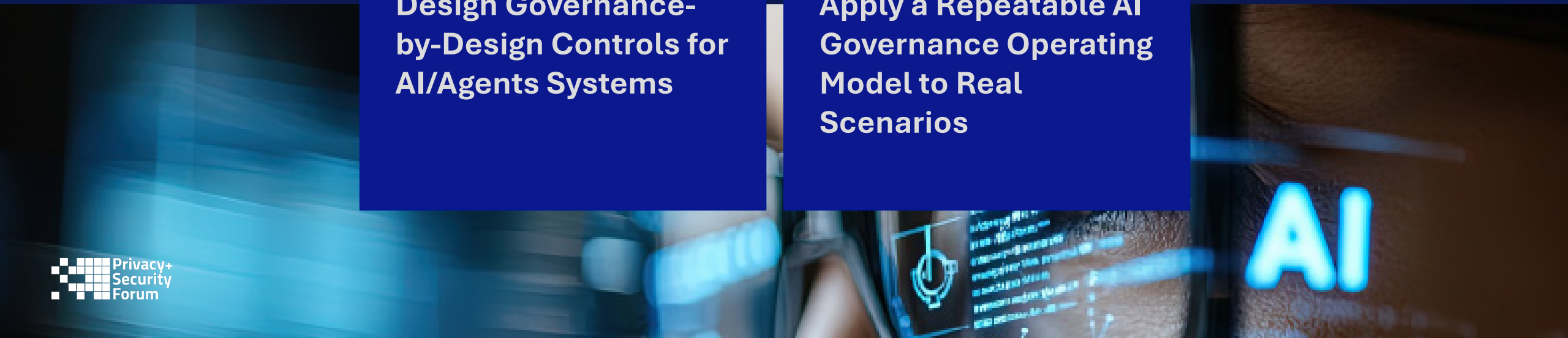
Privacy + Security Forum

Session: AI Governance Workshop
Privacy, Priorities, and Persuasion
May 6, 2026, 8:50 AM – 11:30 AM

Privacy+ Security Forum

Objectives for Today

- Interpret AI Governance Landscape & Enforcement Trends
- Explain what Changes with Agentic AI and Identify Governance Risks
- Design Governance-by-Design Controls for AI/Agents Systems
- Apply a Repeatable AI Governance Operating Model to Real Scenarios



Privacy+ Security Forum

Moderators: BDO Representatives

Karen Schuler
Global Head of Privacy, Data & AI Governance
MC and Panel #1

Taryn Crane
Cyber & Risk Governance Practice Leader
Panel #2

Mark Melnychenko
Privacy Technology Practice Leader
Workshop #1

Ken Frantz
Innovation & AI Leader
Workshop #2

Privacy+ Security Forum

Panelists: Industry Experts

Avery Blank
Senior Legal Counsel

Qualcomm

Maria Buccieri
Senior Director, Enterprise Compliance

Amtrak

Drew Bjerken
Chief Privacy & AI Governance Officer

Marriott Vacations Worldwide

Will Farley
Vice President, Data Strategy

The Reynolds and Reynolds Company

Wesley Taylor
Senior AI & IG Consultant

Southwest Airlines

AI Regulatory & Compliance State of the Union

Building Responsible AI Agents

Privacy+ Security Forum

Session: AI Governance Workshop

Panel Agenda

Timing	Topic
8:50 – 9:30	AI Regulatory & Compliance State of the Union
9:30 – 10:10	Building Responsible AI Agents
10:10 – 10:15	Break

