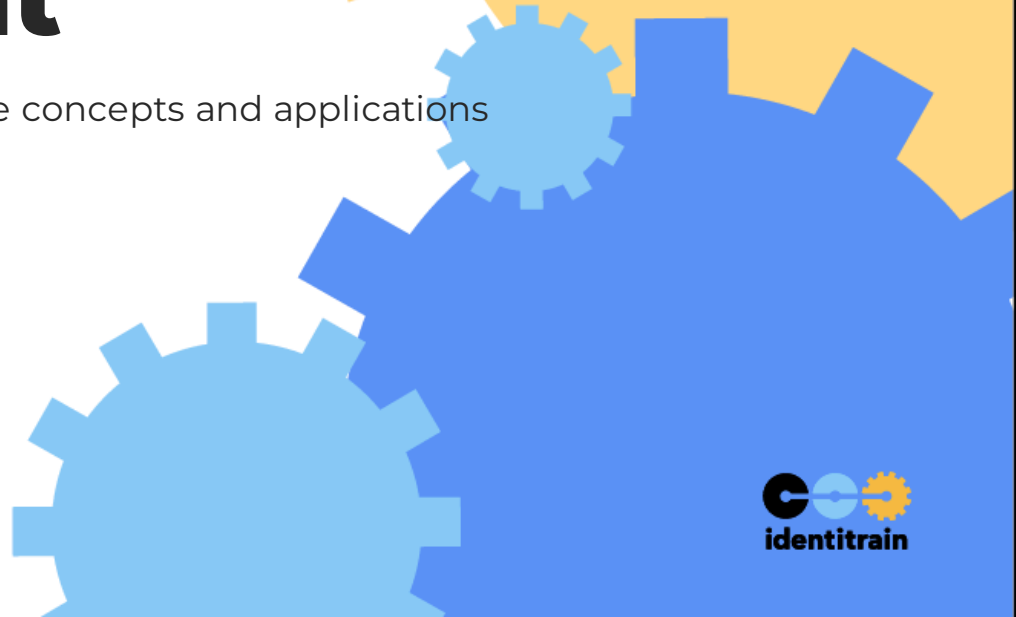
IDT-AI-101 - Leveraging Artificial Intelligence in Identity and Access Management

An introduction to artificial intelligence concepts and applications



identitrain

AI for IAM Professionals

Leveraging Artificial Intelligence in Identity and Access Management

- `<<insert: cover graphic: dark professional colour scheme with a stylised AI neural network branching into identity shield and access key icons; circuit-board lines connect the elements to suggest integration; the overall mood is modern, authoritative, and welcoming>>`

Welcome & Course Goals

- **Welcome!**
- Instructor Introduction (Name, Brief Background)
- **Course Goals:**
  - **Demystify AI:** Understand core AI concepts relevant to IAM — without deep technical jargon.
  - **Explore Applications:** Discover how AI is integrated into IAM platforms *today* and what's coming next.
  - **Learn Practical Skills:** Get hands-on with prompt engineering and AI-assisted IAM workflows.
  - **Assess Risks:** Understand bias, privacy risks, AI-specific threats, and the regulatory landscape as they affect identity systems.
  - **Look Ahead:** Explore emerging challenges: AI agents, non-human identity, shadow AI, and the evolving IAM professional's role.

identitrain

Who Is This Course For?

- **Target Audience:**
  - IAM Analysts & Administrators
  - IAM Engineers & Architects
  - Identity Governance & Compliance Professionals
  - IT Auditors involved in IAM controls
  - IAM Program & Product Managers
  - Anyone whose role involves managing or governing identities and access
- **No Prior AI Expertise Required!**
- **Focus:** Conceptual understanding and practical application in the IAM context.

identitrain

Course Format & Agenda

- **Duration:** ~3 Days
- **Style:** Presentations, demonstrations, hands-on labs, exercises, and scenario analysis.
- **21 Exercises:** Individual activities and end-of-chapter quizzes; 13 exercises use a live AI chatbot (free-tier tools: ChatGPT, Claude, Gemini, etc.).
- **High-Level Agenda:**
  - **Day 1 (AM):** AI Foundations: What is AI? Key concepts, types of learning.
  - **Day 1 (PM):** Working with AI: Prompt engineering, data safety, hands-on labs.
  - **Day 2:** AI in IAM Today: Platform AI features, AI copilots, cost & licensing, non-human identity.
  - **Day 3:** Shadow AI, Responsible AI & Road Ahead: Real-world incidents, bias, regulation, governance, future trends.
- `<<insert: horizontal three-day timeline: each day block contains the chapter title and two or three small representative icons (e.g., Day 1: brain + prompt bubble, Day 2: platform dashboard + agent icon, Day 3: shield + gavel); a progress arrow runs left to right beneath; clean flat style with the course colour palette>>`

identitrain

Why AI Matters for *Your* IAM Role

- **The Numbers Tell the Story:**
  - **70%** of security and identity professionals are actively pursuing AI qualifications (ISC2 2025).
  - **56%** wage premium for professionals with AI skills (PwC Global AI Jobs Barometer 2025).
  - **AI/ML** entered the top-5 in-demand skills across the industry for the first time.
- **This Is a Career Opportunity:**
  - AI is already embedded in the IAM platforms you use and evaluate *today*.
  - Understanding AI makes you more effective at governance, access management, and compliance.
  - Ability to critically evaluate vendor AI claims separates informed buyers from easy targets.
- **Not Optional — Essential:** The question is no longer *whether* to learn about AI, but how quickly you can apply it.

identitrain

The AI + IAM Landscape in 2025–2026

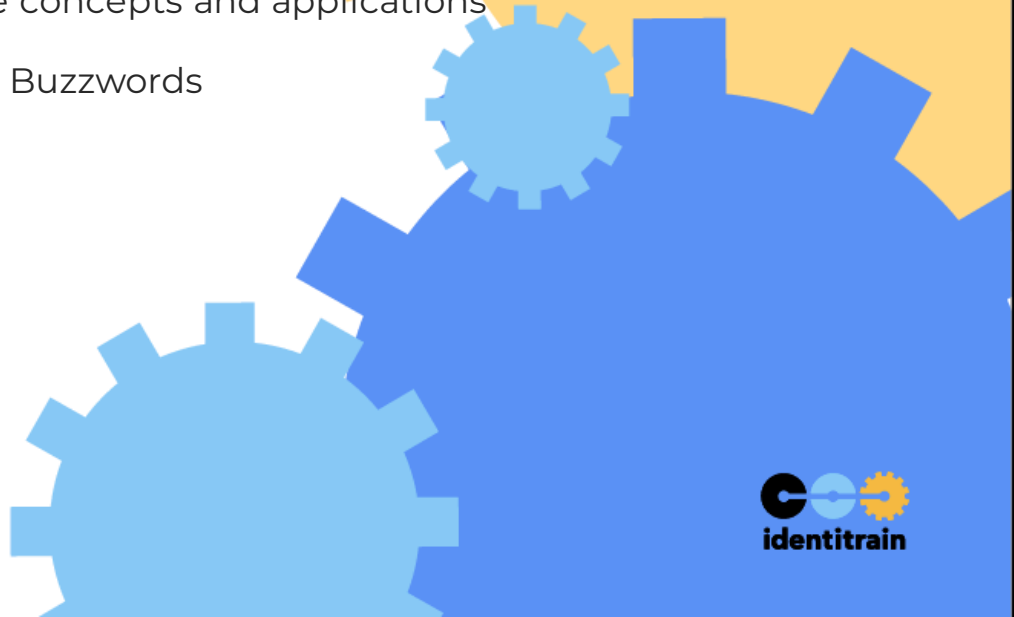
- **AI Is Now Core Infrastructure:** Every major IAM platform ships AI-powered features as standard. AI copilots provide natural-language interfaces to identity data.
- **New Identity Challenges:**
  - Non-human identities outnumber humans **50:1+** — and 97% are overprivileged.
  - **82%** deploying AI agents — only 44% have governance frameworks (SailPoint 2025).
  - Shadow AI: 80%+ of employees using unapproved AI tools with sensitive data.
- **Regulation Is Catching Up:**
  - EU AI Act high-risk obligations arriving **August 2026**.
  - Multi-jurisdictional compliance is the baseline expectation.
- **This Course Addresses This Reality.**

