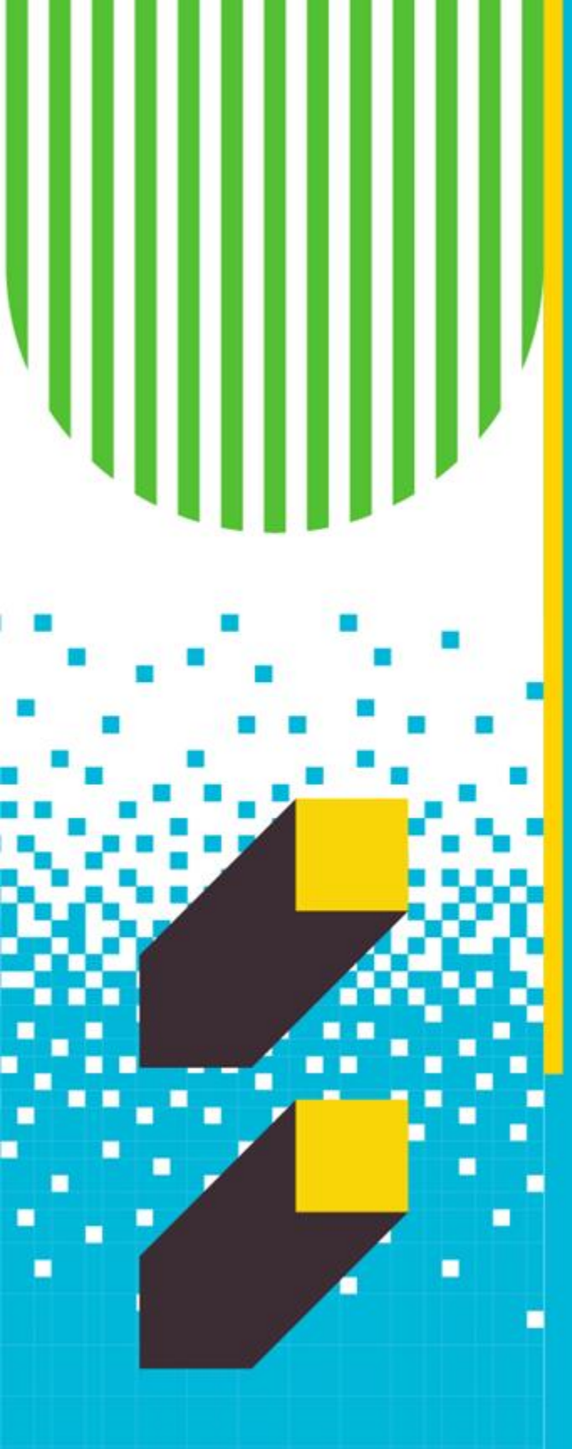


IAPP **Privacy. Security. Risk. 2025**

Training 28-29 October

Workshops 29 October

Conference 30-31 October

SAN DIEGO

#PSR25

AI Governance as a Driver of Innovation Amidst Regulatory Flux

October 31, 2025

WELCOME AND INTRODUCTIONS

Ken Priorie

Deputy General Counsel, Product, Engineering, IP and Partners- Docusign

Bret Cohen

Partner, Data, Privacy, Cybersecurity & AI
Hogan Lovells

Saima Fancy

Data Governance Product Manager
Adobe

Part I - Regulatory Catalysts for Internal Action

EU AI Act

Governance obligations for non-GPAI, “high risk” systems – effective August 2026

Provider obligations

- *Art. 9 - Risk management system* - a “continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating”
- *Art. 10 - Data & data governance* - “Training, validation and testing data sets shall be subject to data governance and management practices appropriate for the intended purpose”
- *Art. 11 / Annex IV - Technical documentation* - 23 requirements
- *Art. 12 - Record-keeping* - Automated logging requirements
- *Art. 14 - Human oversight*
- *Art. 15 - Accuracy, robustness & cybersecurity*
- *Art. 17 - Quality management system* - 13 requirements

Deployer obligations

- Take “appropriate technical and organisational measures” to ensure use in accordance with provider instructions
- Assign human oversight
- Where applicable, undertake a DPIA
- Where the deployer exercises control over input data, ensure that it is relevant and sufficiently representative
- Monitor the operation of the AI system for risks
- If risks detected, notify providers and, in some circumstances, “the relevant market surveillance authority”

Colorado AI Act

Governance obligations for “high risk” systems – effective June 30, 2026

Developer obligations

- Use “reasonable care to protect consumers from any known or reasonably foreseeable risks of algorithmic discrimination”
- Make available documentation for deployers, including:
  - Foreseeable risks of algorithmic discrimination, and measures taken to mitigate those risks
  - How it was evaluated for algorithmic discrimination
  - Data governance measures for training data sets
  - Info needed for deployer to conduct impact assessment

Deployer obligations

- Use “reasonable care to protect consumers from any known or reasonably foreseeable risks of algorithmic discrimination”
- Implement a risk management policy and governance program, “an iterative process planned, implemented, and regularly and systematically reviewed and updated over the life cycle of [the AI system], requiring regular, systematic review and updates”
  - Must be “reasonable considering” the NIST AI RMF, ISO/IEC 42001, “or another nationally or internationally recognized risk management framework for [AI] systems”
- Conduct an annual impact assessment
- Undertake post-deployment monitoring
- Notify the AG of algorithmic discrimination

