WP-026 | June 2026
Zero-Knowledge AI Governance
affix-io.com

WHITEPAPER WP-026 / US STATE AI REGULATION

Colorado, Illinois, and the New US State AI Laws: Building ZK Compliance Architecture Before Enforcement Begins

Colorado SB 24-205 took effect June 2026. Illinois HB 3773 is close behind. Here is how to prove algorithmic compliance without opening your model to auditors.

AffixIO Research | Published June 2026 | Related: WP-015, WP-020, WP-001, WP-011

ABSTRACT

The United States does not yet have a comprehensive federal AI law, but the states are not waiting. Colorado SB 24-205 (the Colorado Artificial Intelligence Act) is the most ambitious state-level AI regulation yet enacted, imposing impact assessment, consumer notification, and annual certification obligations on developers and deployers of high-risk AI systems. Illinois HB 3773 adds algorithmic discrimination protections in employment. At least twelve other states have enacted or are advancing comparable legislation in 2025 and 2026, creating an emerging patchwork that recalls the early years of US state privacy law. This paper sets out what each major framework requires, examines the

technical challenges of demonstrating algorithmic fairness, and explains how zero-knowledge proof infrastructure can support compliance: generating tamper-resistant, independently verifiable evidence of fairness testing results and consequential decision records without exposing model weights, training data, or individual user information.

CONTENTS

SECTION 01

US States Move First on AI Regulation

For several years, US federal AI regulation moved slowly while the European Union passed the AI Act and began building out implementing regulations. The federal picture remains incomplete as of mid-2026, but the gap is now closing at the state level, and at a pace that many organisations have underestimated.

Colorado SB 24-205, known as the Colorado Artificial Intelligence Act, became effective 1 February 2026. It is the most comprehensive AI regulation enacted by any US state to date, covering both developers and deployers of high-risk AI systems and establishing obligations that range from impact assessments before deployment to annual certifications filed with the Colorado Attorney General. The scope is not limited to large technology companies: any organisation that deploys an AI system making consequential decisions affecting Colorado residents is potentially in scope.

Illinois HB 3773 adds algorithmic discrimination protections layered on top of the existing Illinois Human Rights Act, targeting the specific context of employment decisions made or assisted by automated systems. Employers using AI in hiring, promotion, or termination are required to notify affected individuals, conduct annual bias audits, and publish the results. These requirements build on Illinois's established track record as an early mover in technology-specific employment law, including the Artificial Intelligence Video Interview Act enacted in 2020.

At least twelve other US states enacted or actively advanced AI legislation in 2025 and 2026. Texas, Virginia, Connecticut, Tennessee, and California are among those with bills at various stages of the legislative process, covering overlapping but non-identical categories of AI risk and imposing overlapping but non-identical obligations. Companies operating at scale across multiple US states now face a compliance landscape that closely mirrors the early CCPA period, when organisations first realised that US privacy law was no longer a single federal question but an expanding matrix of state requirements, each with its own definitions, thresholds, and enforcement timelines.

This paper focuses on the specific compliance obligations that are now active or imminent at the state level, the technical challenges of demonstrating algorithmic fairness in a way that satisfies those obligations, and the role that zero-knowledge proof infrastructure can play in generating durable, regulator-readable evidence. Readers looking for coverage of EU AI Act compliance are directed to earlier papers in this series, particularly WP-007, WP-010, WP-015, and WP-020.

SECTION 02

Colorado SB 24-205: The Key Requirements

Colorado SB 24-205 applies to two distinct categories of actor: **developers** of high-risk AI systems and **deployers** of high-risk AI systems. The distinction matters because the obligations are different, but both are substantial.

A **high-risk AI system** under the Act is one that makes, or substantially assists in making, consequential decisions. Consequential decisions are explicitly defined to include determinations related to education enrolment or admission, employment or employment opportunity, a financial or lending service, an essential government service, healthcare, housing, insurance, and legal services. The list is broad by design, and the inclusion of "substantially assists" means that AI systems used to generate recommendations or scores that feed into human decisions are likely in scope even if a human makes the final determination.

