

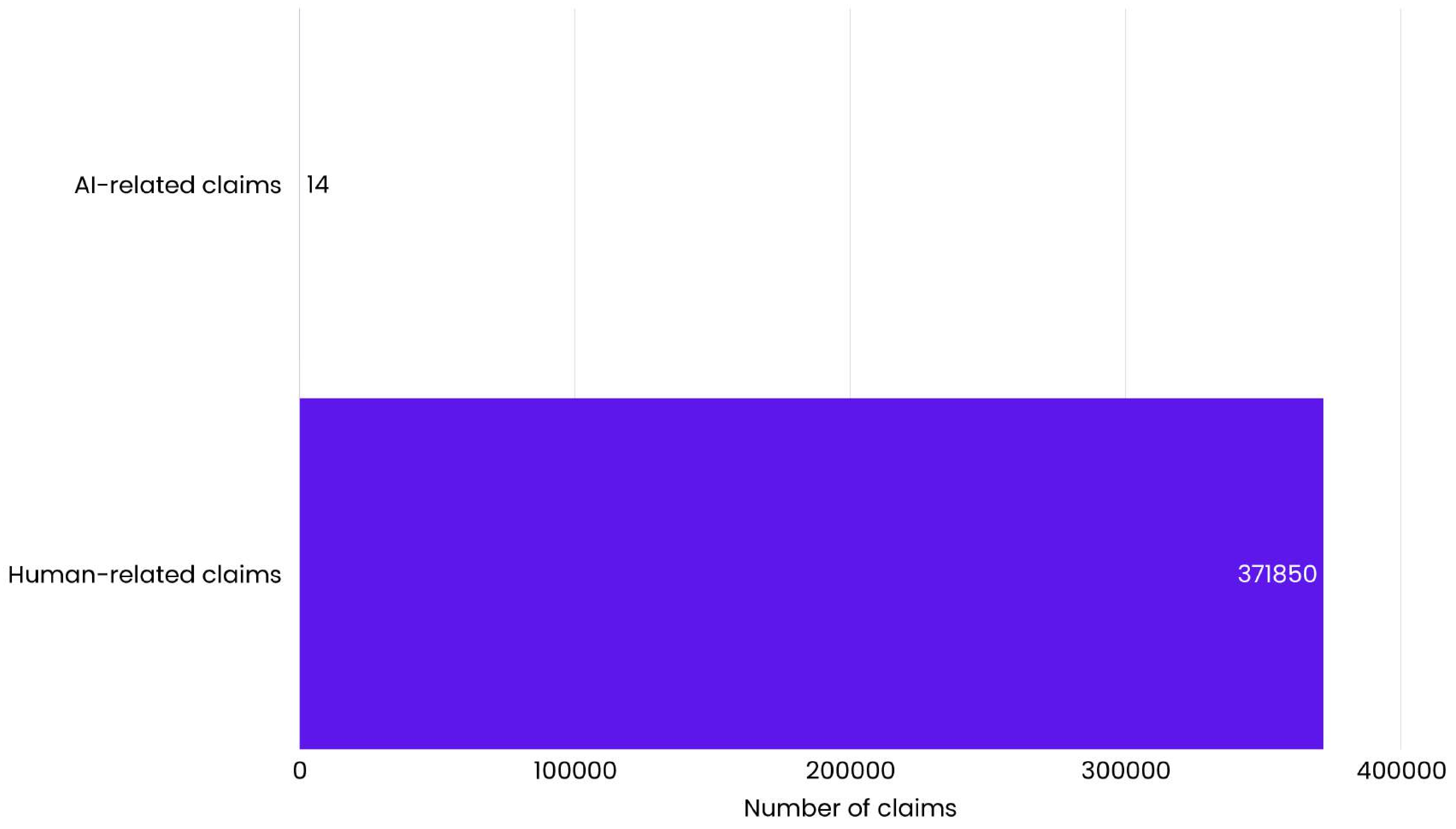
Warden AI

AI vs. Humans

Human bias remains a huge challenge

>99.9% of employment discrimination claims in the last five years have been related to human bias, not AI bias.

Discrimination claims in employment (2020-2024)



What this means

- Each year, around 100,000 employment discrimination claims are submitted in US alone.
- About 100 of these turn into lawsuits.
- The pre-AI status quo is rife with alleged and actual discrimination.

Source: US EEOC Enforcement and Litigation Statistics

Warden AI

AI vs. Humans
How we benchmark human bias

How it works

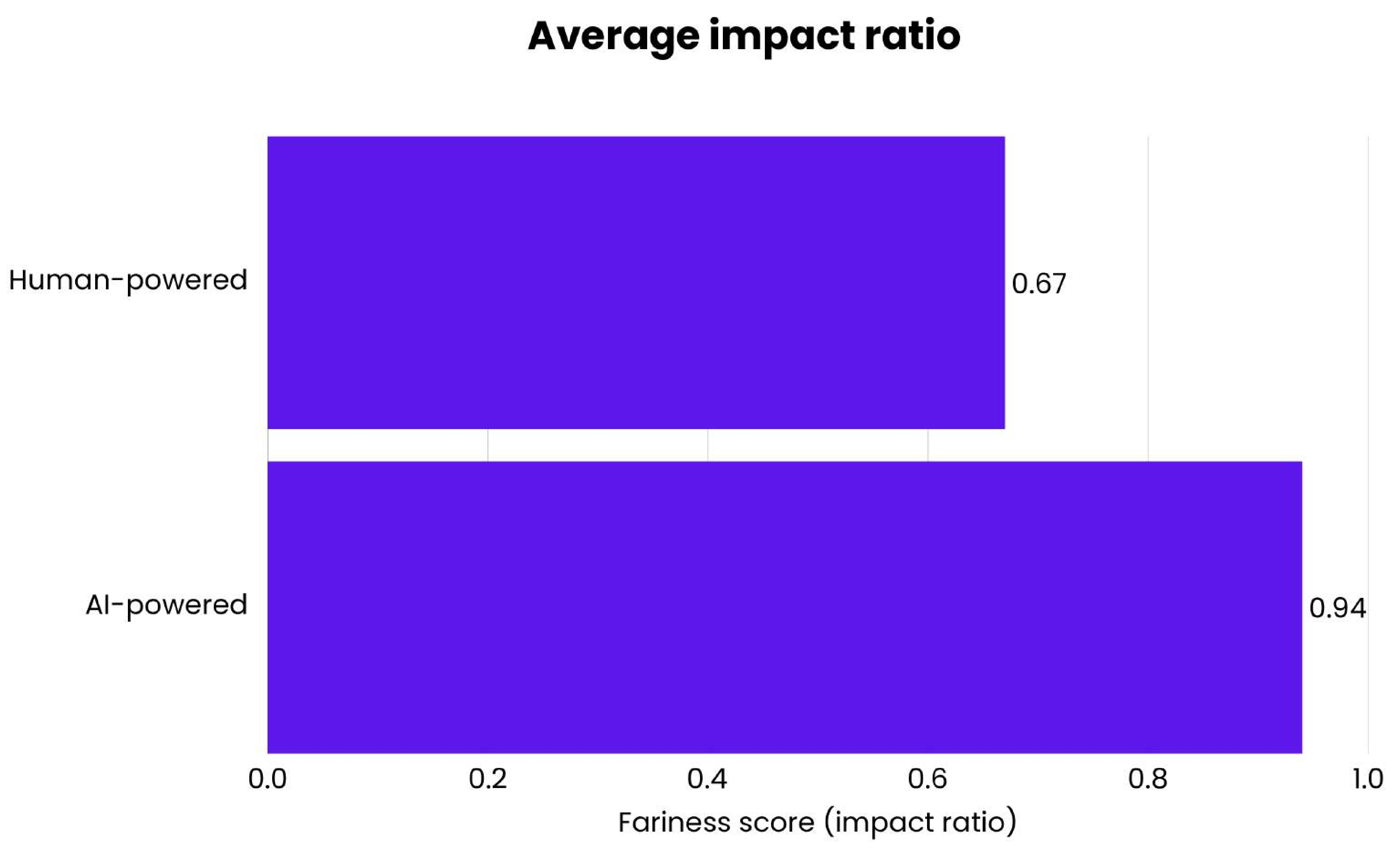
- To compare AI systems with human decision-making, we created a human bias benchmark using data from multiple academic and industry studies examining bias in employment.
- These studies measured how often candidates from different sex and racial groups were selected in real-world hiring.
- We converted these findings into impact ratios, the same fairness metric we apply to AI systems, allowing a direct comparison between human and AI outcomes.