CCPA Regulations on Automated Processing

Governance obligations – effective January 2026 (risk assessments for existing processes by December 31, 2027)

“Automated decisionmaking technology” or “ADMT”

Means any technology that processes personal information and uses computation to replace human decisionmaking or substantially replace human decisionmaking (i.e., use the output to make a decision without human involvement)

Risk assessments required when:

- Using or training ADMT for “significant decisions” about consumers, facial- recognition, emotion-recognition, identity-verification, or physical/biological-profiling
- Using automated processing to infer characteristics about a person based on (1) systematic observation of that person or (2) the person’s presence in a sensitive location

Risk assessments must document 16 requirements and:

- Determine whether the “risks to consumers’ privacy” outweigh benefits to the consumer
- Be made available to downstream deployers of ADMT
- Be refreshed at least once every 3 years, or upon material change
- Be submitted to the CPPA

Enforcement & Regulatory Expectations

Unfair or deceptive acts or practices

- *"Unfair"*: A practice that causes or is likely to cause substantial injury to consumers which is not reasonably avoidable and not outweighed by countervailing benefits to consumers or to competition
- The use of AI or other automated decision-making systems can be unfair if they produce **biased**, **discriminatory**, or **incorrect** results that could have been reasonably avoided

Enforcement & Regulatory Expectations

FTC "unfair" AI enforcement – *Rite Aid*

- Deployed AI-based facial recognition technology to identify customers who may have been shoplifting
- Collected a database of 10,000+ images of "persons of interest" along with names and other criminal background data
- Many of the images were low-quality and came from security cameras, employee phone cameras, and news stories
- System generated "thousands" of false-positive matches (e.g., flagging the same person across the U.S.)

Enforcement & Regulatory Expectations

FTC alleged that there was a failure to appropriately:

- Consider and mitigate risk of misidentifying consumers
- Test, assess, measure, document, or inquire about accuracy before deployment
- Prevent the use of low-quality inputs
- Regularly monitor or test accuracy after deployment
- Train employees

Enforcement & Regulatory Expectations

Questions from actual regulatory inquiries

- Describe how AI technology is integrated into your product, including LLMs used; sources of data used to train/fine-tune; and the software/data architecture
- Regarding the generation of "prohibited outputs" (i.e., results that violated policies by being false, misleading, or biased), demonstrate the steps taken to:
  - identify, assess, or measure the AI system's capacity to generate such "prohibited" or "harmful" outputs
  - address or mitigate risks, including filtering, editing, modifying, supplementing, or anonymizing data sets; training of models; filtering/blocking of inputs/outputs
  - monitor, detect, investigate, or respond to complaints or reports of violations
- Produce all complaints received about prohibited outputs
- Produce all internal communications/assessments about potential to produce prohibited outputs
- Produce all testing, red team exercises, audits, assessments, or analyses regarding the extent to which your AI system generates prohibited outputs
- Provide all prompts or instructions appended to user inputs to the AI system, and describe the dates of deployment and the scenarios to which they are applied

Common Threads Across Laws

- Be prepared!
- Expectation to **risk assess**, **test**, and **monitor** high-risk AI and ADMT systems
- Expectation to **identify** and **explain** why AI systems achieved a certain result
- Expectation that you sought to identify and mitigate **reasonably foreseeable risks**

Part II - Strategic Framework Design & Governance Structures

Three-Pillar Governance Architecture

Building AI governance that actually works—and gives you a competitive edge

Create a Trust Council

Stop the finger-pointing. A cross-functional Trust Council makes it crystal clear who's responsible for AI decisions. No more "I thought Legal was handling it" or "I assumed Product checked that."

Use What You Already Have

You've already invested in privacy tools and data governance. Extend them to cover AI instead of building everything from scratch. Work smarter, not harder.

Trust Assessment

Measures how AI systems affect organizational trust, moving beyond mere risk cataloging. Go beyond checklists. Measure whether your AI systems actually build or erode trust with the people who matter most: your customers, employees, and regulators.

Authority and Accountability Boundaries

Clear boundaries prevent confusion and conflict. Here's who decides what.

1

Trust Council Authority

The Trust Council sets requirements and mandates controls. Operational teams execute within those boundaries.

Executives remain accountable for overall program success. The Trust Council operates with delegated authority to make day-to-day governance decisions.

This creates a balance: the Trust Council has real power to enforce standards, while operational teams retain autonomy within established guardrails.

2

Who Owns What

Accountable vs. Responsible—they're different:

Executives are accountable for the entire AI risk management program (e.g., the CTO owns overall success)

Product Owners are responsible for documenting and managing risks in their specific AI solutions

Trust Council members are responsible for implementation and enforcement of governance decisions

3

Setting and Enforcing Policy

The Trust Council creates binding policies. Examples:

- Prohibiting use of unapproved generative AI tools for decision-making
- Blocking sensitive data input into public AI systems
- Requiring specific security controls for high-risk deployments

Enforcement happens through existing systems: IT governance, access controls, monitoring, and compliance programs.

4

Power to Pause

The Trust Council can stop a release when the trust profile is insufficient or risks exceed tolerance.

Example: If monitoring detects model drift causing biased outcomes, the Trust Council mandates an immediate pause and requires retraining before the system can resume operation.

Escalation path: Issues beyond Trust Council authority go to Executive Leadership for final decision.