Workshop Agenda

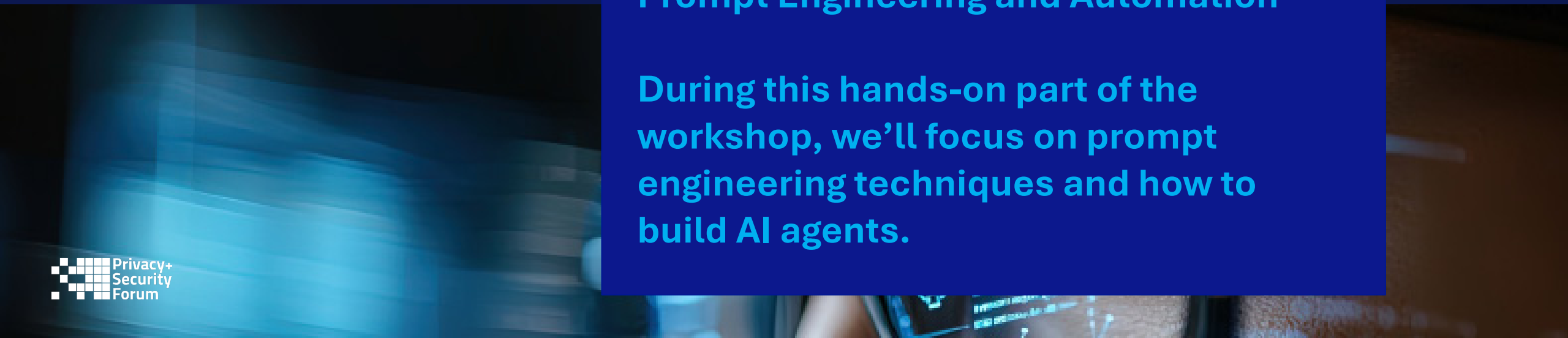
Timing	Topic
10:15 – 11:00	From Shadow AI Detection to Governed Agents: Prompt Engineering
11:00 – 11:30	Building AI Agents

Privacy+ Security Forum

From Shadow AI Detection to Governed Agents

Prompt Engineering and Automation

During this hands-on part of the workshop, we'll focus on prompt engineering techniques and how to build AI agents.



Privacy+ Security Forum

From Shadow AI Detection to Governed Agents

Shadow AI is a challenge for many organizations.

Technical approaches to detecting shadow AI include:

- Network Analysis
- CASB/SSE Discovery
- API Monitoring
- Browser/Endpoint Telemetry
- Expense Monitoring

Today's Focus

Expense Monitoring: How different types of increasingly sophisticated prompts perform for finding suspicious financial transactions

How an AI agent can be built and deployed

Privacy+ Security Forum

The Dude abides using generative AI…but only if it's approved. We've created a list of 27 example expense report line items, from employees across Lebowski Industries. Some are legitimate business expenses. The others? Yeah well, you know, that's just like your opinion, man.

Transaction ID	Employee Name	Transaction Date	Vendor	Description	Expense Category	Transaction Amount	Employee Department
EXP-001	Jeffrey 'The Dude' Lebowski	3/1/2026	Hollywood Star Lanes	Bowling alley rental for team building event	Team Building	$450.00	Facilities
EXP-002	Walter Sobchak	3/2/2026	Staples	Office supplies - binders, paper, pens	Office Supplies	$87.50	Operations
EXP-003	Maude Lebowski	3/3/2026	Blick Art Materials	Canvas and acrylic paint for office murals	Office Decor	$312.00	Creative
EXP-004	Donny Kerabatsos	3/4/2026	In-N-Out Burger	Team lunch - 12 combo meals	Meals & Entertainment	$156.00	Engineering
…	…	…	…	…	…	…	…

We'll compare three prompt engineering approaches against the same set of example expense data:

- **The Lazy Prompt:** Minimal instruction, maximum ambiguity
- **Chain of Thought (CoT):** Structured, multi-step reasoning
- **ReACT:** Reasoning + Acting with simulated tools and iterative passes

Privacy+ Security Forum

For each, we'll evaluate:

- **Detection Accuracy:** ***How many*** instances of shadow AI usage does each approach catch?
- **Explainability:** Can we understand ***why*** each instance was flagged?
- **Auditability:** Does the output include ***structured data*** we can store and review?
- **Cost:** What are the ***token and dollar costs*** of each approach?