Developer Obligations

Developers must use reasonable care to protect consumers from known or reasonably foreseeable risks of algorithmic discrimination. In practice, this translates into a set of documentation and disclosure obligations: developers must be able to provide deployers with meaningful information about the AI system's training data sources, known limitations, evaluation results, and intended use cases. A developer that sells or licenses a high-risk AI system to deployers without this documentation is potentially in breach, and the deployer's compliance will depend on having received it.

The "reasonable care" standard for developers also implies a bias testing programme with documented results. An organisation cannot credibly claim to have used reasonable care to protect against algorithmic discrimination if it has not tested the system for discriminatory outcomes across protected demographic groups. The Act does not specify a particular testing methodology, but the reference to NIST AI RMF as a relevant standard for risk management suggests that the Colorado Attorney General's office will look for structured, documented programmes rather than ad hoc testing.

Deployer Obligations

Deployers face the most operationally demanding requirements. Before deploying a high-risk AI system, a deployer must conduct an **algorithmic impact assessment**: a documented analysis of the system's reasonably foreseeable impacts on consumers, including the potential for algorithmic discrimination. This assessment must be updated whenever the system changes in a material way.

Deployers must also implement a risk management policy, establish a process by which affected consumers can be notified that an AI system was involved in a consequential decision, and provide a mechanism for consumers to appeal or seek correction of AI-driven decisions. The annual certification requirement is perhaps the most novel element: deployers must certify compliance to the Colorado Attorney General each year, providing a summary of impact assessments conducted and any known incidents of algorithmic discrimination.

> The Act defines algorithmic discrimination as using an AI system in a way that results in unlawful differential treatment based on protected characteristics, including race, colour, national origin, sex, age, disability, religion, or genetic information. Notably, discrimination does not require intent; a model that produces disparate outcomes across protected groups may create liability even if the protected characteristic is not an explicit input.

SECTION 03

Illinois HB 3773 and the Broader State AI Landscape

Illinois HB 3773 amends the Illinois Human Rights Act to prohibit the use of automated decision-making systems to discriminate in employment on the basis of protected characteristics. The bill targets a specific and practically important domain: AI systems used in hiring, promotion, and termination

decisions. For employers using AI-assisted recruitment or performance management tools, HB 3773 adds obligations on top of existing federal anti-discrimination law.

The key operational requirements are notification, audit, and transparency. Employers must notify applicants and employees when an automated system is used in a decision affecting them. They must arrange an annual bias audit conducted by an independent auditor, and publish a summary of the audit results. The publication requirement is significant: it creates a public record that consumer advocates, regulators, and job applicants can examine, and it creates reputational risk for employers whose audit results reveal material disparities.

Beyond Colorado and Illinois, the state AI law landscape is evolving rapidly. The table below summarises the major active or advanced state frameworks as of mid-2026.

STATE / LAW	STATUS (MID-2026)	COVERED DECISIONS	KEY OBLIGATIONS
Colorado SB 24-205	In force from Feb 2026	Employment, education, lending, housing, insurance, healthcare, criminal justice	Impact assessments; consumer notification; appeal process; annual AG certification
Illinois HB 3773	Enacted; implementation 2026	Employment (hiring, promotion, termination)	Employee/applicant notification; annual independent bias audit; publication of audit summary
Connecticut SB 2	Advancing in legislature	High-impact automated decisions broadly defined	Developer disclosure obligations; deployer impact assessments; consumer rights
Texas SB 2033	Under consideration	Consequential decisions by high-risk AI	Risk management; bias testing; consumer notification
Virginia HB 2355	Under consideration	High-risk AI in consumer-facing contexts	Impact assessments; transparency reports; consumer appeal rights
Tennessee INSPIRE Act	Enacted 2025	AI-generated content and voice cloning	Disclosure of AI-generated content; consent requirements for voice/likeness

The compliance challenge for organisations operating across multiple US states is not simply the volume of requirements but their heterogeneity. Colorado's definition of a "high-risk AI system" and Illinois's "automated decision-making system" are similar in spirit but differ in their precise boundaries. Connecticut's proposed framework draws on different terminology again. A company that designs its compliance programme around one state's definitions may find that it satisfies the letter but not the intent of another state's law.