Author	Study name	Focus	Link
Kubiak et al. (2023)	Gender equity in hiring	Gender	Link
González, Cortina & Rodríguez-Menés (2018	The Role of Gender Stereotypes in Hiring	Gender	Link
Quillian & Lee (2023)	Trends in racial and ethnic discrimination in hiring	Race/Ethnicity	Link
Banerjee, Reitz & Oreopoulos (2011)	Do Large Employers Treat Racial Minorities More Fairly?	Race/Ethnicity	Link
Bertrand & Mullainathan (2003)	Are Emily and Greg More Employable than Lakisha and Jamal?	Race/Ethnicity	Link
Adamovic & Leibbrandt (2022)	A large-scale field experiment on occupational gender hiring discrimination	Gender	Link
Goldin & Rouse (2000)	Orchestrating Impartiality	Gender	Link
Weichselbaumer (2015)	Discrimination against Female Migrants Wearing Headscarves	Gender, Ethnicity, Religion	Link
Zschirnt & Ruedin (2016)	Ethnic discrimination in hiring decisions	Race/Ethnicity	Link
Neumark, Bank & Van Nort (1996)	Sex Discrimination in Restaurant Hiring	Gender	Link
Quillian et al. (2017)	Meta-analysis of field experiments	Race/Ethnicity	Link

AI vs. Humans

AI systems show fairer treatment on average

The data suggests that, on average, AI systems treat different demographic groups more fairly than humans do.



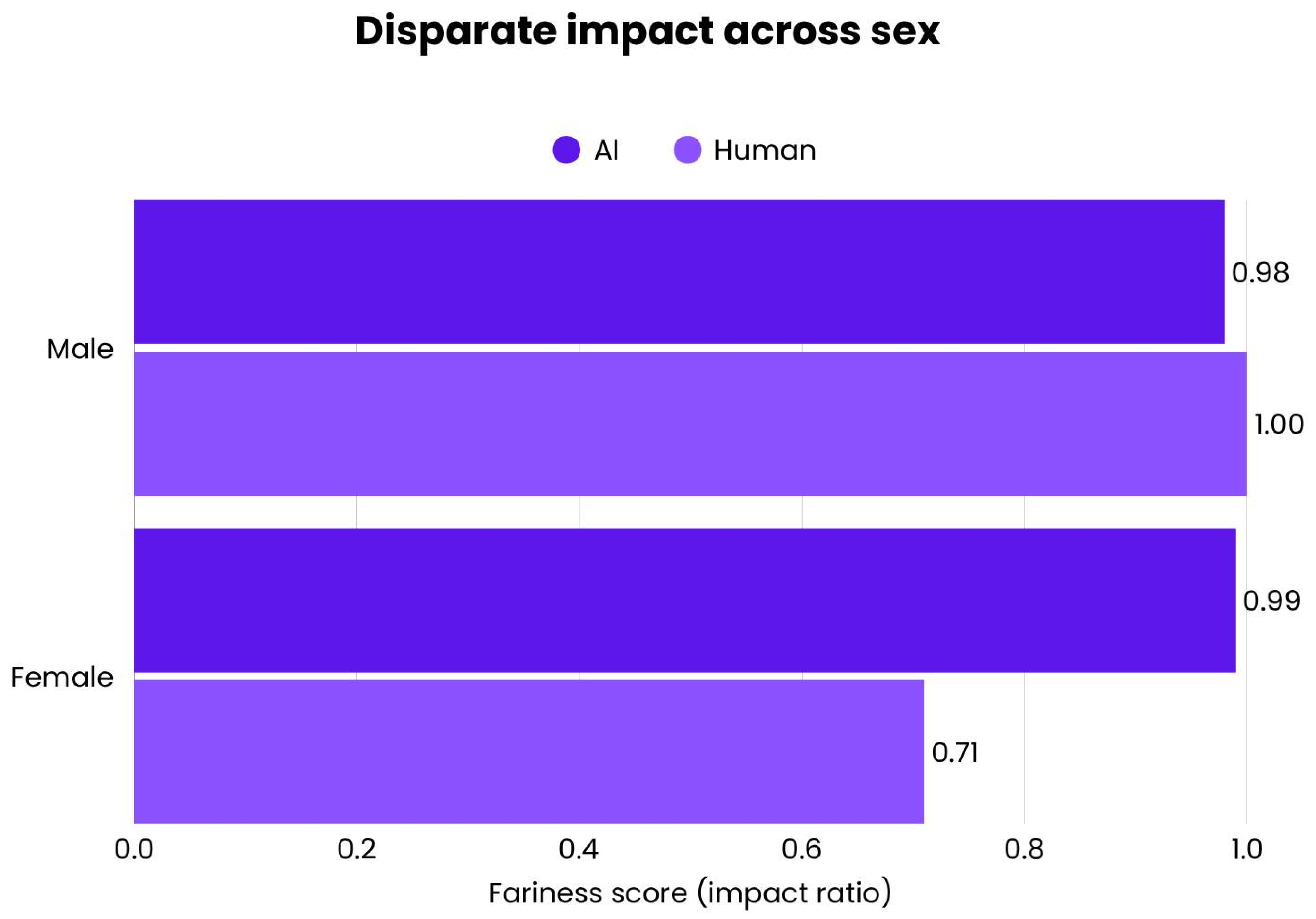
What this means

- While AI systems aren't immune to bias, data suggests they often outperform human decision-makers on average fairness metrics.
- Human decisions, by contrast, tend to reflect longstanding systemic biases, yet are rarely scrutinized to the same degree.
- The data points to a valuable opportunity: rather than avoiding AI for fear of bias, organizations can use it to raise the standard.

Source: Warden AI customer audit data; published online audits; peer-reviewed studies

AI could be closing the gender gap in employment

The data suggests female candidates experience up to **39%** fairer treatment with AI.