identitrain

AI Foundations for Identity and Access Management

An introduction to artificial intelligence concepts and applications

Understanding the Basics Without the Buzzwords



identitrain

What is AI? (Simply Put)

- **Artificial Intelligence (AI) = Systems that Mimic Cognitive Functions**
- Core Idea: Machines performing tasks that typically require human intelligence.
- **Key Capabilities:**
  - **Learning from Data:** Identifying patterns, relationships, and insights within information.
  - **Making Decisions/Predictions:** Using learned patterns to act or forecast future outcomes.
  - **Understanding Language/Images:** Processing and interpreting text, speech, or visual information.
- **Focus:** Not about conscious machines, but about *pattern recognition* and *task automation* at scale.
- `<<insert: simple three-step flow diagram: "Data In" box on left → central "AI Processing" box with gear/brain icon → "Decision / Prediction / Action" box on right; clean connecting arrows between steps; flat design style>>`

identitrain

The IAM Challenge & AI's Role

- **The Modern IAM Challenge:**
  - Increasing complexity (cloud, hybrid IT, more applications & identities to manage).
  - Explosive growth in non-human identities (service accounts, API keys, AI agents).
  - Governance at scale (access reviews, compliance, role management across thousands of users).
  - Balancing strong controls with smooth user experience.
- **AI's Role:** A force multiplier, helping IAM teams manage scale, complexity, and speed more effectively.

identitrain

How AI Helps IAM: Security, Efficiency, UX & Compliance

- **Security:**
  - Detecting sophisticated threats: anomalies in user behaviour that rules-based systems miss.
  - Assessing risk in real time to drive adaptive access decisions.
- **Efficiency:**
  - Automating data-intensive tasks: access review recommendations, log analysis, policy drafting.
  - Accelerating onboarding with AI-powered identity verification.
- **User Experience:**
  - Adaptive authentication: less friction for trusted users, stronger challenges when risk is high.
  - AI copilots providing natural-language access to identity data and workflows.
- **Compliance:**
  - Continuous policy analysis: spotting misconfigurations and toxic access combinations.
  - Identifying outlier entitlements and dormant accounts at scale.

identitrain

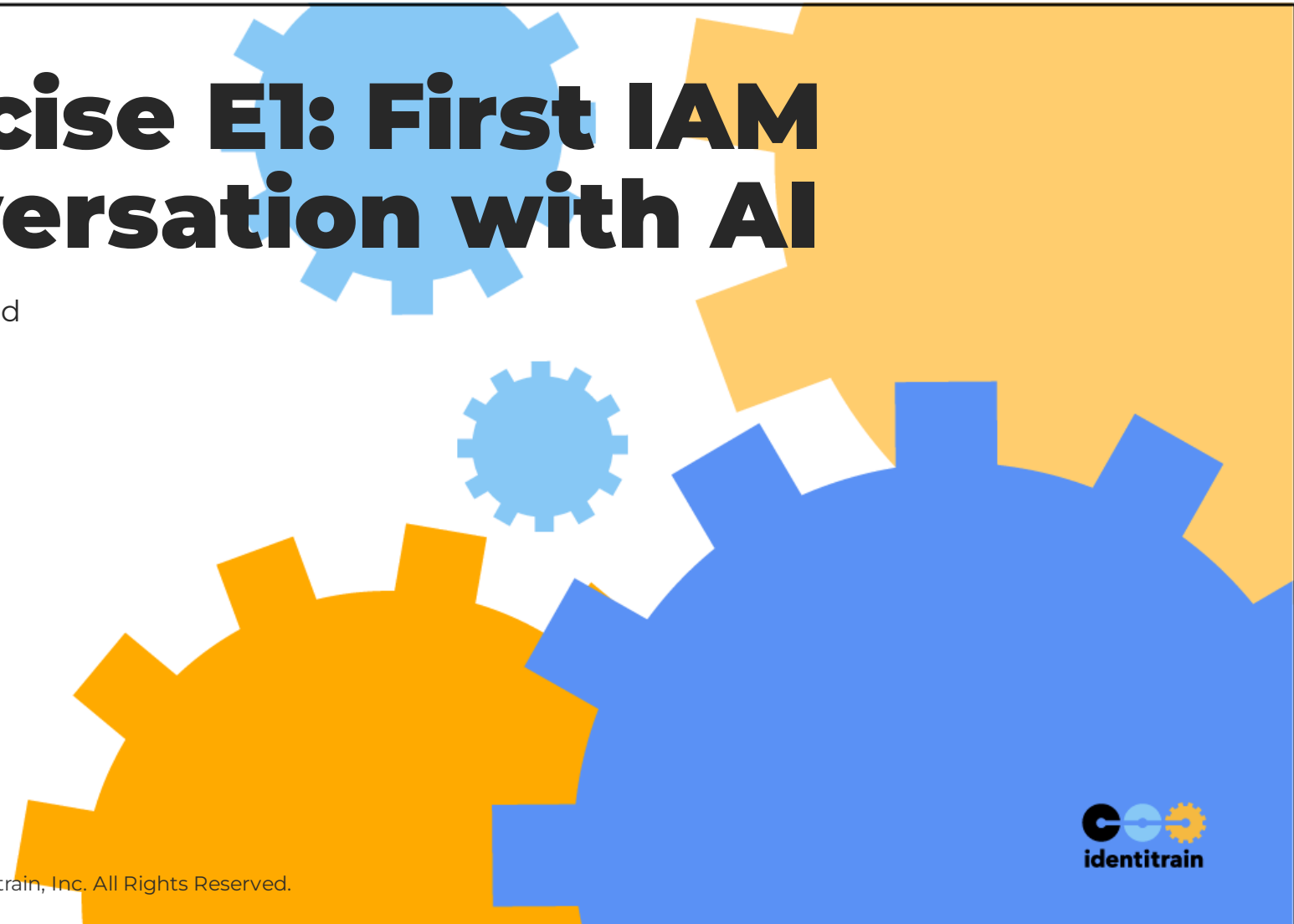
AI in IAM: Concrete Examples

- **Anomaly Detection:** AI flags impossible travel: a user logging in from London and Singapore within 30 minutes.
  - *Real incident:* Arup deepfake video call fraud, $25.6M loss, Hong Kong, 2024. Traditional controls couldn't detect the deception.
- **Identity Verification:** AI verifies ID documents and matches selfies during remote onboarding.
  - *Current reality:* Entrust reports a deepfake attack against identity verification systems every 5 minutes (2025).
- **Access Review Intelligence:** AI recommends approve/revoke based on usage and peer comparison.
  - *Shipping product:* SailPoint Harbor Pilot: AI agents for identity governance, GA March 2025.
- **Adaptive Authentication:** AI adjusts login requirements based on real-time risk assessment.
  - *Shipping product:* PingOne Risk: contextual risk scoring driving adaptive policies.