Trust Council Architecture Framework

The Trust Council operates within a dual-layered governance structure anchored by the NIST AI Risk Management Framework (RMF) functions of Govern, Map, Measure, and Manage.- ***This will depend on the nature of your organization***

Executive Sponsor	CTO, CPO,CIO, CDO, etc	**Accountable (A)** for the entire AI governance program. Sets AI objectives, defines risk thresholds, allocates resources, and owns strategic outcomes.
Data Science/Engineering	Risk Specialists, Senior Engineers, Data Scientists, ML Engineers	**Responsible (R)** for hands-on risk management. Oversees AI development, monitors system performance, implements technical controls, ensures policy compliance, and produces technical documentation (model architecture, data inputs/outputs, training procedures).
Legal Counsel / Privacy Lead	Legal, Compliance, Privacy leads	**Consulted (C) / Responsible (R)** for regulatory compliance. Ensures AI systems meet legal and regulatory requirements, provides data protection expertise, and develops compliance strategies.
Product Owner	Product Management, AI Technical Lead	**Responsible (R)** for product-level governance. Documents risk ownership, manages escalation processes, maintains audit evidence for specific AI solutions, defines product vision, and prioritizes features with governance requirements in mind.
Customer Impact	Customer Sales/Delivery/Satisfaction, Marketing, Messaging Team	**Consulted (C) / Responsible (R)** for stakeholder perspective. Represents the customer voice, ensures alignment between technical capabilities and external messaging, and validates that documentation reflects actual user experience.

Pillar Two: Resource Multiplication Strategy

Add AI requirements to your existing governance systems instead of creating separate AI-only structures. This augmentation approach lets you scale AI governance without proportionally scaling headcount or budget.

Privacy Impact Assessments → AI Impact Assessments

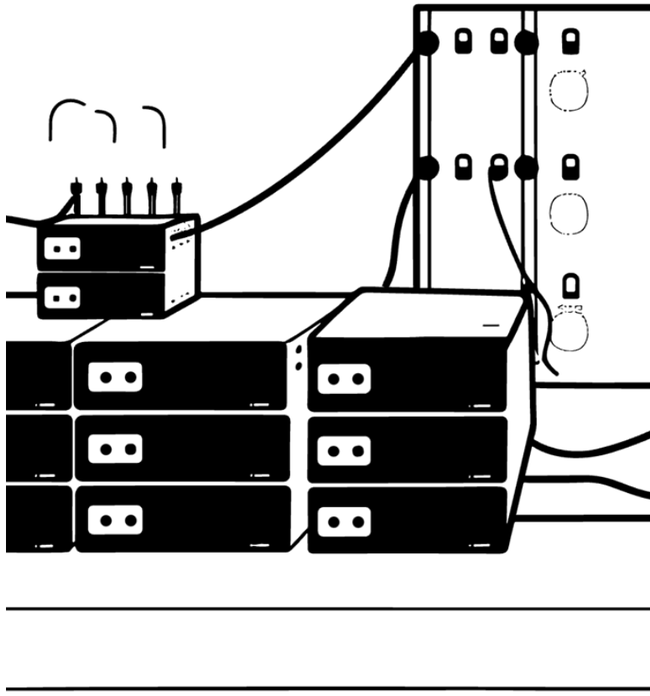
Expand your existing PIAs to cover AI-specific risks: model bias, explainability gaps, algorithmic fairness, and technical robustness—not just data privacy

Data Governance → AI Data Governance

Extend your current data frameworks to track AI training data sources, document data provenance, and implement bias detection and mitigation processes.

Cybersecurity → AI Security

Integrate AI-specific threats into your existing security infrastructure: prompt injection attacks, model extraction, data poisoning, and adversarial inputs



Step 1: Expand Your Risk Assessments

Turn your Data Protection Impact Assessments (DPIAs) into AI Impact Assessments (AIIAs).

Start With What Triggers a DPIA

Use your existing DPIA triggers as the starting point for AI projects:

- Processing personal data
- Systematic monitoring
- Use of innovative technology

If it would trigger a DPIA, it triggers an AIIA. This ensures consistent evaluation across all AI initiatives involving data.

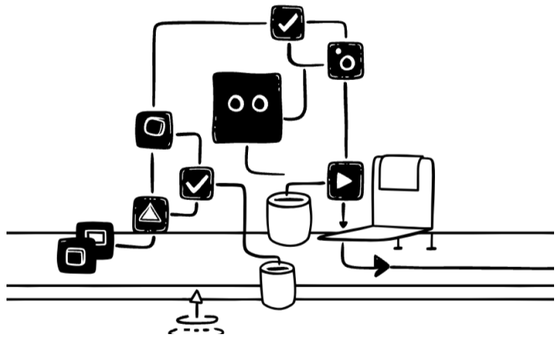
Add AI-Specific Risk Analysis

Integrate threat modeling into your DPIA process to identify AI-specific vulnerabilities:

- Prompt injection attacks
- Data poisoning
- Model extraction
- Adversarial manipulation
- Algorithmic bias
- Explainability gaps

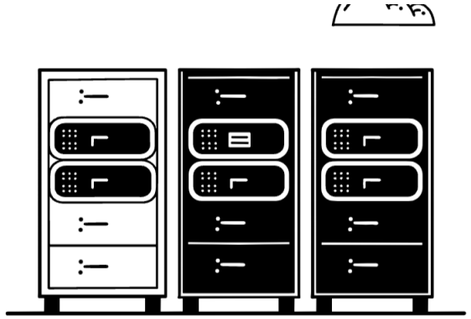
Same assessment structure, expanded scope. Your teams already know how to conduct impact assessments—now they're evaluating additional risk categories.