What this means

- The data indicates that AI systems may be helping to close the gender gap, showing near-parity in outcomes between men and women, on average.
- In contrast, human-led decisions show a pronounced disparity, with women significantly less likely to experience a positive outcome.
- While not definitive, these findings suggest that AI, when properly developed and deployed, has the potential to support fairer outcomes at scale.

Source: Warden AI customer audit data; published online audits; peer-reviewed studies

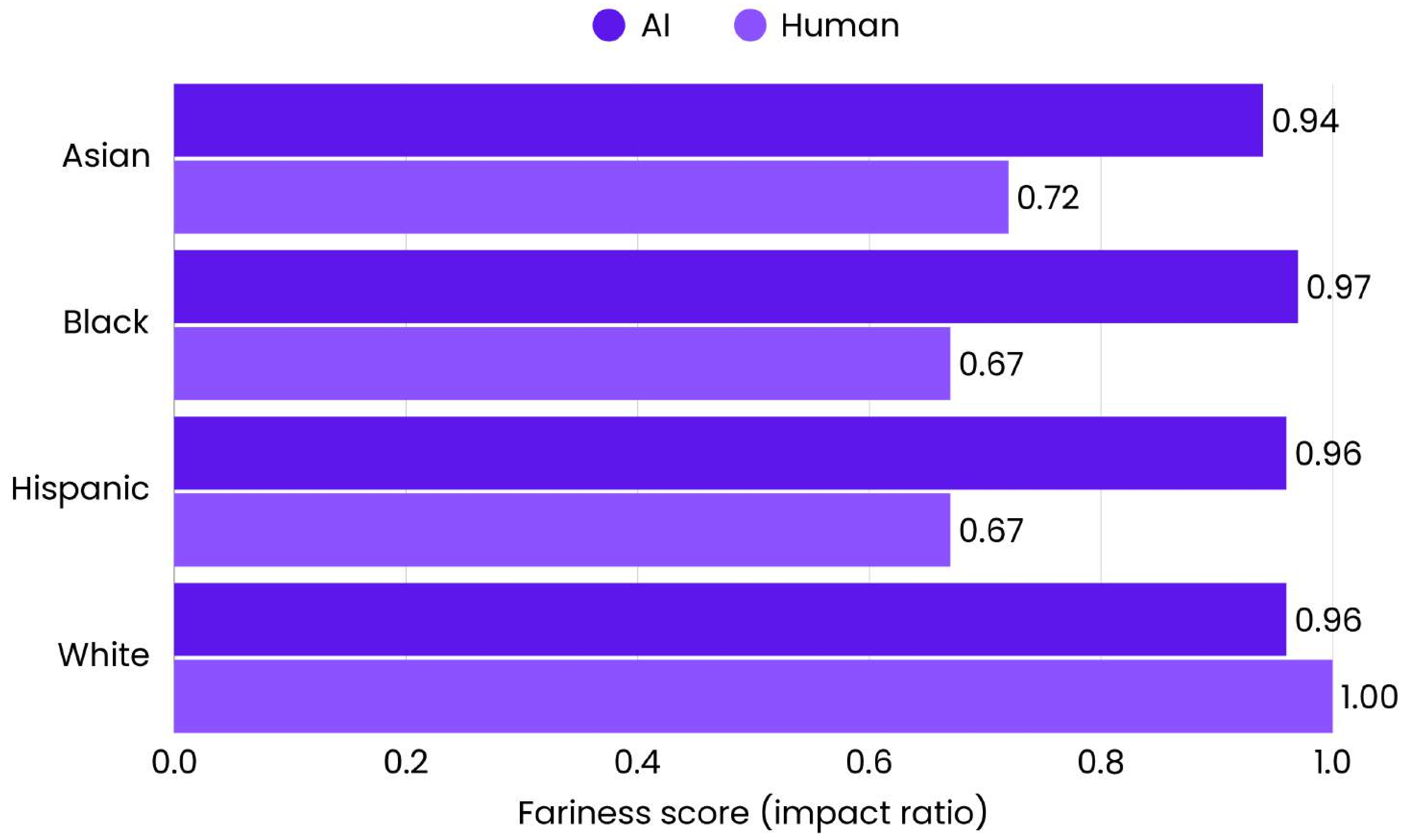
Warden AI

AI vs. Humans

AI could be reducing racial inequalities in employment

The data suggests racial minorities experience up to **45%** fairer treatment with AI.

Disparate impact across race/ethnicity



What this means

- The data indicates that AI systems may be reducing racial disparities, with minority groups seeing fairer outcomes compared to historical human-led decisions.
- In particular, Black and Hispanic candidates show substantially higher fairness scores under AI systems than under human judgment.
- While these findings don't eliminate the need for oversight, they highlight the potential for AI to improve equity when built and used responsibly.

Source: Warden AI customer audit data; published online audits; peer-reviewed studies

AI Compliance

Is the industry ready for AI regulation?

"HR leaders face a patchwork of regulations across regions. Turning this into day-to-day governance is now the differentiator."

Martyn Redstone
Founder,
Eunomia HR

Warden AI

Key AI regulations are emerging for the HR industry

Amid a wave of AI regulations, several have a meaningful impact on the HR/TA space.