This heterogeneity is not unusual in the history of US regulation. The state privacy law landscape following California's CCPA produced exactly the same pattern: overlapping but non-identical rights, definitions, and enforcement mechanisms, eventually pushing many larger organisations toward a highest-common-denominator compliance approach. The same dynamic is likely to play out in AI regulation, with organisations that invest early in flexible, evidence-generating compliance infrastructure better placed than those that build narrow, state-specific programmes.

SECTION 04

The Algorithmic Discrimination Problem

Algorithmic discrimination in practice means a model that produces statistically different outcomes across protected demographic groups, even when the protected characteristic is not an explicit input. This is a more subtle and more pervasive problem than intentional discrimination, and it arises from several interconnected sources that are worth understanding before considering how to measure and demonstrate compliance.

The most common mechanism is **proxy discrimination**. Models trained on historical data may assign predictive weight to variables such as postcode, credit history, device type, or time of day that correlate strongly with protected characteristics without being protected characteristics themselves. A credit scoring model that penalises applicants in certain postcodes is not explicitly using race as an input, but if those postcodes are demographically homogeneous, the effect may be functionally indistinguishable from a model that does. The EU AI Act Annex IV and the US Equal Credit Opportunity Act both recognise this dynamic, but state AI laws like Colorado SB 24-205 place it front and centre by defining algorithmic discrimination to include differential outcomes without requiring discriminatory intent.

The **measurement problem** is equally significant. Detecting algorithmic discrimination requires access to both model outputs and demographic data for the same individuals, so that disparity metrics can be computed across groups. These two datasets are rarely held together in practice, for good

reasons: organisations often deliberately avoid storing demographic data alongside operational data to limit their exposure under anti-discrimination law. Combining them to test for discrimination creates its own legal and privacy risks, particularly where sensitive characteristics such as health data or religion are involved.

Bias Testing Methodologies

Several methodologies exist for detecting and quantifying algorithmic bias, each with its own strengths and technical limitations. **Disparate impact analysis** measures whether a model's outcomes are distributed proportionally across demographic groups, using the four-fifths rule from US employment law as a common threshold: if the positive outcome rate for a protected group is less than 80% of the rate for the most favoured group, a prima facie case of disparate impact may exist. **Counterfactual fairness testing** asks whether a model would produce a different outcome if an individual's protected characteristic were different while all other inputs remained constant. **Slice-based evaluation** disaggregates model performance metrics (accuracy, false positive rate, false negative rate) by demographic subgroup to identify systematic performance gaps.

None of these methodologies is technically perfect. Disparate impact analysis depends on choosing the right comparison groups and reference period. Counterfactual testing requires a model of how protected and non-protected characteristics interact, which is itself contested. Slice-based evaluation can miss discrimination that manifests only at the intersection of multiple characteristics. Organisations that conduct bias testing should document the methodology chosen, the limitations acknowledged, and the rationale for choosing one threshold over another, because regulators and plaintiffs will scrutinise those choices.

The interaction between these technical challenges and Colorado's reasonable care standard creates a genuine compliance tension. Organisations must demonstrate that they have tested for discrimination, but no single testing approach is universally accepted as sufficient. The most defensible position is one that uses multiple complementary methodologies, documents the results in a tamper-resistant record, and can show that the testing informed subsequent model design or retraining decisions.

SECTION 05

Proving Fairness Without Revealing the Model

The core tension in demonstrating algorithmic fairness to a regulator is that the most direct form of evidence, access to the model weights and training data, is precisely what organisations are most reluctant to provide. Model weights are typically trade secrets representing significant investment. Training data may include sensitive personal information, third-party licensed datasets, or proprietary data whose disclosure is restricted by contract. Full output disclosure creates privacy risks for individuals in the test set. The challenge is to generate evidence that is credible and verifiable without requiring any of these disclosures.