Step 2: Repurposing Data Management and Data Subject Rights Programs



Privacy by Design for Data Minimization

Integrate Privacy by Design (PbD) at every stage of AI development. Minimize the collection of personal data by applying PbD principles from the outset to reduce unnecessary data.



Anonymization for Training Data

Utilize anonymization techniques to effectively remove personal identifiers from training datasets. Prioritize publicly available data or use synthetic data to reduce reliance on sensitive proprietary information.

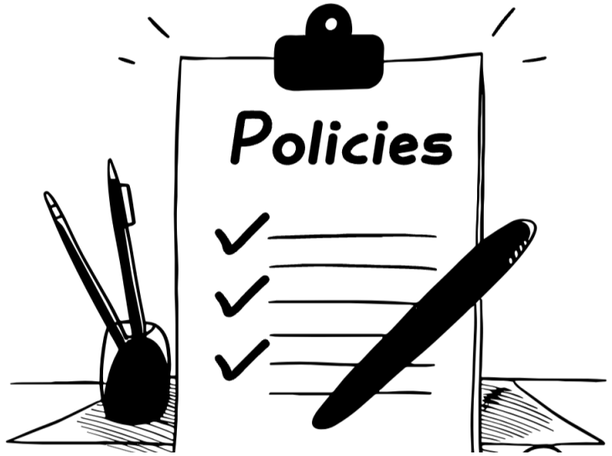
RAG Systems with Privacy Protection

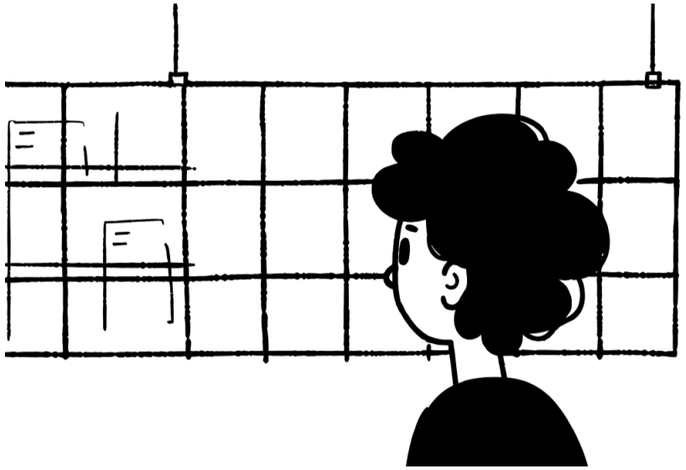
Apply anonymization and pseudonymization techniques to data used in RAG systems before processing. Implement robust security measures to ensure confidentiality and data integrity, adhering to necessity and proportionality principles.

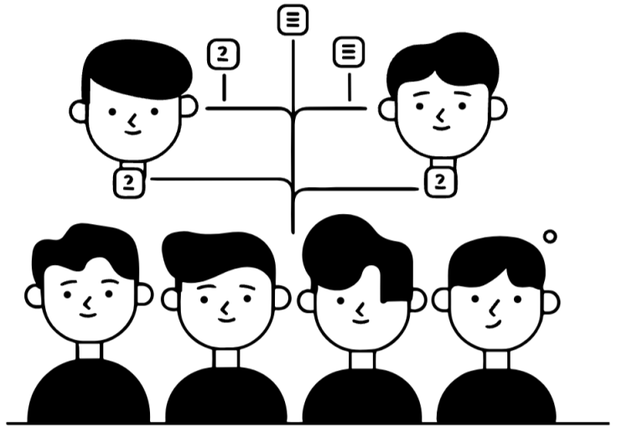
Existing data minimization and data subject rights programs can be extended to cover the specifics of AI training data and RAG systems.

Step 3: Integrating Governance and Oversight Structures

Rather than building a separate AI governance board, embed AI-specific roles and responsibilities within your existing risk governance framework. This integration approach maximizes organizational efficiency while ensuring comprehensive oversight.







Policy Integration

Incorporate AI governance into existing policies: incident response, data use, vendor due diligence, acquisition, and procurement.

Example: Update vendor management programs to mandate contractual prohibitions against vendors using proprietary data for AI training.

Screen vendors against vulnerability databases and verify model license compliance.

RACI Integration

Assign AI-specific responsibilities (e.g., AI Impact Assessment, model testing, risk acceptance) to existing cross-functional teams like legal, security, and product management.

Management Oversight

The management team leverages its operational role to implement the AI risk management program and ensure comprehensive oversight.

Step 4: Extending Cybersecurity Controls to AI-Specific Threats

Existing investments in enterprise-wide security measures (e.g., DLP, access controls, monitoring) can be extended to cover AI's new attack surface, encompassing data, models, and outputs.

Deploy Prompt Security & AI-DLP

Implement robust prompt security measures to prevent injection attacks and data leaks. Integrate AI-specific Data Loss Prevention (DLP) strategies with existing enterprise security to enhance overall protection across the AI lifecycle.