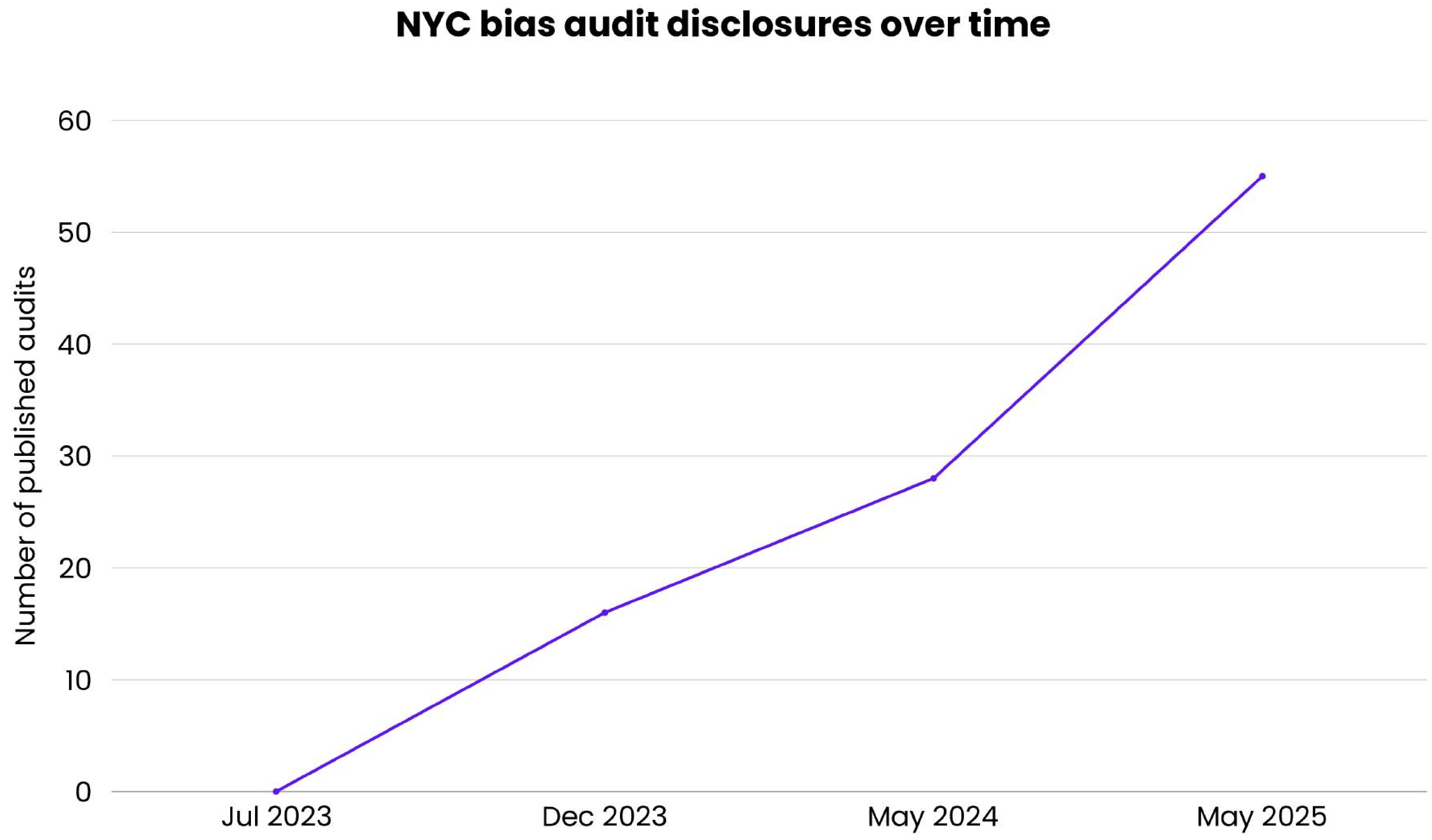
Regulation	Type	Region	In effect	Applies to
NYC Local Law 144	Regulation	US (NYC)	Jan 2023	Deployers (employers using AI tools)
Colorado SB205	Regulation	US (CO)	Feb 2026	Developers and deployers (broad scope)
EU AI Act	Regulation	EU	August 2026	Developers and deployers of high-risk AI
ISO 42001	Standard	Global	2024	Primarily developers (voluntary)

What this means

- Despite the wave of 100s of AI regulations & guidance, only a few have big impact on the HR/TA space.
- These regulations will likely set the requirements that others will follow (e.g. California).
- ISO 42001 is gradually emerging as a standard for process management of AI systems (in the HR space and beyond).

AI Compliance
Compliance lags behind regulation

Two years since NYC LL 144 went into affect and compliance is growing but still lagging behind.



What this means

- After two years of the NYC LL 144 regulation being in effect and compliance is slowly growing despite a slow start.
- The NYC regulation while well-intentioned has been highly flawed and easy for people to circumvent.
- Despite this, organizations are increasingly putting pressure on vendors to demonstrate compliance and slowly the regulation is having the intended effect.

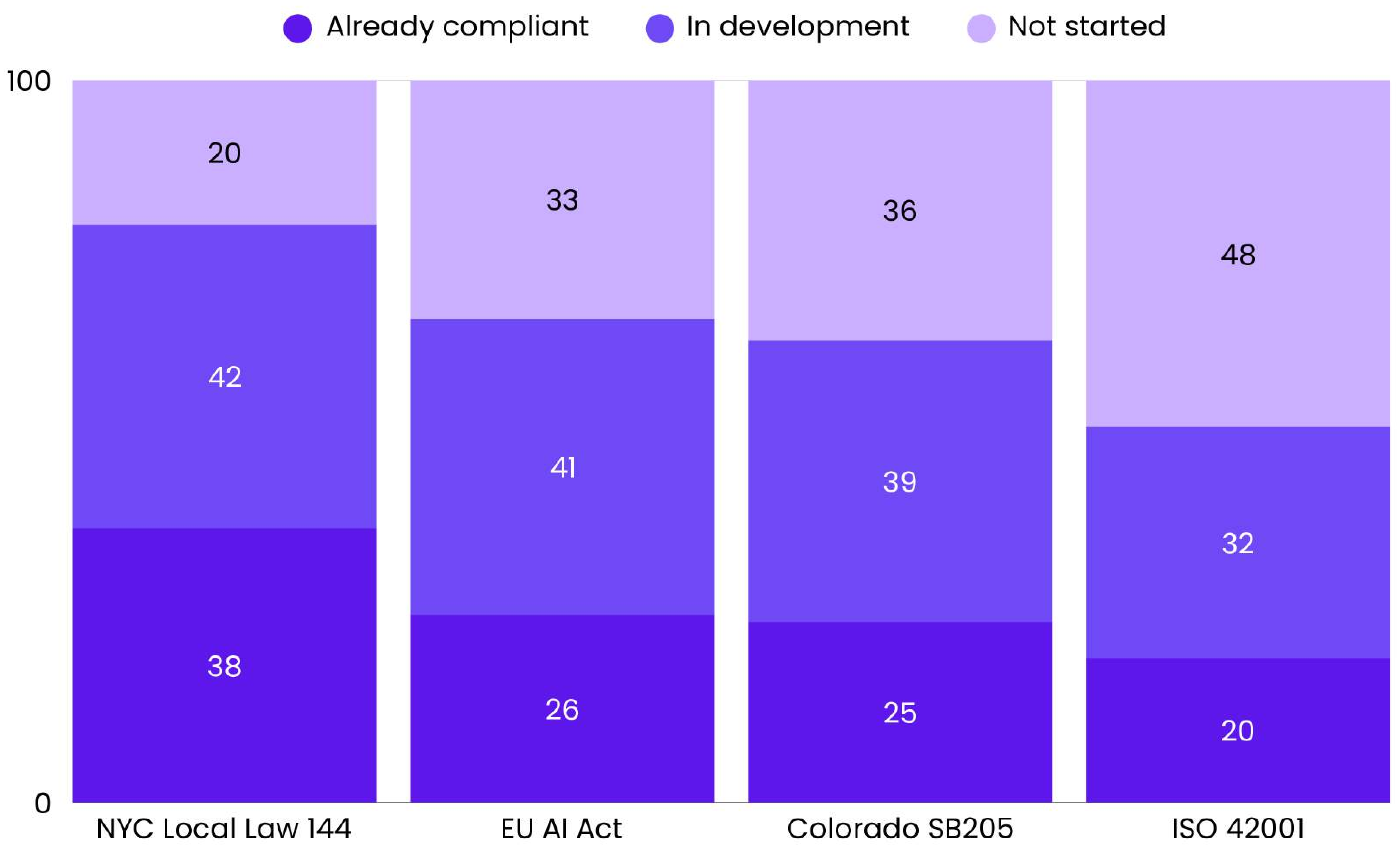
Source: ACLU AEDT Tracking Repository

Warden AI

Many are preparing for regulation but few are fully ready

Many vendors are working on compliance with key regulations, but few are fully ready today.