Zero-knowledge proofs offer a technically rigorous path through this tension. In a ZK proof system, a **prover** (the organisation running the model) can demonstrate to a **verifier** (a regulator, auditor, or consumer advocate) that a specific mathematical statement is true, without revealing the underlying data or computation that makes it true. The proof is generated by running a defined circuit over private inputs and producing a compact cryptographic object that can be checked efficiently by anyone with the corresponding verification key.

Applied to fairness testing, the approach works as follows. The model owner specifies a fairness metric: for example, the difference in positive prediction rates between demographic group A and demographic group B across a defined evaluation dataset. The model is run over that dataset, and a ZK proof is generated that attests: the disparity in positive prediction rates between group A and group B on this evaluation set is less than X%, where X is the agreed threshold. The proof is then provided to the regulator. The regulator can verify the proof using the verification key. The verification confirms that the statement is true without revealing the model weights, the evaluation dataset, or any individual's data.

> This is structurally analogous to a financial audit. An auditor certifies that a company's accounts give a true and fair view without publishing every underlying transaction. The audit opinion is credible because the auditor has access to the underlying records and their certification is staked on

> their professional reputation and legal liability. A ZK fairness proof achieves a similar outcome, but the credibility is grounded in cryptographic guarantees rather than professional trust alone.

The practical requirements for implementing ZK fairness proofs include: a precisely specified fairness metric and threshold, agreed in advance with the relevant regulator or auditing body; an evaluation dataset that is representative and, ideally, independently curated; a ZK circuit capable of computing the chosen fairness metric and comparing it to the threshold; and a process for generating and storing the proof alongside the governance record for the evaluation run.

It is important to be clear about what ZK fairness proofs do and do not demonstrate. A proof that a model's disparity metric fell below a defined threshold on a specific evaluation dataset is a meaningful and verifiable compliance statement. It does not prove that the model is fair in every possible real-world deployment context, that the evaluation dataset was truly representative of the deployment population, or that no individual was discriminated against. Regulators should, and likely will, treat ZK fairness proofs as one component of a broader compliance programme, not as a substitute for it. Organisations should use them accordingly.

SECTION 06

Consumer AI Disclosure: The Cryptographic Approach

Colorado SB 24-205 and Illinois HB 3773 both require that consumers be notified when an AI system makes or substantially assists a consequential decision affecting them. This is a transparency obligation, but it is more demanding than it may initially appear. A notification that merely states "an AI system was involved in this decision" is unlikely to satisfy the requirements, because it provides no basis for the consumer to exercise their appeal or correction rights intelligently. The notification needs to be meaningful enough to support that exercise.

The challenge is that meaningful disclosure is in tension with several other legitimate interests. Detailed disclosure of how a model reached a decision could expose proprietary model logic. Disclosure of which input features were weighted most heavily could be gamed by applicants adjusting their behaviour to manipulate the score. And disclosure at the individual level, multiplied across many decisions, creates a corpus of information that could in aggregate reveal things about the model that the organisation has a right to protect.

A cryptographic approach resolves this tension by creating a structured **governance record** for each consequential AI decision. The governance record is generated at the time of the decision and contains: a decision reference identifier; the model version hash, uniquely identifying which version of the model was used; the timestamp of the decision; the category of consequential decision (employment, lending, housing, and so on); the outcome (positive or negative, approved or declined); and a ZK proof attesting that the decision was generated by the specified model version at the specified time, following a process that met defined fairness criteria.

The consumer receives a reference number at the time of notification. They can present this reference number when exercising their appeal or correction right. The reviewer uses the governance record associated with that reference to verify which model version made the decision and when, without accessing any other consumer's data or any internals of the model. This gives the consumer a meaningful basis for appeal and gives the reviewer a tamper-resistant record to work with, while keeping the model's internal logic protected.

The governance record is anchored in a Merkle tree, so that the organisation can later demonstrate, at aggregate level, that all consequential decisions made during a period are accounted for. A consumer advocacy organisation or regulator that receives the Merkle root for a reporting period can verify that any specific governance record they are given is genuinely part of the complete set, with no records having been altered or silently removed. The entire Merkle tree structure is signed with ML-DSA-65, a post-quantum digital signature scheme, providing long-term tamper resistance against both classical and future quantum computing attacks.