identitrain

Exercise E1: First IAM Conversation with AI

See: Chapter_1.md



Understanding the Engines: Key AI Concepts

- Peeking Under the Hood (Without Needing an Engineering Degree)
- Introducing Neural Networks and Large Language Models

Introducing Artificial Neural Networks (ANNs)

- **Inspired by the Brain:** Computer systems loosely modelled on biological neurons and their connections.
- **Core Idea:** Learning complex patterns from data through interconnected layers of simple processing units ("nodes").
- Think of layers processing information sequentially, like stages in an assembly line.
- `<<insert: simplified neural network diagram: three vertical columns of circles labelled "Input Layer", "Hidden Layers", "Output Layer"; lines connect every circle in one column to every circle in the next; arrows indicate data flow left to right; clean textbook-style illustration>>`

identitrain

How Neural Networks Learn (Conceptually)

- **Learning Process:**
  1. **Input:** Data is fed into the first layer.
  2. **Processing:** Nodes in each layer perform calculations and pass signals forward.
  3. **Output:** The final layer produces a result (e.g., a prediction, classification).
  4. **Feedback:** The result is compared to the "correct" answer (if available).
  5. **Adjustment:** Internal parameters are tuned to improve accuracy.
- **"Strength of Connections" (Weights):** Imagine dials on each connection. During learning, the AI slightly turns these dials up or down based on feedback, making influential connections stronger.
- **"Node Threshold" (Biases):** Think of this as a threshold each node has. If the incoming signal is strong enough, the node "fires" and passes the signal forward.
- `<<insert: annotated neural network close-up: a few nodes and connections are highlighted; each connection has a small "weight dial" icon; each node shows a "bias threshold" slider; callout labels explain "Weight = how much influence" and "Bias = threshold to fire"; builds on the previous NN diagram>>`

identitrain

Why NNs Matter for IAM

- **Key Strength:** Excellent at finding intricate, non-obvious patterns in large, complex datasets.
- **IAM Relevance: The Pattern Recognition Powerhouse**
  - **Anomaly Detection (UEBA):** Identifying subtle deviations from normal user behaviour patterns.
  - **Biometrics:** Analysing facial features, voice patterns, or typing rhythm for identity verification.
  - **Risk Scoring:** Calculating access risk by combining many complex contextual factors.
  - **Peer Group Analysis:** Identifying users with outlier access compared to similar roles.
- NNs often power the core "intelligence" behind AI-driven IAM platform features.

identitrain

Introducing Large Language Models (LLMs)

- **What are they?** Specialised, extremely large Neural Networks.
- **Training Data:** Trained on *enormous* amounts of text and code (internet, books, etc.).
- **Core Capability:** Highly skilled at predicting the next word (or "token") in a sequence.
  - Given "The cat sat on the...", it predicts "mat" (or "roof", "chair", etc.).
- This core skill unlocks broader language understanding and generation.
- `<<insert: LLM prediction diagram: left side shows a partial sentence "The cat sat on the…" entering an LLM box; right side shows a horizontal bar chart of candidate next tokens ("mat" 62%, "roof" 18%, "chair" 12%, …) ranked by probability; clean infographic style>>`

LLMs: Emergent Abilities & Examples

- **Emergent Abilities (from predicting next word):**
  - Generating human-like text (emails, reports, policies).
  - Answering questions based on training data.
  - Summarising long documents accurately.
  - Translating between languages.
  - Writing computer code in various languages.
  - Engaging in coherent conversations (chatbots).
- **Well-known Examples:**
  - OpenAI's GPT series (powering ChatGPT)
  - Anthropic's Claude series
  - Google's Gemini
  - Meta's Llama series
  - Mistral AI's models

identitrain

Why LLMs Matter for IAM

- **IAM Relevance: The Language Specialists**
- Primarily power **Generative AI** tasks in IAM:
  - **Drafting:** Access policies, procedures, user guides, compliance documentation.
  - **Summarising:** Audit logs, access review data, compliance reports (requires careful data handling!).
  - **Explaining:** Complex policies or regulatory requirements in plain language.
  - **Interacting:** Powering chatbots for user self-service (access requests, password help) or analyst assistance.
  - **AI Copilots:** Enabling natural-language interfaces to IAM platforms: "Show me all users with admin access who haven’t logged in for 90 days."

identitrain

AI Concepts Recap + Other Models

- **Neural Networks (NNs): The Pattern Finders**
  - Finding complex patterns in data → UEBA, biometrics, risk scoring, peer analysis.
  - Think: *Analysis and detection* based on learned patterns.
- **Large Language Models (LLMs): The Language Specialists**
  - Understanding and generating text → policy drafting, summarisation, copilots, chatbots.
  - Think: *Language tasks:* creation, summarisation, interaction.
- **Other Approaches (Brief Mention):**
  - *Diffusion Models:* Generate realistic images. IAM relevance: understanding deepfake threats to identity verification.
  - *Reinforcement Learning:* AI learns through trial and error. Potential future use: optimising adaptive authentication strategies.
- **Key Takeaway:** "AI" isn't monolithic. Different tools for different jobs.