Continuous Monitoring & Auditing

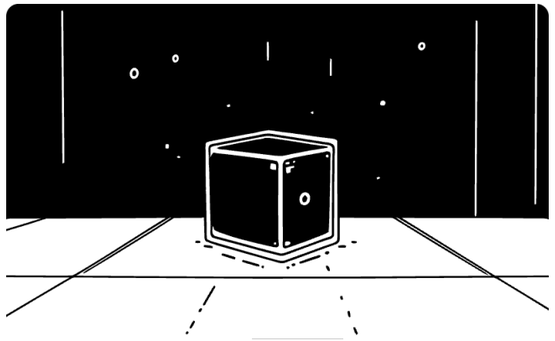
Extend existing IT monitoring frameworks to include AI-specific behaviors. Implement system logs and dashboards to track usage patterns, API calls, and flag unusual access to internal AI systems, inspired by financial trading controls.

Model Risk Evidence

Mandate continuous monitoring to detect model drift, biases, and compliance issues. Collect evaluation results from metrics, adversarial testing, and Red Teaming exercises as robust evidence for AI risk management decisions.

Pillar Three: Trust Assessment

The Trust Assessment framework measures how AI systems affect organizational trust, moving beyond mere risk cataloging. It connects technical risk analysis directly to what truly matters: whether customers, employees, and regulators can trust your AI systems.



Measure Trust Impact

Move beyond cataloging risks to actively measure how AI systems influence organizational trust among customers, employees, and regulators.

Reframe Risk with a Trust Lens

Re-evaluate existing risks (e.g., privacy, bias, opacity) by asking: "What does this risk mean for trust?" This changes how risk data is interpreted.

Stakeholder Trust Considerations

Understand that a privacy breach damages customer trust, biased outputs erode employee confidence, and model opacity makes stakeholders question AI insights.

Technical vs. Trust Risk

A "moderate" technical risk can have severe trust implications if it impacts a customer-facing system, while a "high" technical risk in an internal tool might have limited trust impact.

Phase 1: Mapping Risk to Trust Dimensions

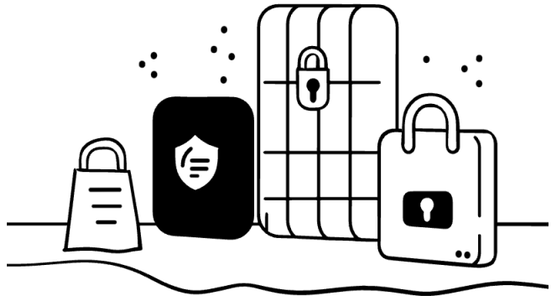
Your existing risk assessments already contain trust signals—you just need to surface them. When the impact assessment flags potential bias, that's data about fairness trust. When monitoring detects model drift, that's data about reliability trust. When security testing finds vulnerabilities, that's data about safety trust.

Safety Trust

Privacy and security risks: Can we protect people?

Equity Trust

Fairness and bias risks: Does this treat people fairly?



Reliability Trust

Performance and accuracy risks: Can people count on this?

Understanding Trust

Transparency and explainability gaps: Can people make sense of this?

Each AI system will have a trust profile showing where it's strong and where it's vulnerable. A high-stakes decision system needs high marks across all dimensions. An experimental internal tool might only need strong safety trust.

Phase 2: Quantifying Trust Impact

Use the same scoring you'd apply to risk severity, but frame it around trust degradation:

Limited trust impact

Issues are contained, stakeholders remain confident

Moderate trust impact

Concerns emerge but can be addressed through communication

Significant trust impact

Confidence wavers, requires active trust rebuilding

Severe trust impact

Trust breaks down, relationship repair needed

This quantification then informs a broader strategic analysis, guiding decisions on how to allocate resources and evaluate opportunities based on potential trust implications. The probability assessment shifts too. Instead of "likelihood of occurrence," think about "likelihood of trust exposure." An internal model error might have low exposure if caught in testing. The same error in production, visible to customers, has high exposure.

Prioritization Decisions

The Trust Council leverages these impact scores to prioritize mitigation efforts, focusing resources where trust degradation poses the greatest risk to stakeholders and the organization.

Resource Allocation

High trust impact necessitates immediate and significant investment in controls, design changes, or transparent communication strategies to maintain confidence.

Strategic Opportunity Evaluation

Proposed AI initiatives are evaluated not just on business potential, but also on their inherent trust profile and the projected trust impact, ensuring sustainable growth.

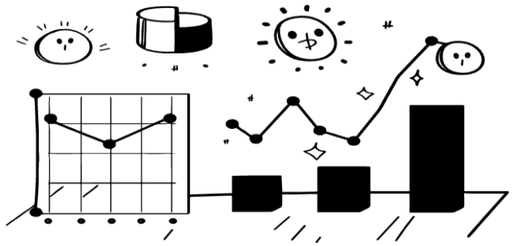
Business Value & Trust

Understanding trust impact helps articulate the direct link between building and maintaining trust and generating long-term business value, protecting reputation and customer loyalty.

This integrated approach allows us to proactively manage potential downsides while seizing opportunities for innovation responsibly. Each AI system will have a trust profile showing where it's strong and where it's vulnerable. For example, a system might show:

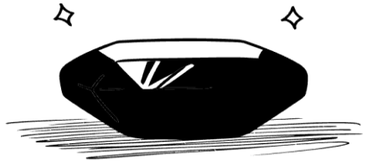
Phase 3: Trust Investment Decisions

Now the cost-benefit analysis makes sense in business terms. You're not just weighing "mitigation cost vs. residual risk." You're asking: What investment restores trust, and is that worth it for this use case?