SECTION 07

Generating Regulator-Readable Evidence for Algorithmic Accountability

Colorado SB 24-205's annual certification requirement is operationally novel. Deployers must certify to the Colorado Attorney General that they have conducted required impact assessments, implemented required risk management policies, and complied with consumer notification and appeal obligations. The AG's office has not yet published technical guidance on the format that certifications must take, but it is reasonable to expect that certifications will need to be substantiated by underlying evidence if the AG exercises its enforcement authority.

ZK governance records are well suited to supporting this certification. A deployer seeking to certify annual compliance can provide the AG with a package containing: a summary of the AI systems covered, the impact assessments conducted, and the fairness testing results; Merkle roots for each reporting period, allowing verification that all consequential decisions in scope have corresponding governance records; ZK fairness proof references, allowing spot verification of claimed fairness metrics; and an ML-DSA-65 signature over the entire package, establishing tamper resistance for the submission itself.

This approach gives the regulator meaningful verifiability without requiring access to the underlying decision records or model internals. The AG's office can verify, with mathematical certainty, that the Merkle tree attested by a given root contains a specific number of governance records, that those records have not been altered since the tree was constructed, and that the claimed fairness metrics were verified at the time they were claimed. This is substantially more robust than a narrative compliance summary, which can be assembled after the fact and is difficult for a regulator to audit efficiently.

> The Merkle-anchored audit trail enables what might be called **completeness verification**: the regulator can confirm that all consequential AI decisions within a defined scope and period have corresponding governance records, without needing to read any individual

> record. The combination of completeness verification and per-record integrity provides a foundation for meaningful regulatory oversight that is difficult to achieve with traditional logging or reporting approaches.

A practical note on the structure of regulator-readable evidence packages: the most useful format is likely to be a structured document containing the certification statement, the Merkle roots and signing keys for each system and reporting period, the ZK proof references for fairness evaluations, and human-readable summaries of impact assessments. The technical cryptographic elements provide the verifiability; the human-readable elements provide the context that a regulator needs to understand what they are being shown. Neither is sufficient alone.

For the annual certification specifically, organisations should also document the process by which governance records were generated, the chain of custody for Merkle roots between reporting periods, and the version history of each AI system covered. This documentation supports the regulatorys ability to trace any specific decision record back to the model version and evaluation results that were current at the time, which may be important if an enforcement investigation is opened months or years after the fact.

SECTION 08

Cross-Jurisdictional Compliance: Colorado, Illinois, EU AI Act, and NIST AI RMF

Organisations that operate in both the United States and the European Union now face an overlapping matrix of AI governance obligations from multiple directions. The good news is that these frameworks share more structural similarity than their surface differences suggest. The bad news is that the differences that do exist, in scope, definitions, and enforcement mechanisms, require careful attention when designing a compliance programme intended to satisfy multiple frameworks simultaneously.

REQUIREMENT	COLORADO SB 24-205	ILLINOIS HB 3773	EU AI ACT	NIST AI RMF
Impact assessment	Required before deployment; updated on material change	Not explicitly required; implied by audit obligations	Conformity assessment (Article 9); required before placing on market	Govern and Map function voluntary but referenced b Colorado
Bias testing	Implied by reasonable care standard; no prescribed methodology	Annual independent bias audit required; summary published	Annex IV testing requirements; accuracy and robustness across subgroups	Measure function; demographic disaggregatic recommende
Consumer notification	Required for each consequential decision	Required for covered employment decisions	Article 14 (human oversight); transparency to affected persons	Govern function; transparency and accountabilit
Audit trail / record-keeping	Required to support AG certification and consumer appeal	Required to support annual audit	Article 12 (logging); full record-keeping for high-risk AI	Manage function; incident tracking and documentati
Post-market monitoring	Annual AG certification; incident reporting	Annual bias audit	Article 61 (post-market monitoring); serious incident reporting	Measure and Manage functions; ongoing monitoring

NIST AI RMF (AI Risk Management Framework) provides a voluntary US federal baseline that is directly referenced in Colorado SB 24-205 as a relevant standard for risk management policies. The framework's four functions, Govern, Map, Measure, and Manage, map reasonably well to the obligations in Colorado's Act. Organisations that have already aligned their AI risk management to NIST AI RMF are well positioned for Colorado compliance: the

primary additional requirements are the specific consumer notification process and the annual AG certification, neither of which has a direct NIST AI RMF equivalent.