identitrain

Exercise E2: Evaluate Vendor AI Claims with AI

See: Chapter_1.md

How AI Learns: Types of Learning

- Different Approaches for Different Problems
- Supervised, Unsupervised, and Generative AI
- `<<insert: three equal panels side by side: Panel 1 "Supervised": labelled data examples (/ tags) flowing into a model; Panel 2 "Unsupervised": untagged data points being grouped into clusters by a model; Panel 3 "Generative": a text prompt entering a model that outputs new content; consistent visual style across all three>>`

identitrain

Supervised Learning: Learning with Labels

- **Concept:** AI learns from data where humans have provided the "correct answers."
- **Analogy:** Flashcards: the AI sees questions with known answers and learns the patterns.
- **How it Works:** Feed labelled examples (e.g., logins marked "Fraudulent" / "Legitimate") → AI learns patterns → accurately labels *new, unseen* data.
- **IAM Examples:**
  - **Phishing Detection:** Training on emails labelled "Phishing" / "Not Phishing."
  - **Known Fraud Patterns:** Flagging logins matching historically labelled fraud cases.
  - **Compliance Prediction:** Training on access requests labelled "Compliant" / "Non-Compliant."
- **Key Strength:** Effective for **classification** and **prediction** when you have reliable, labelled historical data.
- `<<insert: scatter plot with data points colour-coded "Fraudulent" (red) and "Legitimate" (green) feeding into an AI model box; the model outputs a curved decision boundary separating the two classes; clean 2D chart style>>`

identitrain

Unsupervised Learning: Finding Hidden Patterns

- **Concept:** AI learns from data *without* labels; discovers structure on its own.
- **Analogy:** Exploring a new city without a map, where you notice patterns and unusual spots yourself.
- **Key Technique: Anomaly Detection:** finding data points that differ significantly from the "norm."
- **Major IAM Relevance: UEBA**
  - Learns baseline "normal" for each user/entity; flags significant deviations (unusual hour, unfamiliar resources, atypical API calls).
- **Other IAM Uses:**
  - **Role Mining:** Clustering users with similar access → discovering potential roles.
  - **Outlier Detection:** Users with excessive entitlements vs. peers.
  - **Novel Threats:** Detecting *previously unseen* attack patterns.
- **Key Strength:** Discovers what you didn't know to look for. Most IAM-relevant learning type.
- `<<insert: scatter plot of unlabelled data points; the AI has drawn dashed boundaries around three natural clusters and circled two isolated outlier points in red; a legend labels "Cluster" and "Outlier"; visually conveys pattern discovery without prior labels>>`

identitrain

Generative AI: Creating New Content

- **Concept:** AI models designed to *create* new, original content: text, images, code, audio.
- **How:** Learns patterns and structure from vast training datasets, then generates novel content that follows those patterns.
- **Underlying Technology:** Primarily powered by LLMs (for text) and other neural network architectures (for images, audio).
- **IAM Relevance (Significant & Growing):**
  - **Content Creation:** Drafting access policies, procedures, training materials, FAQs.
  - **Summarisation:** Condensing audit logs, access review data, compliance reports (data privacy is CRITICAL; covered in Chapter 2).
  - **Interaction:** Powering chatbots for user self-service and AI copilots for IAM professionals.
  - **Data Augmentation:** Generating synthetic data for testing identity systems.
- **Key Characteristic:** Focus on *creating and synthesising* information, not just analysing what exists.

identitrain

Choosing the Right AI Type

- **It Depends On the Problem and the Data:**
  - **Want to classify into known categories?** (phishing vs. not, compliant vs. not) → **Supervised Learning** (needs labelled data)
  - **Want to find unknown patterns or outliers?** (novel threats, role discovery, outlier access) → **Unsupervised Learning** (works with raw data)
  - **Want to create new content?** (draft policy, summarise logs, copilot query) → **Generative AI** (needs prompt/input)
- **Real-World Systems Often Combine Approaches:**
  - Unsupervised finds anomalies → humans label them → supervised learns to spot them faster.
  - Generative AI drafts a policy → humans review → refined policy feeds governance system.
- **You Don't Need to Choose, But Understanding Helps You:**
  - Evaluate vendor claims ("What type of AI powers this feature?")
  - Understand limitations ("This needs labelled data — do we have that?")
  - Ask better questions during procurement and architecture reviews.

identitrain

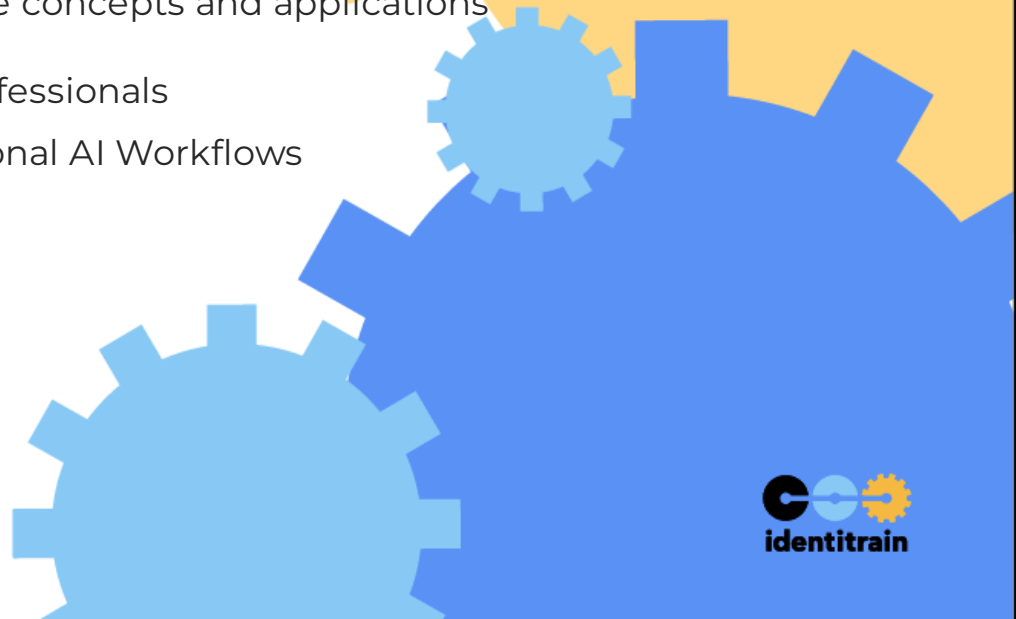
Exercise E3: Match the IAM Scenario to the AI Type

See: Chapter_1.md

Working with AI: From Prompts to Context Engineering

An introduction to artificial intelligence concepts and applications

- Mastering AI Interaction for IAM Professionals
- From Effective Prompts to Professional AI Workflows



identitrain

From Prompts to Context Engineering

- **Prompt Engineering (2023-2024):** Craft the perfect question, get a good answer.
- **Context Engineering (2025-2026):** Design the *entire information environment* the AI works within.
- **Context = Everything the AI Knows When You Speak:**
  - **System instructions:** Persistent rules and persona ("You are an IAM policy reviewer...")
  - **Retrieved data:** Documents, policies, logs fed via RAG
  - **Tool definitions:** What external systems the AI can query or act on
  - **Conversation history:** Prior messages that shape understanding
  - **Your prompt:** The specific question or instruction *right now*
- **The Prompt Is Still Important** – but it's one piece of a larger picture.
- **This Chapter:** Learn to write effective prompts AND configure the full context for professional results.
- `<<insert: concentric circles diagram: innermost circle labelled "Your Prompt", surrounded by rings labelled "Conversation History", "System Instructions", "Retrieved Data (RAG)", "Tool Definitions"; the full diagram is labelled "Context"; clean, minimal style>>`