But first…

Would any Dudeists from the audience like to try their own prompt?

Demo - Jupyter Notebooks

A Jupyter Notebook is an open-source web application used to create and share interactive documents that combine live code, computational results, rich text, and visualizations. It is a standard tool in data science, machine learning, and scientific research because it allows users to execute code in small "cells" and see the output immediately.

They are very useful for those with some background in programming, but don't worry. Later we'll show you techniques for building AI agents which require little to no coding.

Privacy+ Security Forum

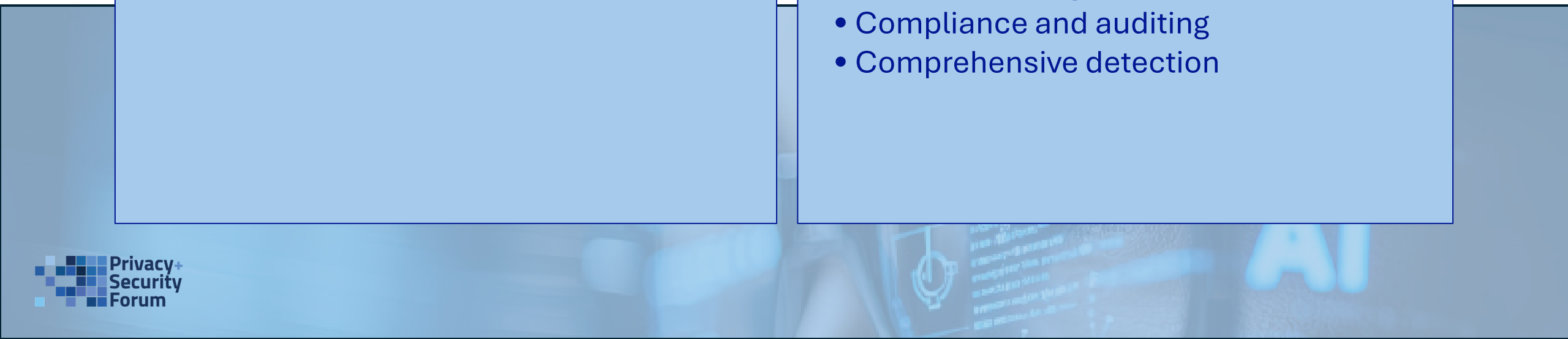
What we Tested

We evaluated three prompt engineering approaches against the same dataset of 27 expense entries containing 12 simulated shadow AI expenses (4 obvious, 3 moderate, 5 subtle) representing 10 distinct shadow AI schemes.

Key Takeaways

Prompt engineering is not just about getting answers - it's about getting the right answers in a format you can trust, explain, and audit for:

- Quick screening
- Compliance and auditing
- Comprehensive detection



Privacy+ Security Forum

Up next…

Governed Agents

ACT 1 — THE PROBLEM

Live Demo: An Agent in Minutes

We’re building an agent together to gain credibility in governance discussions.
You wouldn’t build an enterprise application, but you can build a personal agent in real time — pointed at a sanitized internal document repository.
No formal governance. No security review. Just a few clicks or lines of code.

The Scenario

Internal spend & contract review agent using a fictional finance document store

The Data Sources

Sanitized policy docs, mock employee handbook, fictional contract templates

The Point

Powerful agents can be created in under 10 minutes — with zero governance oversight

Privacy+ Security Forum

How Agentic Risk Differs from Traditional Software

1

Traditional App

Static data flows, defined permissions, human-initiated actions, predictable behavior

2

Early AI Tools

Generative outputs, human reviews results, limited autonomy, contained scope

3

AI Agents

Autonomous action, multi-source correlation, self-directed tool use, continuous operation

4

Agentic Networks

Agent-to-agent orchestration, compounding risk, emergent behavior, minimal human oversight

Audience Poll: Where Are You Today?

Poll Question

"Where is your organization in its AI agent journey?"