Companies that have invested in EU AI Act compliance are similarly well positioned. The EU Act's Article 12 record-keeping requirements and Article 9 conformity assessment obligations are structurally similar to Colorado's audit trail and impact assessment requirements. ZK governance records designed to satisfy Article 12 can, with appropriate structuring, also satisfy Colorado's documentation requirements. The main divergence is the AG certification process, which has no direct EU analogue, and the consumer appeal mechanism, which is more explicit in Colorado's Act than in the EU Act's human oversight provisions.

The practical advice for organisations facing multiple jurisdictions is to build governance infrastructure around the most demanding requirements and verify that it also satisfies the less demanding ones, rather than building separate programmes for each jurisdiction. A ZK audit trail that satisfies EU AI Act Article 12 logging requirements, supports Colorado's impact assessment documentation needs, and generates the fairness proof references needed for Illinois's annual bias audit is more efficient and more durable than three separate compliance programmes built independently.

SECTION 09

Developer and Deployer Compliance Checklist for Colorado and Illinois

The following checklists are intended as a starting point for compliance planning, not as legal advice. Organisations should seek legal counsel on the application of Colorado SB 24-205 and Illinois HB 3773 to their specific circumstances.

For AI Developers (Colorado SB 24-205)

1. Identify whether your AI system makes or substantially assists consequential decisions as defined by the Act. If your system provides recommendations or scores that feed into covered decisions, it is likely in scope even if it does not make the final determination.
2. Document training data sources, including provenance, date of collection, known limitations, and any filtering or pre-processing applied. This documentation will need to be provided to deployers.
3. Document known limitations of the system: performance degradation on specific population subgroups, input domains where the system has not been validated, edge cases identified during testing.
4. Implement a structured bias testing programme. Run disparate impact analysis, counterfactual testing, or slice-based evaluation across protected characteristic groups. Document the methodology, results, and any remediation steps taken.
5. Establish a process for providing deployers with updates to the above documentation when the system changes materially.

For Deployers (Colorado SB 24-205)

1. Conduct an algorithmic impact assessment before deploying any high-risk AI system. The assessment should document the intended use, the consequential decisions the system will assist, the potential for algorithmic discrimination, and the mitigations implemented.
2. Implement a written risk management policy covering the system's deployment context, the monitoring approach, and the escalation process for potential algorithmic discrimination incidents.
3. Establish a consumer notification process: every consequential decision assisted by the AI system must generate a notification to the affected consumer identifying the AI involvement and explaining how to exercise their appeal or correction right.
4. Implement a consumer appeal and correction process. Document the process and ensure it is accessible to Colorado residents in a timely manner.
5. Prepare for annual AG certification: establish an internal process for compiling impact assessment summaries, governance record counts, and

fairness testing references into the annual certification submission.

For Employers in Illinois (HB 3773)

1. Identify all AI systems used in hiring, promotion, or termination decisions. This includes systems provided by third-party HR technology vendors.
2. Implement notification for all applicants and employees when an automated system contributes to a decision affecting them. The notification should be provided at the time of the decision, not retrospectively.
3. Engage an independent auditor to conduct an annual bias audit. The audit should assess whether the system produces disparate outcomes across protected characteristic groups in the Illinois employment context.
4. Publish a summary of the annual audit results. The publication must be accessible to the public; most organisations will publish on their careers website or in their annual report.

ZK audit infrastructure is particularly relevant to checklist items covering bias testing documentation (developer checklist item 4), consumer notification governance records (deployer checklist items 3 and 4), and the evidence base for annual AG certification (deployer checklist item 5). These are the requirements that most benefit from tamper-resistant, independently verifiable records.