identitrain

The Golden Rule: Data Privacy with AI Tools

- **NEVER Input Sensitive or Confidential Data into Public AI Tools!**
- **What is Sensitive in IAM Context?** PII (usernames, emails), internal system names/IPs, credentials, API keys, incident details, proprietary policies, raw logs, access review data.
- **Why?** Potential data retention by providers, use for model training, accidental exposure, regulatory violations.
- **The Scale of the Problem:**
  - 80%+ of employees use unapproved AI tools at work (shadow AI).
  - 34.8% of employees enter sensitive data into AI tools.
  - Source code is the #1 data type leaked to AI tools at 18.7% of sensitive inputs.
- **Regulatory Exposure:** GDPR (EU), CCPA (US-CA), LGPD (Brazil), PIPL (China), PDPA (Singapore): all apply to personal data entered into AI tools.
- **The Only Safe Approach:** Anonymise, sanitise, generalise. Use conceptual descriptions or redacted/mocked-up data ONLY. If unsure, DON'T INPUT IT.
- `<<insert: large red-orange alert triangle icon centred on the slide, with bold white text inside reading "STOP: NO SENSITIVE DATA"; beneath it a single-line caption: "Public AI tools are not safe for confidential information"; high visual impact, impossible to miss>>`

identitrain

The 2026 AI Toolkit: What's Changed

- **Reasoning / Thinking Modes:** All major platforms now offer modes where the AI "thinks harder" — allocating extra compute to reason step by step before responding.
- **Multimodal Input:** Share images, screenshots, documents, and diagrams alongside text. Ask the AI to analyse what it sees.
- **Million-Token Contexts:** Process entire policy frameworks, codebases, or audit logs in a single conversation (1M+ tokens on leading platforms).
- **Structured Outputs:** Request machine-readable responses (JSON, tables, specific schemas). AI output can feed directly into downstream tools and workflows.
- **Tool Use & Function Calling:** AI can query databases, call APIs, search the web, and execute actions within defined boundaries.
- **Key Takeaway:** AI tools in 2026 are fundamentally more capable than even 18 months ago. Understanding these capabilities changes how you use them.
- `<<insert: five capability icons arranged horizontally: brain-with-gears (reasoning), eye-with-image (multimodal), document-stack (long context), JSON-brackets (structured output), tool-wrench (tool use); each with a short label beneath; clean infographic style>>`

identitrain

The Foundations: From Zero-Shot to Meta-Prompting

- **These are the building blocks** that modern AI systems automate and build upon. Understanding them helps you debug when things go wrong.
- **Zero-Shot:** Ask directly, no examples. Relies on AI's training data. Good for straightforward tasks.
- **Few-Shot:** Provide 1-5 input/output examples to demonstrate the desired pattern. Dramatically improves consistency and format control.
- **Option Generation & Evaluation:** Ask for multiple solutions, then ask the AI to compare them. Powerful in sequence for decisions with trade-offs.
- **Meta-Prompting:** Ask the AI to help you write a better prompt. Bootstraps your way into complex tasks.
- **IAM Examples:**
  - Zero-shot: `"Explain the principle of least privilege."`
  - Few-shot: Give 2 examples of "technical finding -> user notification," then ask for a new one.
  - Options: `"Propose three approaches for temporary emergency access. Evaluate each on security risk vs. effort."`

identitrain

Prompt Chaining: Breaking Down Complex Tasks

- **Concept:** Break a complex task into sequential steps. Use output from one step as input for the next.
- **Why It Matters:**
  - Tackles problems too complex for a single prompt.
  - Keeps the AI focused on one manageable task at a time.
  - Enables human review and correction at intermediate stages.
- **IAM Example – Analysing a Security Alert:**
  - **Step 1:** `"Summarise the key facts in this sanitised alert: [paste data]"`
  - **Step 2:** *(Using Step 1 output)* `"Identify 2-3 potential security risks or policy violations."`
  - **Step 3:** *(Using Step 2 output)* `"Suggest immediate actions and one longer-term remediation."`
- **The Bigger Picture:** Prompt chaining is the manual version of what agentic AI systems do automatically. The AI agents you'll encounter in Chapter 3 chain reasoning and tool calls in autonomous workflows. Understanding chaining helps you understand and govern those agents.

identitrain

RAG, MCP & Enterprise AI

- **The Core Problem:** Public AI tools don't know your organisation's policies, users, or configurations.
- **RAG (Retrieval-Augmented Generation):**
  - System *retrieves* relevant data from YOUR knowledge base, feeds it + your query to the LLM. Answer *grounded in your actual data*.
  - Powers: Microsoft Security Copilot, SailPoint Harbor Pilot, and most enterprise AI copilots.
- **MCP (Model Context Protocol):** Open standard for connecting AI to external data sources, tools, and workflows — **"USB-C for AI."**
  - Adopted by all major AI providers. Defines how AI agents connect to your systems — a critical IAM governance surface.
- **Key Takeaway:** RAG makes AI organisation-aware. MCP standardises how it connects. Both directly relevant to IAM.
- `<<insert: two-part diagram: left side shows RAG flow (query → retrieval from knowledge base → LLM → grounded response); right side shows MCP architecture (AI Host connected via MCP Client to multiple MCP Servers providing tools and data); a bridge arrow connects the two, labelled "Enterprise AI Architecture"; clean, professional style>>`

identitrain

Why Prompt Quality Matters – and Common Pitfalls

- **In IAM, precision is non-negotiable.** Vague prompts → vague outputs. In access governance and compliance, vague means wrong.
- **Common Pitfalls:**
  1. **Ambiguity:** `"Tell me about access control"` — RBAC? In AWS? For what audience?
  2. **Missing context:** `"Write an S3 bucket policy"` — For whom? What data? Produces insecure output.
  3. **Implicit assumptions:** `"Summarise the latest incident"` — AI doesn't know your internal context.
  4. **Too much at once:** `"Redesign our IAM architecture"` — Break into steps (chaining).
  5. **No format/audience:** Getting an essay when you need bullet points for executives.
- **The Fix:** Be specific. Provide context. Define format. Break complexity into steps.
- `<<insert: two-column layout: left column "Poor Prompt" shows a vague one-line prompt with red X; right column "Better Prompt" shows a specific, context-rich prompt with green checkmark; bold visual contrast>>`

identitrain

Exercise E4: From Vague to Specific: Refining an IAM Prompt

See: Chapter_2.md

Thinking Modes: Making AI Reason Through Complex Problems

- **What Are Thinking Modes?** AI can allocate extra compute to reason step by step before responding, rather than generating an immediate answer. All major platforms now offer this.
- **How It Works:** The model "thinks" internally (hidden scratchpad / chain-of-thought reasoning), then provides a more thorough final answer.
  - Most platforms let you control the **effort level** or **thinking budget** — from quick reasoning to deep, extended analysis.
- **When to Use:** Complex analysis with multiple variables or competing requirements — multi-regulation compliance mapping, risk assessments, vendor evaluation, policy gap analysis.
- **When NOT to Use:** Simple lookups, routine drafting, straightforward questions. Thinking modes are slower and use more resources.
- **Key Insight:** For complex analytical work, thinking modes produce dramatically better results than standard prompting.