Vendor status across key AI regulations



What this means

- Despite promising progress on compliance, the majority of vendors are still not compliant with the existing NYC regulation and few are ready for the new ones.
- On the other hand, there has been a clear wake-up call and a large proportion are actively preparing for the coming wave of regulations.

Source: State of responsible AI survey, 2025

Vendor Practices

How mature are vendors on Responsible AI?

“Strong governance frameworks build the confidence that enables rapid AI adoption. This is what turns innovation into execution.”

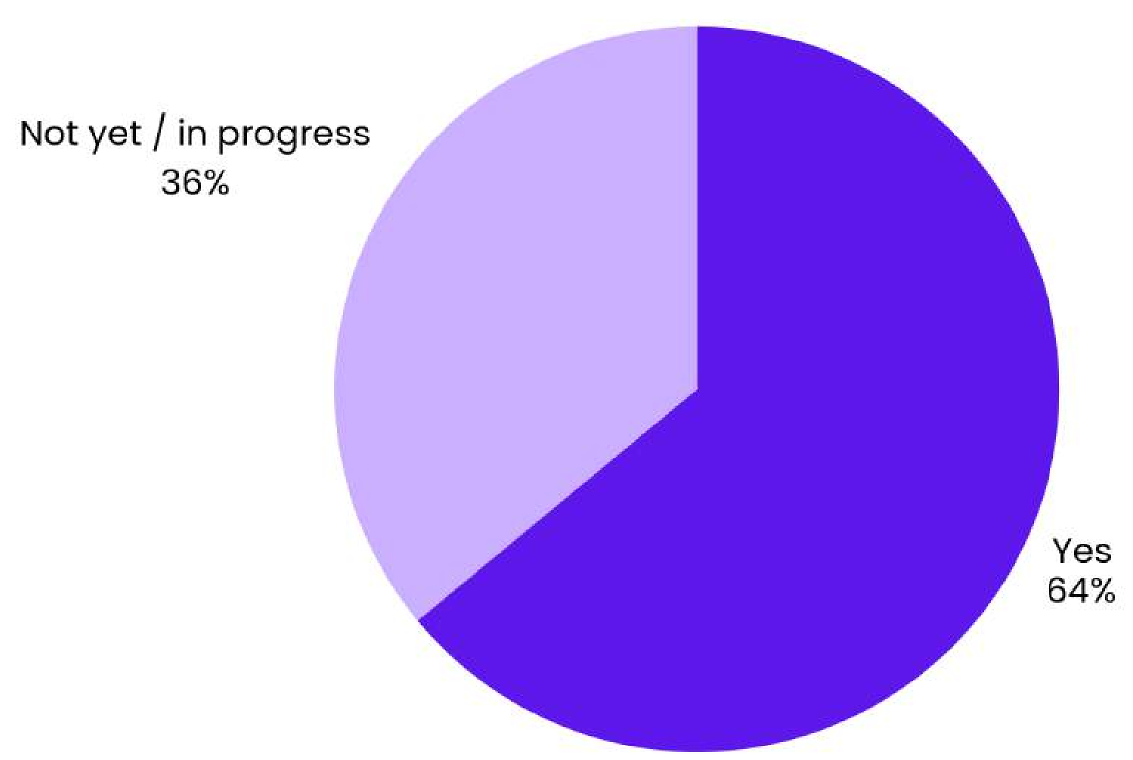
Trent Cotton
Head of TA Insights,
iCIMS

Warden AI

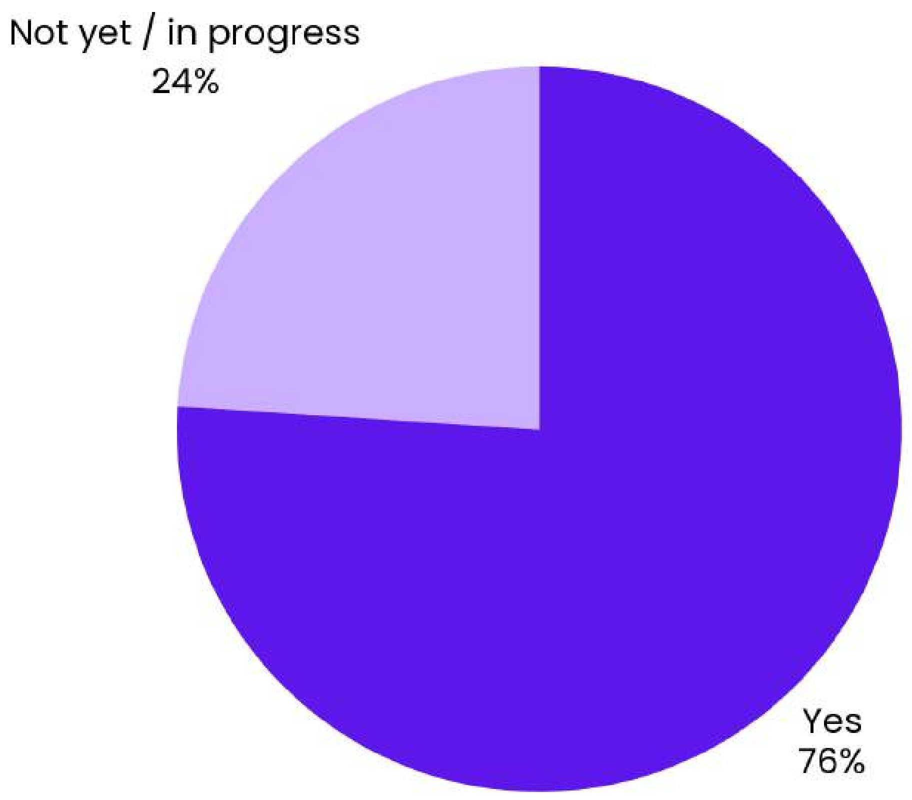
Vendor Practices
AI governance foundations are becoming standard

Most vendors have defined AI principles, implemented an AI policy or framework, and assigned clear ownership for AI governance.