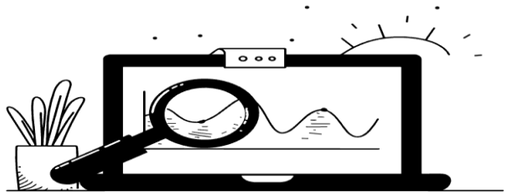
Additional Training Data

From underrepresented groups to improve model fairness.



New Fairness Constraints

Implement new constraints directly into the model's architecture.



Enhanced Monitoring

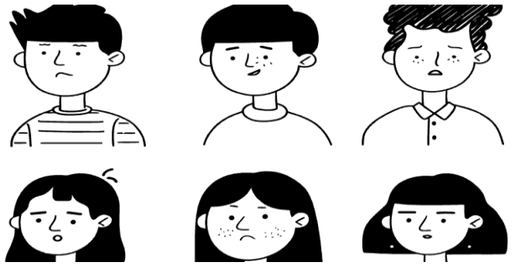
Continuously monitor for disparate outcomes and potential biases.

Clear Documentation

Provide **transparent** documentation of model capabilities and limitations.

Real Costs

These have real costs in time, resources, and sometimes performance trade-offs.



Restored Stakeholder Confidence

The benefit is restored confidence that the system treats people fairly.

#PSR25

Part III - Operational Implementation & Measurement

Data Sources

Mitigations

Bias Assessment

Transparency

Data Governance for AI: Lifecycle Management

AI systems require rigorous oversight across three distinct data categories, each with unique governance requirements and compliance implications.

Training Data

Source validation, consent verification, quality assessment, bias evaluation, retention policies, and audit trails for all data used to build models

Inference Data

Real-time input validation, purpose limitation enforcement, access logging, temporary storage controls, and deletion protocols for operational data

Derived Data

Output classification, secondary use restrictions, aggregation rules, re-identification risk assessment, and long-term retention decisions

Public sector organizations face additional constraints including Freedom of Information requirements, data sovereignty rules, and procurement transparency obligations that must be embedded into governance workflows.

Operational Control Framework

Effective AI governance requires a multi-layered approach that addresses risk identification, data lifecycle management, privacy engineering, and continuous monitoring. These four pillars work together to create a robust framework that meets regulatory expectations while enabling innovation.

AI Risk Registers

Extend traditional privacy impact assessments to capture AI-specific risks including model bias, data drift, and algorithmic transparency gaps

Data Use Governance

Manage training data, inference data, and derived data lifecycles within public sector constraints and regulatory boundaries

Privacy Engineering

Apply data minimization, purpose limitation, and transparency principles systematically throughout AI system design

Continuous Monitoring

Integrate real-time governance controls directly into AI operations for immediate detection and response

AI Risk Registers

Document model bias, drift, and transparency gaps

Data Use Governance

Manage training, inference, and derived data lifecycles

Privacy Engineering

Apply minimization, purpose limits, and transparency

Continuous Monitoring

Real-time controls for detection and response

Privacy Engineering Principles in Practice

01

Data Minimization Assessment

Evaluate whether each data element is strictly necessary for the AI system's purpose. Remove or aggregate data that doesn't directly contribute to the model's legitimate function.

02

Purpose Limitation Enforcement

Define explicit, legitimate purposes for AI processing. Implement technical controls that prevent data from being used for secondary purposes without explicit consent or legal basis.

03

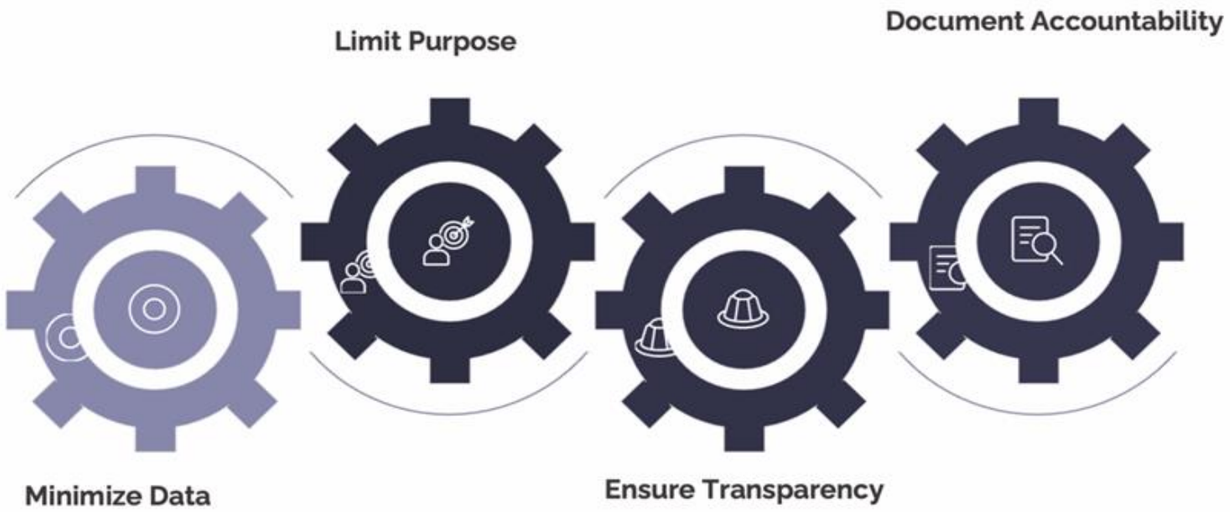
Transparency Mechanisms

Build explainability into model architecture. Create user-facing interfaces that communicate how decisions are made, what data influences outcomes, and how to challenge results.