identitrain

Structured Outputs & Multimodal Input

- **Structured Outputs: Machine-Readable Results**
  - Request responses in specific formats: JSON, CSV, tables, YAML.
  - AI output can feed directly into SIEM tools, GRC platforms, or provisioning workflows.
  - *IAM Example:* `"Analyse these 5 access review findings and return a JSON array where each object has: finding_id, risk_level (high/medium/low), affected_roles, and recommended_action."`
- **Multimodal Input: Beyond Text**
  - Share screenshots, diagrams, PDFs, and photos alongside your text prompts.
  - The AI analyses what it sees and responds accordingly.
  - *IAM Examples:*
    - Screenshot a Conditional Access policy configuration and ask AI to identify coverage gaps.
    - Share an architecture diagram and ask AI to map data flows for compliance.
    - Upload a vendor security questionnaire (PDF) and ask AI to draft responses.
- **Key Takeaway:** Text-in, text-out is just the starting point. Structured outputs and multimodal input turn AI into a workflow component, not just a writing assistant.

identitrain

System Instructions & AI Projects

- **The Idea:** Instead of repeating context in every prompt, configure it once and reuse it.
- **All Major Platforms Offer Persistent Workspaces:**
  - **Claude Projects:** System instructions + uploaded knowledge documents. Strong privacy model.
  - **Custom GPTs (OpenAI):** Task-specific assistants with instructions, knowledge files, and tool integrations.
  - **Gemini Gems:** Configured AI within Google Workspace, integrated with Docs/Sheets/Drive.
- **What This Means for IAM:**
  - Create an "IAM Policy Reviewer" with your governance framework as context.
  - Build a "Compliance Analyst" that knows your regulatory obligations.
  - Set up an "Access Review Assistant" with your certification criteria pre-loaded.
- **Key Insight:** System instructions are the most powerful lever in context engineering. Configure the AI's environment once, benefit every time.

identitrain

Exercise E5: Hands-On: Draft a Policy Section with AI

See: Chapter_2.md

Exercise E6: Prompt Chaining: Security Alert Analysis

See: Chapter_2.md

Building Reusable Templates

- **Why Templates?** If you do a task regularly, build a prompt template once and reuse it every time.
- **Four Components of a Good Template:**
  - **Purpose:** What is this template for? (One sentence.)
  - **Context fields:** Placeholders you fill in each time: `[Platform]`, `[Role Name]`, `[Compliance Framework]`, `[Audience]`.
  - **Safety checklist:** Verify before every use: no PII, no system names, no credentials, output reviewed by human.
  - **Example prompt:** A completed example showing all fields filled with safe, anonymised content.
- **IAM Template Example:**
  - *Purpose:* Draft a quarterly access review summary for management.
  - *Fields:* `[Department]`, `[Review Period]`, `[Total Entitlements]`, `[Key Findings -- anonymised]`, `[Audience]`
  - *Safety:* No real names. No system names. No credentials. Human review before distribution.
- **From Template to System Instruction:** A good template can become the system instructions for an AI Project. Configure once, use for every conversation.

identitrain

From Chat to Workflow: Building Professional AI Processes

- **The Mindset Shift:** Stop "chatting with AI." Start building AI workflows.
- **The Right Balance:**
  - **Leverage AI's strengths:** reasoning, analysis, language understanding, pattern recognition.
  - **Apply constraints:** system instructions, safety checklists, structured output schemas, human review checkpoints.
  - Too few constraints → unreliable. Too many → you've built a rigid form, not an AI workflow.
- **Workflows Are Living Processes:**
  - Start simple: a template for a common task. Add constraints as you learn. Refine as tools evolve.
  - Document workflows so your team can use and improve them. Embed safety into the process itself.
- **Your Goal:** Repeatable, safe, improvable AI-assisted processes — not one-off experiments.

identitrain

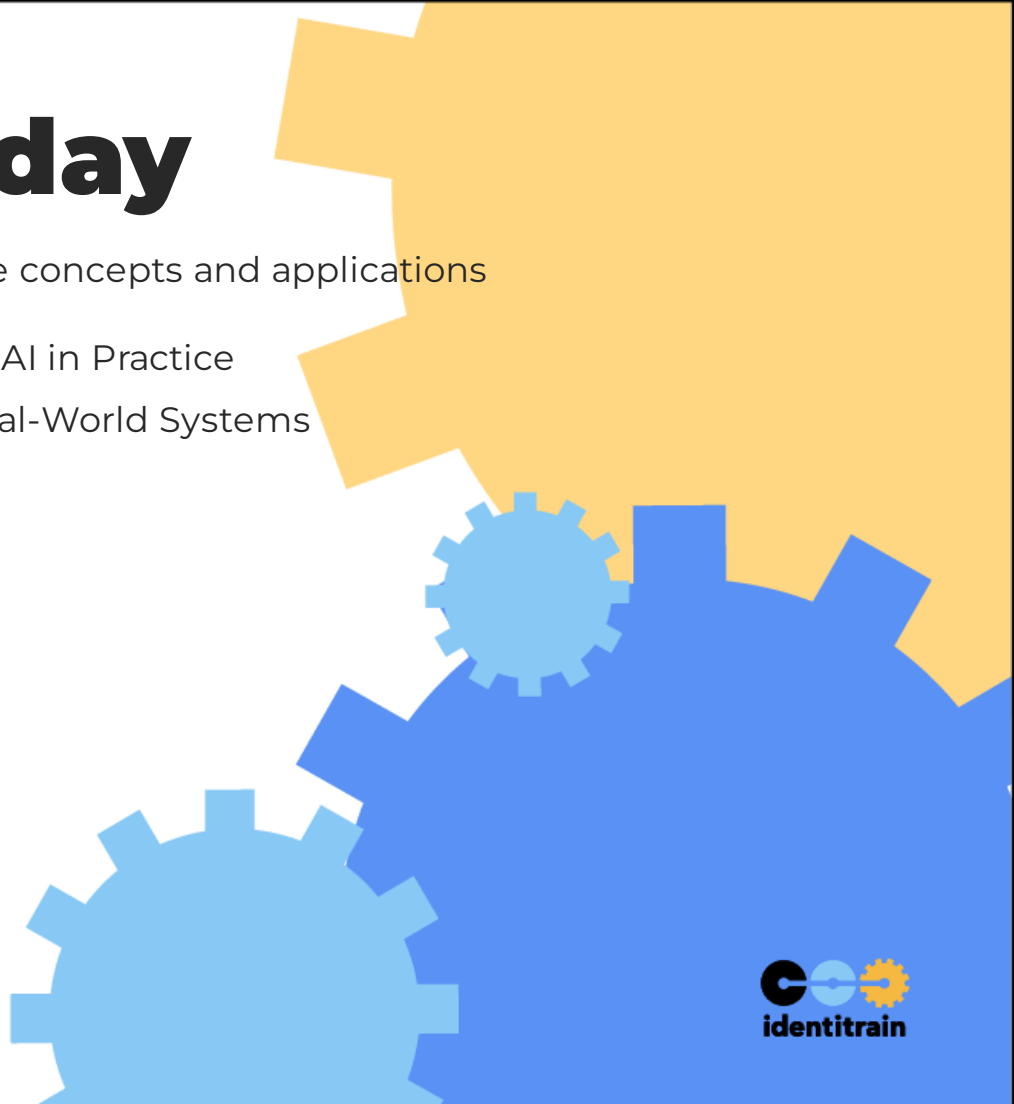
Exercise E7: Build Your IAM Prompt Template