01

Heard of Agents

Aware but no formal activity yet

02

Piloting

Experimenting with 1–2 use cases, limited governance

03

Active Strategy

Deployed agents with some controls in place

04

Actively Using In Daily Work

Deployed agents with some controls in place

Vote: What agent would you see being deployed in your organization?

Low Risk

Internal knowledge assistant — answers questions from approved internal docs

Medium Risk

HR/Finance helper — summarizes policies, flags anomalies in employee or financial data

Higher Risk

Semi-autonomous operations agent — executes workflows, modifies records, sends notifications

Facilitator: Use poll results to calibrate depth — foundational guardrails for early-stage orgs; scaling and refinement for mature ones.

Privacy+ Security Forum

ACT 3 — GOVERNANCE DESIGN

Guardrail Design Session

Translate role-based concerns into concrete guardrails using the formula: **"This agent is approved if X, Y, and Z are true."**

1	**Data Access** Agent only accesses predefined, approved datasets — no ad-hoc data source connections
2	**Purpose Limitation** Agent operates strictly within its declared use case; scope changes require re-review
3	**Human-in-the-Loop** Agent cannot send external communications or modify records without human approval
4	**Logging & Auditability** All actions are logged, reviewable, and retained per data retention policy
5	**Incident Response** A kill-switch exists; responsible owner is identified and reachable 24/7

ACT 2 — WHY THIS IS DIFFERENT

Defining Agentic Risk

Privacy Risks

- Greater likelihood of discovering and aggregating sensitive or regulated data
- Increased re-identification risk at scale
- Cross-border data movement without adequate controls

Security Risks

- Prompt injection and adversarial manipulation
- Compromised tools enabling data exfiltration
- Lateral movement through APIs and integrations

Autonomy Risks

- Agents changing access rights without human review
- Sending external communications autonomously
- Modifying records that traditionally require approval

Existing DPIA and Threat Modeling approaches must be adapted — agentic systems introduce risks that traditional software frameworks were not designed to address.

CONCRETE GUARDRAILS

Guardrail #1: Approved Tools & Models

Only use **trusted, reviewed models and agent platforms**. The approved path must be fast, well-communicated, and easier than going rogue — this is the primary antidote to shadow AI.

Fast Approval Path Includes:

- Risk assessment and DPIA
- Vendor security review
- Documentation of data storage and processing locations
- Confirmation of enterprise controls support

Approval Criteria

- → Supports enterprise-grade logging and audit trails
- → Regional data residency options available
- → Robust role-based access control
- → Vendor security posture independently verified

CONCRETE GUARDRAILS

Guardrail #2: End-to-End Agent SDLC

Privacy, security, and compliance controls must be **baked in from the start** — not bolted on after deployment. Use this as a reusable checklist.

1. **Design**
   Data minimization, purpose specification, DPIA trigger assessment
2. **Develop**
   Least-privilege access, secure credential management, input validation
3. **Test**
   Red-teaming for prompt injection, data leakage testing, adversarial scenarios
4. **Deploy**
   Approved platform only, access controls confirmed, owner documented
5. **Monitor**
   Drift detection, behavior monitoring, anomaly alerting

Sunset phase: revoke all credentials, archive logs per retention policy, formally decommission the agent from inventory.

CONCRETE GUARDRAILS

Agent SDLC: Monitoring Through Sunset

Monitor

Continuous drift detection, behavioral anomaly alerts, usage pattern review

Maintain

Periodic re-review against updated regulations, model updates, scope changes

Sunset

Revoke credentials, archive logs, remove from agent inventory, notify stakeholders

The SDLC is a living document — revisit it whenever the agent's data sources, capabilities, or organizational context change significantly.