04

Accountability Documentation

Maintain comprehensive records of design decisions, risk assessments, mitigation measures, and ongoing monitoring results to demonstrate regulatory compliance.



Continuous Monitoring Architecture

Real-time governance controls integrate directly with AI system operations, enabling immediate detection of drift, bias, or compliance violations before they impact patients or operations.

Automated Alerting

Threshold-based notifications for model performance degradation, bias indicators, unusual data patterns, access violations, and regulatory trigger events

Audit Trail Capture

Comprehensive logging of all AI system interactions including data access, model invocations, decision outputs, user overrides, and system configuration changes

Incident Response Integration

Automated workflows that trigger investigation, containment, and remediation processes when governance violations or anomalies are detected

Capture Audit Trail

Log access, outputs, overrides

Automate Alerts

Thresholds for performance and bias

Monitor Continuously

Real-time governance and detection

Document Decisions

Record design, risks, mitigations

Implementation Tools for Regulated Environments

Specialized software solutions address the unique requirements of healthcare and public sector AI governance, providing end-to-end visibility and control.

Data Catalog & Traceability

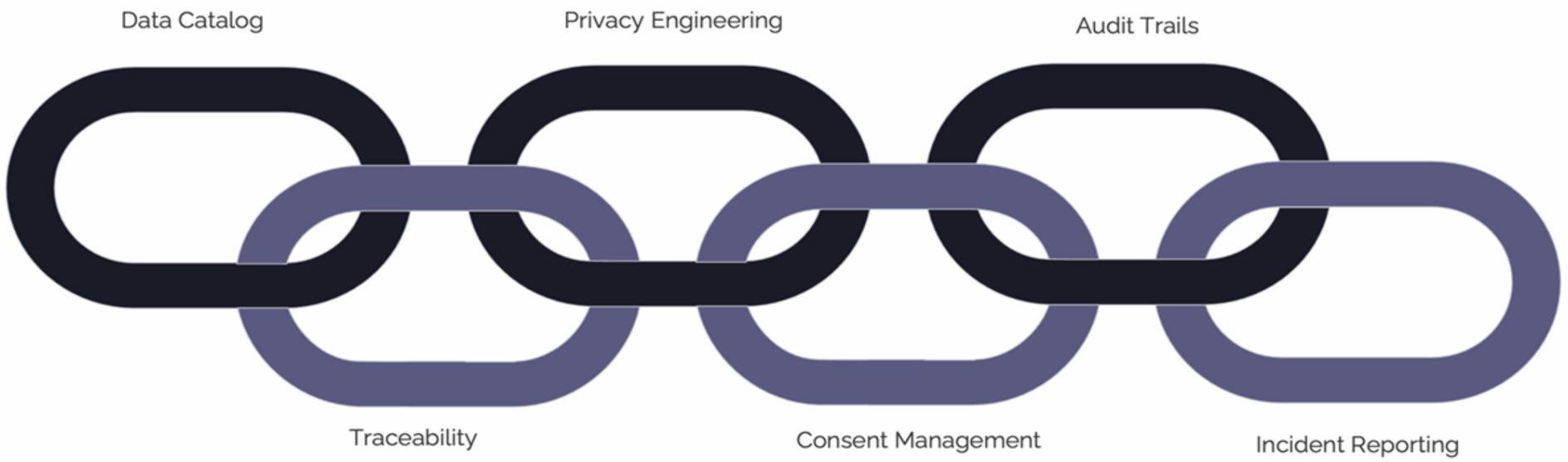
Document data sources, transformations, and lineage throughout the AI lifecycle. Track consent, legal basis, and retention requirements for each dataset. Enable rapid response to data subject requests and regulatory inquiries.

Privacy Engineering Toolkits

Implement de-identification techniques, manage dynamic consent workflows, enforce granular access controls, and apply differential privacy methods. Automate privacy-by-design assessments during development.

Incident Reporting & Audit Trails

Capture AI-triggered events, document oversight measures, maintain immutable audit logs, and generate regulatory reports. Integrate with existing quality management and risk systems for seamless compliance.



Real-World Measurement: Industry Examples

Organizations across regulated industries demonstrate measurable value through sophisticated tracking of governance outcomes.

Healthcare

Hospitals track time-to-incident-report and audit readiness with automated alerts, cutting compliance risks by 40% and improving patient trust scores by 25% through transparent AI governance practices.

Financial Services

Banks measure suspicious transaction detection rates and AML resolution time, reducing regulatory costs by $3M annually while boosting transparency and customer confidence in AI-driven decisions.

Manufacturing

Manufacturers monitor supplier certification gaps and fine reduction rates, streamlining corrective actions and avoiding penalties totaling $1.5M while maintaining supply chain compliance.

Technology

Tech companies measure IP breach detection speed and response times, safeguarding innovation assets worth $50M+ annually while maintaining compliance across global regulatory environments.

Practical Tools: Automated compliance platforms (Sprinto), enterprise risk management software (Diligent), integrated KPI dashboards, and custom API-based reporting systems enable real-time measurement, transparency, and data-driven decisions that align regulatory objectives with business outcomes.

Thank you!