SECTION 10

What Regulators Will Actually Accept

Colorado's Attorney General has not yet published detailed technical guidance on what format impact assessments, bias testing results, and compliance certifications must take. The AG's office is in the early stages of building the expertise and processes needed to receive and evaluate hundreds or thousands of annual certifications from the organisations in scope. This means that organisations filing the first round of certifications will be operating in conditions of regulatory uncertainty about what level of technical detail will be required and what will trigger enforcement attention.

The pattern from EU AI Act implementation offers a useful reference point, though not a perfect analogy. When the EU Act's conformity assessment requirements first came into force, national market surveillance authorities varied significantly in the depth and format of technical documentation they expected. Regulators initially accepted a range of formats and levels of detail. As they built technical capacity and began reviewing submissions in volume, expectations tightened, and organisations that had submitted thin narrative summaries found themselves asked to provide more substantive technical evidence on short notice.

The safest compliance posture is therefore to generate records that are more robust than the minimum currently expected, not less. A compliance programme based on narrative summaries, produced after the fact by the compliance function, is easier to assemble quickly but creates fragility: if enforcement scrutiny increases, or if a specific incident draws regulatory attention, a narrative summary is difficult to defend against a regulator who wants to see underlying evidence. Tamper-resistant records generated contemporaneously with the decisions they document are substantially more defensible.

ZK governance records provide a baseline of tamper resistance and independent verifiability that is likely to be viewed favourably by regulators as their expectations clarify. They are not the only compliant approach: a well-maintained conventional logging system with appropriate access controls and integrity protections may also satisfy the requirements. However, ZK governance records have the advantage of allowing verification without access, which is particularly valuable in the context of state AI law where regulators may not have the staffing to conduct detailed forensic reviews of every submission but will want assurance that the records they receive are genuine.

Organisations should also consider the interaction between regulatory scrutiny and litigation risk. Consumer advocacy groups and plaintiffs' attorneys are likely to use state AI law provisions as the basis for civil actions against deployers whose AI systems produce discriminatory outcomes. In that context, the existence of contemporaneous, tamper-resistant governance records demonstrating that fairness testing was conducted and that defined

thresholds were met will be a significant advantage. Organisations that can produce such records in discovery are in a materially better position than those who cannot.

SECTION 11

Known Gaps in the Current Framework

An honest assessment of the current state of US state AI regulation and the compliance approaches available requires acknowledging several significant gaps and uncertainties. These are not reasons to delay compliance planning; they are reasons to build compliance infrastructure that is flexible enough to adapt as the framework evolves.

Colorado SB 24-205's definition of a "high-risk AI system" is broad, and the application of that definition to specific AI systems in specific contexts is not always clear. A recommendation engine that surfaces employment opportunities to job seekers: is it making or substantially assisting a consequential decision? A credit scoring system used by a lender in Colorado but hosted by a technology vendor outside Colorado: who is the developer and who is the deployer for purposes of the Act? These boundary questions will be resolved over time through enforcement actions, guidance from the AG's office, and potentially litigation. Organisations in ambiguous positions should seek legal advice on how to characterise their systems rather than making unilateral interpretive decisions.

The annual AG certification requirement is administratively novel, and there is no established process for what the AG's office will do with the certifications it receives, how it will prioritise enforcement attention, or what will constitute a satisfactory versus an inadequate submission. It is possible that the AG will initially focus on high-profile sectors (employment, lending, insurance) where algorithmic discrimination has the most direct impact on consumers, and will develop sector-specific guidance over time. Organisations in those sectors should monitor AG communications closely.

The interaction between state AI laws and potential future federal AI legislation creates genuine long-term uncertainty. If Congress enacts a comprehensive federal AI law, it may preempt state laws, either wholly or partially. It may instead choose a minimum standards approach that preserves state authority to impose additional obligations. The outcome will significantly affect the long-term compliance landscape. Organisations should design their compliance programmes to be modular: state-specific obligations can be layered on top of a federal baseline or stripped out if preemption occurs, without requiring a complete rebuild of the underlying governance infrastructure.