CONCRETE GUARDRAILS

Guardrail #3: Technical Enforcement & Visibility

Control Capabilities

Permissions & RBAC

Role-based access control scoped to each agent's declared purpose

Action Logging

Full audit trails of every agent action, queryable and tamper-evident

Anomaly Detection

Behavioral monitoring flags deviations from baseline agent activity

Kill-Switch

Immediate pause or revocation capability for any deployed agent

Why It Matters

Technical controls satisfy core privacy and security requirements:

- **Accountability** — clear ownership and traceable actions
- **Traceability** — full audit trail for regulatory inquiries
- **Incident Readiness** — rapid containment when something goes wrong

Tools like Microsoft 365 agent platforms, Cranium, and HiddenLayer provide these capabilities — align them to your guardrails.

CONCRETE GUARDRAILS

Guardrail #4: Policy Framework & Autonomy Levels

Every agent needs **clear ownership, defined risk thresholds, and explicit autonomy criteria.** Policy decisions must tie back to the Agent SDLC, approved tools, and technical enforcement layers.

Fully Autonomous

Operates within tight, pre-approved boundaries — highest technical controls required

Semi-Autonomous

Human approval required for high-impact actions (e.g., sending emails, modifying records)

Recommendation Only

Agent surfaces insights; all decisions made by humans — lowest risk tier

Privacy and security requirements scale with autonomy level. Higher autonomy = stricter DPIA, more rigorous logging, and mandatory human escalation paths.

Privacy+ Security Forum

From Shadow AI to Governed Agents: The Journey

Shadow AI

Uncontrolled, untracked agents created ad-hoc across business units

Awareness

Organization recognizes agent sprawl; begins informal inventory

Initial Controls

Approved platform defined; basic SDLC and ownership requirements established

Governed Agents

Full guardrail framework, autonomy tiers, cross-functional governance group active

Scaled Governance

Automated enforcement, continuous monitoring, policy as code across the enterprise

Most organizations today sit between **Shadow AI** and **Initial Controls**. This workshop gives you the tools to accelerate that journey.

Privacy+ Security Forum

Key Takeaways

1 Agent Sprawl Is Inevitable — and Different

Agents are not apps. They can be created in minutes, connected to sensitive data, and operate autonomously at scale — breaking every assumption of traditional app-centric governance.

2 Treat Agents as Amplified Users

The privacy and security risk is not just data access — it's the speed, scale, and depth of inference. Agents are tireless super-users that require a new risk lens.

3 Practical Guardrails Exist Today

Approved tools, an Agent SDLC, technical enforcement, and a clear policy framework give you a coherent, actionable governance model you can start building now.

Your Next Steps After This Workshop

1. Inventory

Audit existing agents and shadow AI usage across your organization — the footprint is larger than you think

2. Define Policy

Draft an initial agent policy with autonomy tiers, ownership requirements, and risk thresholds

3. Pilot

Stand up an approved agent platform with embedded controls for one low-risk use case

4. Govern

Establish a cross-functional governance group: Privacy, Security, IT, Legal, and a business representative

Start with the Agent SDLC checklist from this workshop — adapt it to your organization's risk tolerance and regulatory context before your next agent deployment.

Privacy+ Security Forum

The Agent SDLC Checklist

Take this checklist back to your organization. Adapt it to your context. Make it a mandatory gate before any agent goes to production.

Pre-Deployment Gates

- [ ] Purpose and data scope documented
- [ ] DPIA completed or scoped
- [ ] Approved platform confirmed
- [ ] Vendor security review passed
- [ ] Business and technical owner assigned
- [ ] Autonomy tier defined and approved
- [ ] Red-team / prompt injection testing completed

Operational Requirements

- [ ] Action logging active and verified
- [ ] Anomaly detection configured
- [ ] Kill-switch tested and documented
- [ ] Incident response plan in place
- [ ] Drift detection scheduled
- [ ] Periodic re-review cadence set
- [ ] Sunset plan documented

Thank you for your time and participation today.

To stay in touch, contact:

Karen Schuler
kschuler@bdo.com

Taryn Crane
tcrane@bdo.com

Privacy+ Security Forum