Q & As

How Did Things Go?
(We Really Want To Know)

Did you enjoy this session? Is there any way we could make it better? Let us know by filling out a speaker evaluation.

1. Open the IAPP Events app.
2. Select **IAPP Privacy. Security. Risk. 2025**
3. Tap "Schedule" on the bottom navigation bar.
4. Find this session. Click "Rate this Session" within the description.
5. Once you've answered all three questions, tap "Done".

Thank you!

Appendix

Meeting Structure and Decision Protocols

The Trust Council maintains oversight through structured meetings and clear decision-making rules.

Governance Gates	**Mandatory approval checkpoints** at critical stages: Plan & Design, Deploy & Use, and others as defined by your organization.	**No product advances** to the next stage without documented risk assessment and formal Trust Council approval.
Risk Thresholds	**The Trust Council classifies each AI system** by risk level: minimal, low, moderate, high, or very high—based on the system's maximum potential risks.	**High-risk systems require Trust Council approval** for any significant changes. This ensures all modifications stay within organizational risk tolerance.
Mandatory Assessments	**Every new AI capability undergoes an AI Impact Assessment** before deployment. This assessment covers AI-specific risks beyond traditional privacy concerns: bias, explainability, model robustness, and more..	**Documentation and mitigation plans are required** for Trust Council approval. No assessment, no approval. No exceptions.

Pillar One: Establish the Trust Council

The primary goal is to centralize responsibility and eliminate the diffusion of responsibility, creating clear accountability structures across organizational functions.

Define Decision-Making Authority

Executive Leadership sets the organization's AI risk tolerance and strategic direction. They own the outcomes.
Trust Council implements and enforces those decisions. This cross-functional team (Legal, Product, Engineering, Security, Privacy) makes day-to-day governance calls.

Document roles and responsibilities using RACI or similar frameworks. Every key AI activity has a clear owner—no ambiguity about who decides what.

Connect AI Governance to Existing Controls

Integrate with current frameworks. AI governance connects to—not replaces—your existing legal, privacy, and IT governance structures.
Use your AI policy as a bridge. It links AI-specific requirements to the compliance systems already in place.
Implement Governance Gates. These are mandatory checkpoints that stop AI systems from advancing when risk thresholds are exceeded. No exceptions.

Give the Trust Council final authority. When risks exceed organizational tolerance, the Trust Council decides whether to proceed, modify, or halt—and that decision stands.

AI Governance as a Driver of Innovation Amidst Regulatory Flux

2030 Vision: Governance as Competitive Advantage

By 2030, governance maturity becomes the primary competitive advantage—automated risk assessments integrate into development pipelines, accelerating rather than delaying deployment.

- Real-time monitoring detects bias and model drift, triggering automatic adjustments before production issues arise.
- Predictive governance dashboards enable executives to forecast compliance risks and model outcomes, transforming oversight from reactive burden to strategic asset.

Build Forward: Evolve Existing Assets, Don't Start Over

Build upon existing privacy tools (DPIAs, data lineage tracking, security controls).

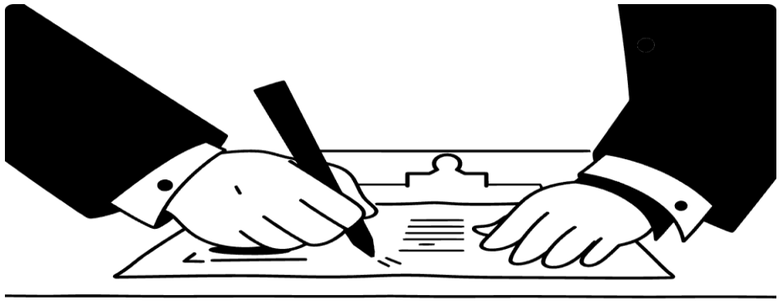
- Augment proven assets into AI Impact Assessments and model provenance systems.
- Evolution requires augmenting what already works, not replacing it.

From Brake to Accelerator: Building Governance That Drives Innovation:

Governance is a product, not just compliance overhead.

- It connects Legal, Product, and Engineering to embed trust and guardrails.
- Traditional risk management is insufficient for rapidly evolving AI.
- Proactive governance transforms compliance into a competitive asset.

Acceptance, Monitoring & Trust-Building Documentation



Acceptance with a Trust Lens

Final acceptance looks different through the trust lens. Instead of "Is residual risk acceptable?" ask "Is the trust profile sufficient for this deployment?" A system with moderate reliability trust might be fine for suggesting meeting times. The same trust level is insufficient for medical diagnosis support. Context determines what "sufficient" means.

Ongoing Trust Monitoring

Ongoing monitoring tracks trust indicators, not just technical metrics. Are users acting on AI recommendations? Are they raising concerns? Is adoption growing or stalling? These behavioral signals tell you whether trust is holding or eroding. When trust metrics decline—even if technical performance looks stable—that's your signal to investigate. Sometimes the issue is communication, not capability. Other times, it reveals problems your technical monitoring missed.

Documentation That Builds Trust

All of this generates documentation, but make it work for trust-building, not just compliance. Your trust assessment should be something you could share with a skeptical stakeholder to show:

- We identified what could affect trust
- We measured the actual impact
- We made intentional choices about investment
- We're monitoring the signals that matter

This isn't about creating perfect systems—it's about building confidence that you're managing AI systems thoughtfully, with eyes open to how they affect the people who need to trust them.