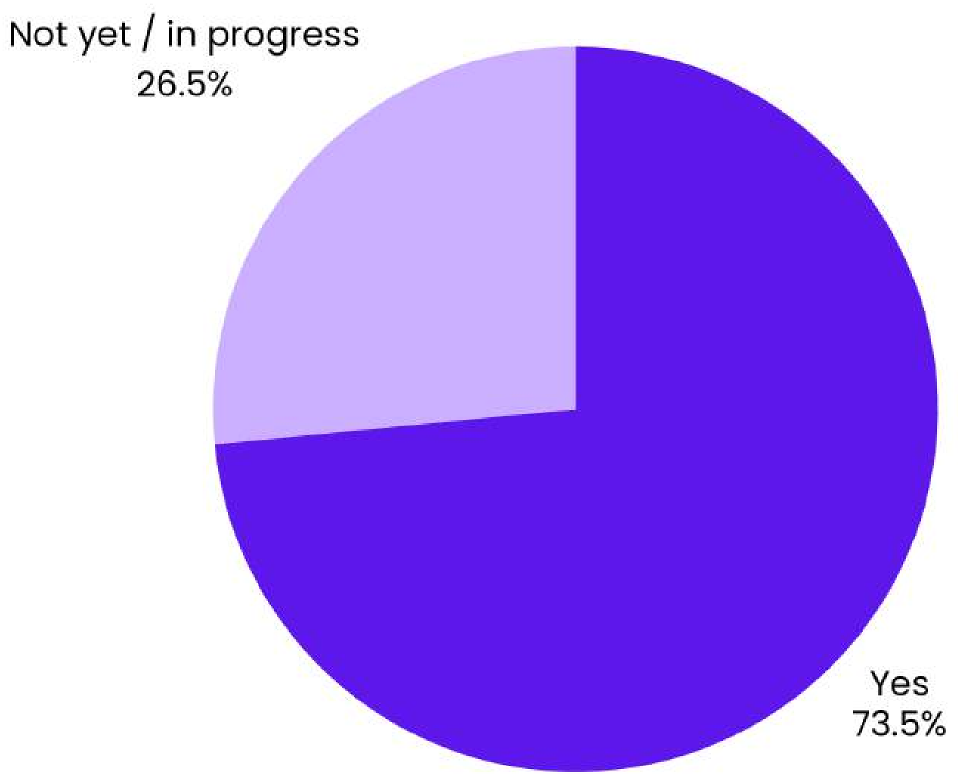
AI principles defined



AI governance policy documented



Named owners for AI governance



What this means

- These governance foundations are a positive sign that Responsible AI is being taken seriously across the ecosystem.
- But they represent the starting point, not the finish line.
- The hard part is turning principles into practice: operationalizing standards, embedding controls, and maintaining accountability in live systems.

Source: State of responsible AI survey, 2025

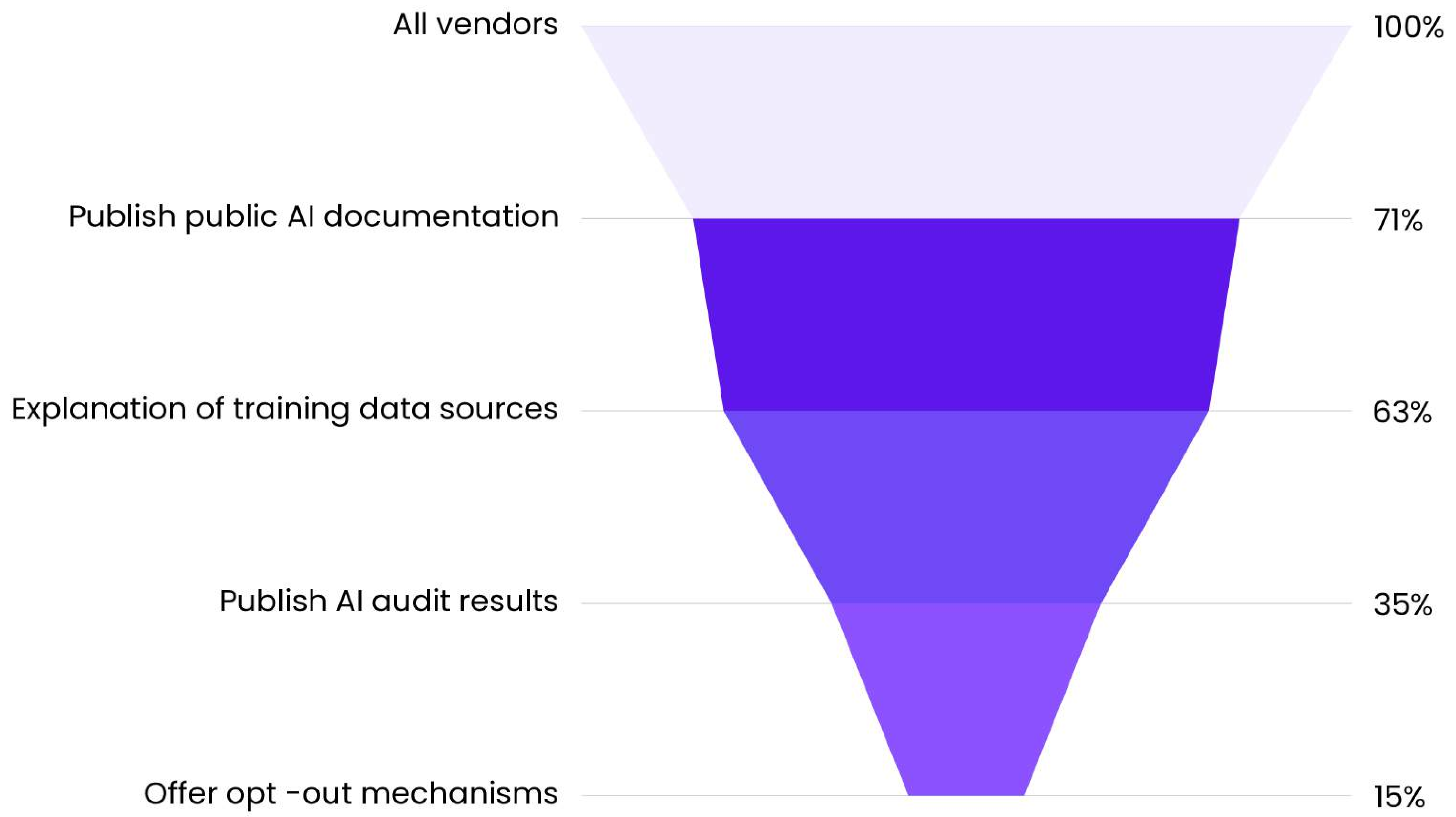
Warden AI

Vendor Practices

There's a promising level of vendor-buyer transparency

Most vendors publish detailed documentation but few offer opt-outs to end-users.

Depth of transparency practices among AI vendors



What this means

- Driven by employer pressure, many vendors are now proactively publishing documentation (e.g. AI explainability statements).
- But deeper transparency practices, like offering opt-outs, remain rare and unevenly adopted.
- To build real confidence with buyers and regulators, transparency needs to move beyond surface-level disclosures toward meaningful visibility and control.