Finally, and most importantly: ZK governance records prove that certain processes were followed at certain times. They cannot prove that those processes were designed correctly, that the evaluation dataset was representative, or that the model was genuinely fair in its real-world deployment. A biased model with excellent audit records is still a biased model, and the existence of audit records will not be a complete defence against discrimination claims if the underlying model produces harmful outcomes at scale. Governance infrastructure is a necessary component of a responsible AI programme, but it is not a substitute for investing in model quality, evaluation rigour, and ongoing monitoring of real-world outcomes.

SECTION 12

Conclusion

The US state AI law landscape is moving at CCPA speed. Multiple states, overlapping requirements, enforcement timelines that feel distant until they are not. Colorado SB 24-205 is already in force. Illinois HB 3773 is implementing. Connecticut, Texas, Virginia, and others are following. Organisations that wait for regulatory clarity before building compliance infrastructure will find that the clarity arrives at the same time as the first enforcement actions.

The common thread running through all of these frameworks is **demonstrable accountability**: evidence that your AI system was designed responsibly, tested for discrimination, and generates records that support investigation

when things go wrong. This is not a new concept in regulated industries. Financial services, pharmaceuticals, and critical infrastructure have long operated under accountability regimes that require contemporaneous, tamper-resistant records of consequential decisions. AI governance frameworks are applying the same principle to a new domain.

Companies that build ZK governance infrastructure now are building something that will serve them across Colorado, Illinois, and whatever state laws follow, as well as across the Atlantic for EU AI Act compliance. The investment is not in compliance with a specific statute; it is in a capability to generate credible, verifiable evidence of responsible AI operation. That capability will remain valuable regardless of how the specific statutory landscape evolves.

The key practical steps are: catalogue your AI systems against the Colorado consequential decision categories; conduct and document impact assessments before or shortly after deployment; implement consumer notification and appeal processes; establish a bias testing programme with documented results; and put in place the record-keeping infrastructure needed to support annual AG certification. These steps, taken together, create the foundation for compliance with Colorado's Act and a defensible position as other state frameworks develop.

AffixIO's governance infrastructure provides one component of that foundation: tamper-resistant, post-quantum secure, regulator-readable records of AI decisions, bias testing results, and compliance certifications. It is not the only component an organisation needs, and it is not the right choice for every organisation. But for organisations that have reached the scale where AI systems are making consequential decisions at volume, the question of how to generate durable, verifiable evidence of responsible operation is one that deserves a technical answer, not just a narrative one.

FAQ

Frequently Asked Questions

Does Colorado SB 24-205 apply to AI systems used only internally?

Yes, if the AI system makes or substantially assists a consequential decision affecting a Colorado resident, it is in scope regardless of whether the system is customer-facing or internal. An internal HR AI system used to make hiring decisions about Colorado employees is covered by the Act. The relevant test is the nature of the decision and whether it affects a Colorado resident, not the outward-facing character of the system.

What is an algorithmic impact assessment?

A documented analysis of an AI system's potential for causing algorithmic discrimination, conducted before deployment and updated when the system changes significantly. Colorado SB 24-205 requires deployers of high-risk AI to conduct these assessments. A well-constructed algorithmic impact assessment should document the system's intended use, the consequential decision categories it assists, the population of affected consumers, the demographic groups at risk of discriminatory outcomes, the bias testing results, and the mitigations implemented. It is a living document, not a one-time exercise.

How does a ZK fairness proof work in practice?

A model owner runs a specified fairness evaluation, for example measuring the difference in positive prediction rates across demographic groups on an independent test set, and generates a ZK proof that the disparity metric falls below a defined threshold. The proof is verifiable by any party with access to the verification key, without revealing the test data or model weights. The proof attests to a specific mathematical statement about a specific evaluation run, and can be stored as part of a governance record alongside the model version hash and the evaluation dataset identifier, creating a durable, auditable record of the fairness testing conducted.

RELATED READING

Related Whitepapers

- WP-001: Cryptographic AI Governance: foundational architecture for tamper-resistant AI decision records
- WP-011: Merkle Tree Audit Architecture: how Merkle-anchored audit trails enable completeness verification
- WP-015: Agentic AI Governance: governance records for multi-step agentic AI systems
- WP-020: DORA, MiCA, and AI Compliance in Financial Services: cross-framework compliance in regulated financial contexts