Source: State of responsible AI survey, 2025

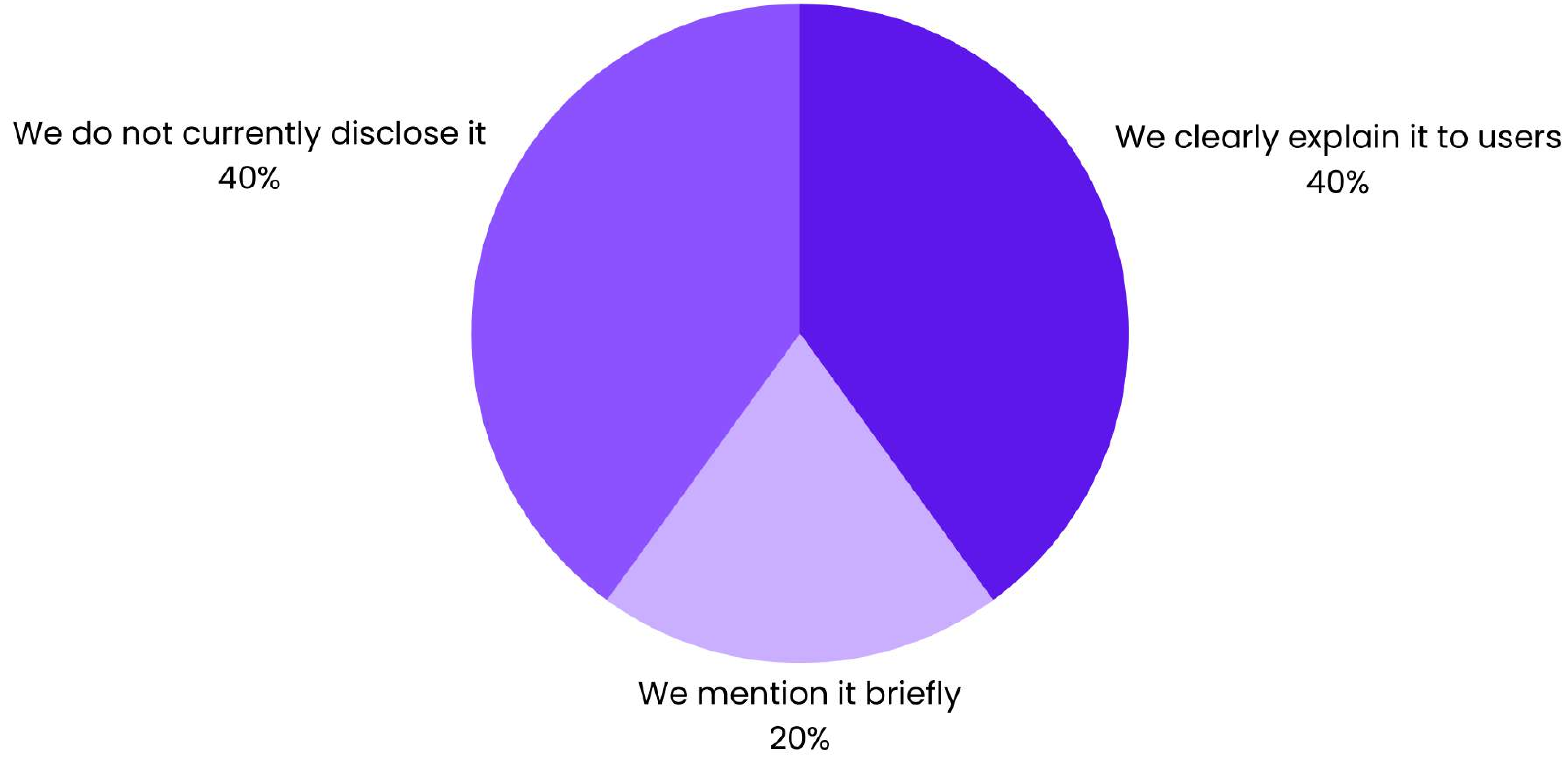
Warden AI

Vendor Practices

But end-users often lack awareness of AI use

Few vendors and HR/TA teams clearly explain how AI is being used, leaving many users unaware that AI plays a role at all.

How transparent are AI/automated TA systems to candidates or employees?



What this means

- This lack of transparency can undermine trust, even when the underlying system is fair or well-governed.
- There's a tension in accountability here, with vendors saying it's up to the organizations deploying the system.
- Many AI regulations explicitly require opt-out rights for human review (e.g. GDPR, EU AI Act, Colorado SB205).

Source: State of responsible AI vendor survey, 2025

Warden AI

“The best way to mitigate risk is to design and deploy AI responsibly from day one.”

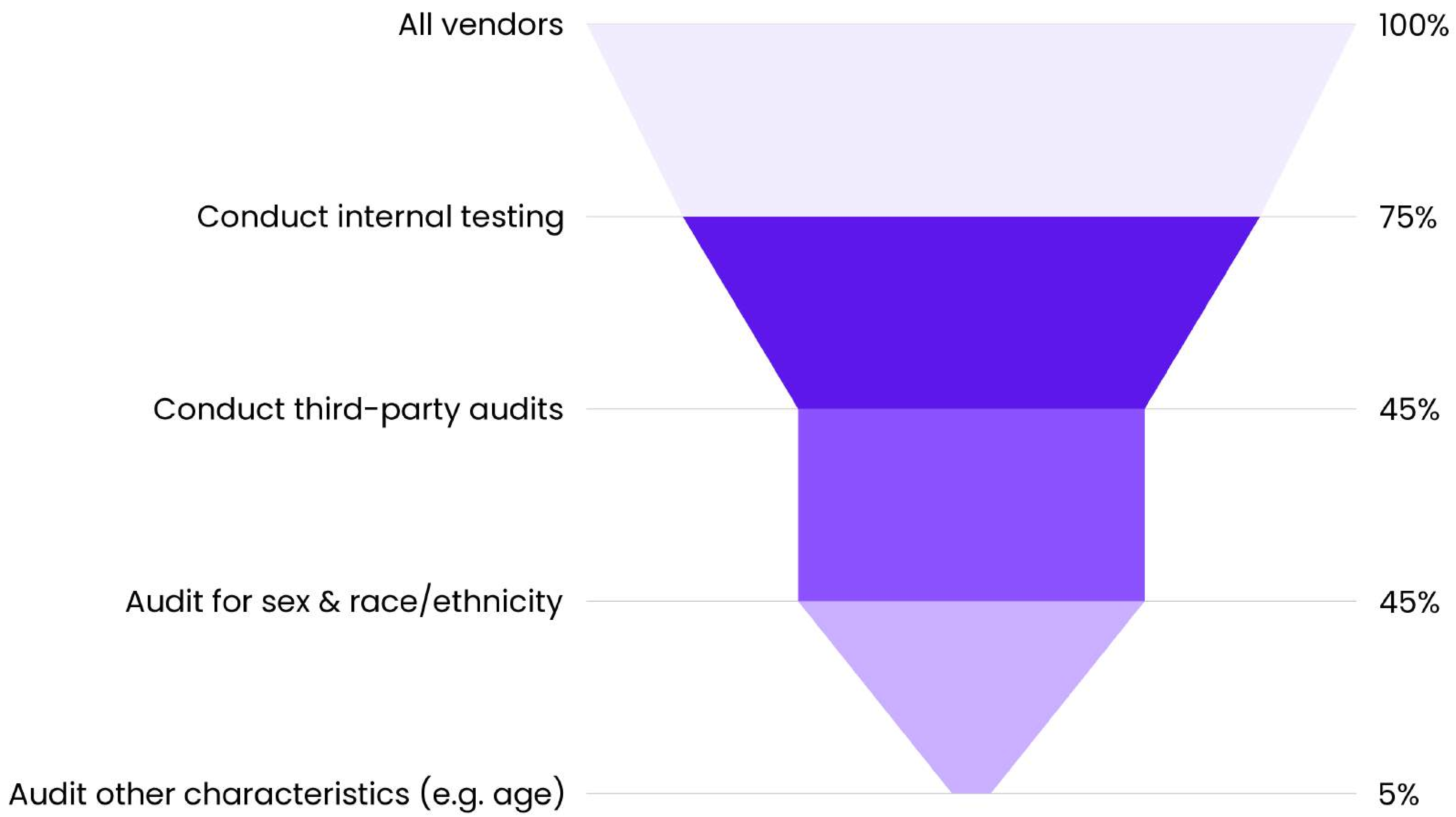
Sultan Saidov
Co-founder and President,
Beamery

Vendor Practices

Bias testing and auditing is becoming standard

Bias testing is now common, at least internally if not with third-party auditing.

Depth of bias evaluation



What this means

- Bias testing is no longer niche with a growing number of vendors doing it.
- In 2025 we predict a shift from about 20% audited vendors to 80%, as a tipping point is reached and it becomes a table stakes expectation.
- This is driven in part by the NYC LL 144 and more significantly by buyer expectations.

Source: State of responsible AI survey, 2025

Warden AI

Vendor Practices

But few audits test beyond sex and race

Only **5%** of bias audits cover other demographic groups like age and disability.

Protected group	% of audits	Common AI audits?	Notes
Sex	100%	✅ Yes	Standard across most audits
Race/ethnicity	100%	✅ Yes	Standard across most audits
Age	5%	⚠️ Rare	Increasingly important (e.g. Mobley case)
Disability	5%	⚠️ Rare	Increasingly important (e.g. Mobley case)
Religion	2%	❌ Very rare	Typically excluded
Sexual orientation	2%	❌ Very rare	Typically excluded
National origin	1%	❌ Very rare	Typically excluded

What this means

- 95% of bias audits focus on the core minimum demographic groups of sex and race/ethnicity.
- This is driven by NYC LL 144 which focuses on those two protected classes.
- Age discrimination is a particularly common risk and often overlooked (e.g. Mobley vs. Workday; iTutorGroup)

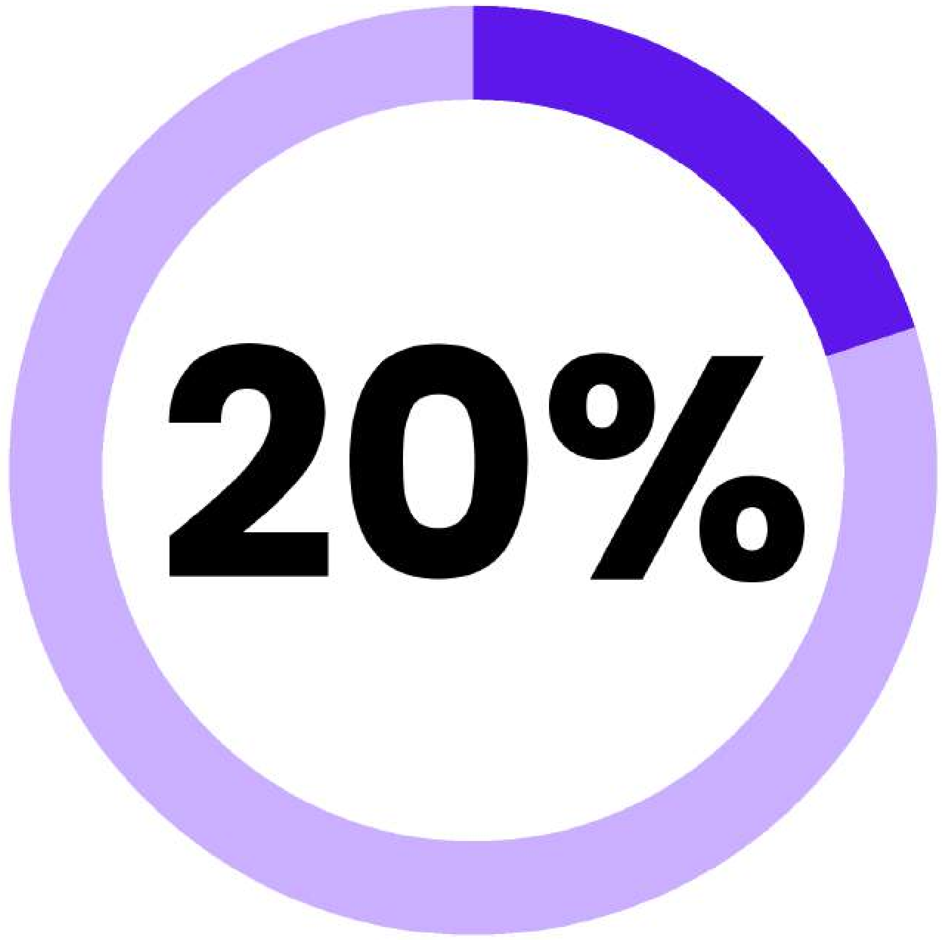
Source: Warden AI customer audit data and published online audits

Warden AI

Vendor Practices

A growing number of vendors are leading the way

20% of vendors follow best-practices for responsible AI development.



of vendors follow all best practices

- Define and document AI governance
- Publish Responsible AI documentation
- Share audit results publicly
- Prepare for AI regulations

What this means

- Not all vendors are equal when it comes to mitigating AI risks and being compliant.
- As scrutiny of AI systems grows, partnering with vendors who meet higher standards will be key to managing risk.
- Buyers should choose vendors who demonstrate a strong posture towards AI risk and compliance.

Source: State of responsible AI survey, 2025

Wrapping it up

Conclusions

The four implications we're seeing for AI use in HR and employment

1

The stakes of getting AI wrong are growing

Mounting regulation, emerging lawsuits, and rising adoption pressure mean the cost of getting it wrong has never been higher.

2

Bias is becoming a shared responsibility

As AI becomes more embedded in TA processes, responsibility for bias is increasingly shared between vendors and the organizations that deploy the tools.

3

AI can improve workplace equality

When designed and deployed responsibly, AI systems can reduce inequality and help close longstanding gaps in workplace outcomes.

4

Vendor choice is now a key risk decision

As AI systems take on higher-stakes decisions, vendor selection now directly shapes legal and reputational risk within the organization.

Methodology

Data sources

Where is our data from?

150+ AI audits using structured test datasets, covering tools across multiple TA use-cases such as screening, sourcing, and interviewing.

Over 1 million test samples, drawn from a combination of proprietary and public datasets.

50+ vendor/practitioner survey responses across TA Tech.

100+ publicly available audit and transparency reports from vendor websites.

Bias audit methodology

Disparate impact

Measures whether outcomes differ across demographic groups (e.g. male vs. female). Systems with any group scoring <0.8 relative to the top group are flagged as potentially biased (per the "four-fifths rule").

Counterfactual consistency

Swaps protected characteristics (e.g. gender, ethnicity) in otherwise identical profiles to check if outcomes change. Systems with <97.5% outcome stability are flagged as potentially biased.

Human bias benchmark

How did we derive the human benchmark?

We compiled a benchmark of "typical human bias" from 10 academic and industry studies that analyzed real-world outcomes (e.g. interview callbacks, progression decisions). These results were converted into the same fairness metrics (impact ratios) to allow direct comparison with AI systems.

About Warden AI

Warden is the platform enabling HR to adopt AI with confidence.

We audit and certify AI systems for fairness and compliance, helping vendors and enterprises as they develop, deploy, and defend AI solutions in HR.

- ✔ Continuously audit AI systems to mitigate risks in real-time.
- ✔ Audit all protected classes and comply with Colorado SB205 with Warden's real-world dataset.
- ✔ Guard against legal claims with evidence-based audit trails.
- ✔ Win stakeholder buy-in with real-time dashboards and reports.

Warden AI