See: Chapter_2.md

AI in IAM Today

An introduction to artificial intelligence concepts and applications

- Where the Rubber Meets the Road: AI in Practice
- From Concepts to Capabilities in Real-World Systems



identitrain

AI Integrated into IAM: Platforms and Capabilities

- Connecting Chapter 1 Concepts to Real Platform Features
- UEBA, Adaptive Auth, IDV, IGA, Policy Analysis: Where They Live
- Every Major Vendor Now Ships AI as Core Infrastructure

Threat Detection & Risk Scoring (UEBA)

- **Focus Area:** Access Management Security
- **Capability:** AI analyses context signals in real time to detect anomalies and calculate risk scores
- **Context Signals:** Location, device fingerprint, time of day, IP reputation, user behaviour history, peer group activity, impossible travel
- **How It Appears:**
  - "Risk Score" (Low / Medium / High) in authentication logs and dashboards
  - Alerts: "Impossible Travel," "Anomalous Location," "Unfamiliar Device," "Unusual API Activity"
- **Why It Matters Now:** Traditional rule-based detection cannot keep pace with attacks like the Arup deepfake video call ($25.6M loss, Hong Kong, 2024). AI-driven behavioural analysis is the primary defence layer
- **Platform Examples:** Microsoft Entra ID Protection, PingOne Risk, Okta Behavior Detection, AWS GuardDuty
- `<<insert: diagram showing six context signals (Location, Device, Time, IP Reputation, User History, Peer Activity) as labelled arrows feeding into a central "AI Risk Engine" box, which outputs a colour-coded risk score gauge (Low/Medium/High); clean architectural style>>`

identitrain

Adaptive Authentication / Risk-Based MFA

- **Focus Area:** Access Management: Security + User Experience
- **Capability:** Uses AI-driven risk scores to dynamically adjust authentication requirements
- **How It Appears:**
  - Policy rules: "IF Risk = High THEN Block" / "IF Risk = Medium THEN Require MFA" / "IF Risk = Low AND Known Device THEN Passwordless"
  - Less friction for trusted contexts, stronger controls for risky ones
- **The Flow:** Context Signals → AI Risk Engine → Risk Score → Adaptive Policy Decision (Allow / Step-Up / Block)
- **Platform Examples:** Microsoft Entra Conditional Access (using sign-in risk levels), Okta Adaptive MFA, PingOne adaptive policies, AWS Cognito Adaptive Authentication
- `<<insert: horizontal flow diagram: "Context Signals" box → "AI Risk Engine" box → "Risk Score" diamond → three branches: green arrow to "Allow / Passwordless", amber arrow to "Step-Up MFA", red arrow to "Block"; each branch clearly labelled with the condition>>`

identitrain

Identity Verification (IDV)

- **Focus Area:** Identity Management: Onboarding & Proofing
- **Capability:** AI verifies government-issued ID documents and matches them to user selfies
- **AI Components:**
  - Document authenticity analysis (security features, tampering detection)
  - Facial recognition (comparing document photo to selfie)
  - Liveness detection (ensuring real person, not photo/video/mask)
- **The Threat Is Escalating:**
  - Deepfake attacks on IDV systems now occur **every 5 minutes** (Entrust 2025)
  - Digital document forgeries up **244% year-over-year**, now **57% of all document fraud**
- **Platform Examples:** Integrations with Jumio, Onfido, iProov, Trulioo via APIs in Entra Verified ID, Cognito, CIAM platforms
- `<<insert: graphic showing an ID document and a selfie photo entering an "AI Verification Engine" box with three check stages (Document Authenticity → Facial Match → Liveness Detection); output is a Confidence Score with Pass/Fail indicator; a red "Deepfake Detection" shield overlays the liveness stage>>`

identitrain

Intelligent Identity Governance (IGA)

- **Focus Area:** Identity Governance & Administration
- **Capability:** AI analyses user entitlements, roles, access usage, and peer activity to improve governance
- **How It Appears:**
  - **Access Review Recommendations:** Flagging risky or outlier access, suggesting approve/revoke ("AI Recommendation: Revoke. Unused 90+ days, low peer usage")
  - **Role Mining Suggestions:** Identifying potential new roles by clustering users with similar access patterns
  - **Outlier Detection:** Highlighting users with excessive or unusual permissions vs. peers
- **Why It Matters:** Combats "rubber-stamping": reviewers approving everything because the volume is overwhelming
- **Platform Examples:** SailPoint Harbor Pilot and Predictive Identity, Saviynt Intelligent Analytics, ForgeRock Autonomous Identity

identitrain

Automated Policy Analysis

- **Focus Area:** Governance, Risk Management, Compliance
- **Capability:** AI and automated reasoning analyse access policies to find risks, misconfigurations, and violations
- **What It Finds:**
  - Overly permissive rules granting broader access than intended
  - Unintended external or public access to internal resources
  - Separation of duties violations and toxic permission combinations
  - Unused access that violates least privilege
- **Platform Examples:** AWS IAM Access Analyzer (external/public access checks, unused access identification, policy generation from CloudTrail), capabilities within CSPM tools analysing IAM configurations
- **Key Distinction:** Automated reasoning applies logical rules exhaustively; it provides mathematical certainty, not probabilistic results

identitrain

Platform Spotlight: Ping Identity

- **PingOne Risk:** UEBA with contextual risk scoring, impossible travel detection, bot detection
- **PingOne Protect:** Broader threat detection across the authentication journey: fraud signals, device intelligence, IP reputation
- **PingOne / PingFederate:** Adaptive authentication policies consuming risk signals
- **Identity for AI (Announced Nov 2025, GA Planned Early 2026):**
  - Centralised control plane for AI agent identity lifecycle management
  - MCP Gateway for securing AI agent interactions
  - Secretless agentic identity: agents authenticate without stored credentials
  - Positions Ping at the forefront of the emerging AI agent identity challenge
- `<<insert: Ping Identity logo>>`

identitrain

Platform Spotlight: Microsoft Entra

- **Entra ID Protection:** ML-driven risky sign-in and risky user detection, feeding risk levels to policies
- **Conditional Access:** Consumes risk levels as policy conditions: the policy engine for adaptive authentication
- **Verified ID:** Verifiable credentials framework, integrating with AI-powered IDV partners
- **Security Copilot (GA 2025):** Natural language interface for IAM/security queries: "Show me all users with high risk scores this week" (more in §3.2)
- **Entra Agent ID (Public Preview May 2025, Expanded Nov 2025):**
  - First-class identity for AI agents: Agent Registry, Conditional Access for agents, Identity Governance for agents
  - Part of Microsoft Agent 365 platform
  - Uses same controls as human identities: Conditional Access, Identity Protection, network controls
- `<<insert: Microsoft Entra logo>>`

identitrain

Platform Spotlight: Okta & AWS

Okta:

- **Adaptive MFA + Behavior Detection:** Risk-based authentication policies
- **Identity Threat Protection:** Real-time threat detection across the platform
- **Auth0 for AI Agents (GA Oct 2025):** Auth, authorisation, token vault, and audit for agents
- **Cross App Access (XAA):** New protocol extending OAuth for agent-to-app interactions

AWS:

- **IAM Access Analyzer:** Automated reasoning for policy validation and least-privilege generation
- **GuardDuty:** ML anomaly detection for IAM principals
- **Cognito:** Adaptive authentication for customer-facing apps
- **Bedrock AgentCore (GA Oct 2025):** Enterprise agent platform with AgentCore Identity
- `<<insert: Okta and AWS logos>>`

identitrain

Platform Spotlight: SailPoint & Industry Trends

- **SailPoint:**
  - **Harbor Pilot (GA March 2025):** AI agents for identity security: natural language workflows, access review recommendations, identity intelligence. 50% of customers onboarded within 30 days of launch.
  - **Predictive Identity:** AI recommendations for access certifications, role mining, peer group analysis, outlier detection
  - **Agentic Identity Security (Announced 2025):** Managing AI agents as first-class identities within IGA
- **Also Shipping AI:** CrowdStrike Falcon Identity Protection, Saviynt Intelligent Analytics, ForgeRock/Ping Autonomous Identity, Delinea, BeyondTrust, CyberArk
- **The Industry Trend:** AI is no longer a differentiator — it's table stakes. Every major IAM vendor ships AI capabilities. The differentiator is *how well* the AI works and *how transparently* it operates.
- `<<insert: SailPoint logo + graphic showing "AI = Standard" across IAM landscape>>`

identitrain

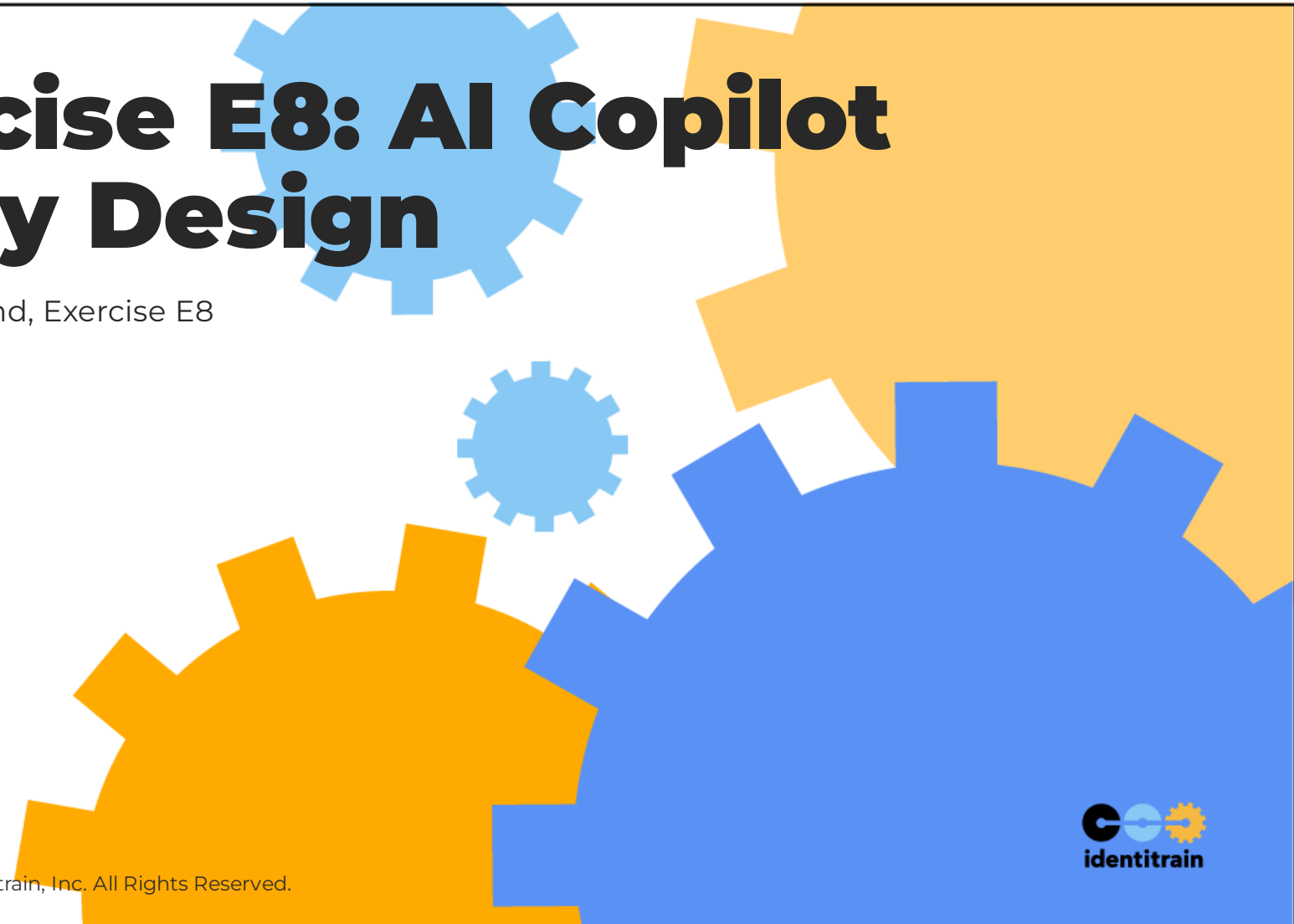
The Cost of AI in IAM: What Managers Need to Know

- **Beyond the Subscription Fee:** AI adoption is a full-lifecycle investment, not just a licence cost.
- **Licensing Models:** Per-user, per-consumption/SCU, per-transaction, freemium vs. enterprise — models vary widely across vendors.
- **Integration & Data Readiness:** Deployment, API development, SSO/SCIM setup. Plus **data cleanup** — the overpermissioning problem means you can't just "turn on" copilots.
- **Ongoing Costs:** Model updates, bias auditing, governance overhead, staff upskilling.
- **Hidden Costs:** Vendor lock-in and data migration. **Cost of NOT acting:** shadow AI carries an average **$670K cost premium** over standard breaches.
- **Key Message:** TCO = Licensing + Integration + Data Readiness + Operations + Governance.
- `<<insert: stacked horizontal bar chart showing Total Cost of Ownership components: Licensing (blue), Integration (teal), Data Readiness (amber), Operations (grey), Governance (purple); with a separate callout box showing "Cost of Inaction: Shadow AI breach premium $670K"; professional chart style>>`

identitrain

Exercise E8: AI Copilot Query Design

See: Chapter_3.md, Exercise E8



identitrain

Exercise E9: Simulating Risk-Based Policy Decisions with AI

See: Chapter_3.md, Exercise E9

The Rise of AI Copilots in IAM

- **What Are AI Copilots?** Natural language interfaces layered on top of IAM and security platforms
- **How They Work:** RAG + LLM + Platform APIs: the copilot retrieves your data, then uses an LLM to generate answers grounded in your environment
- **Leading Examples:**
  - **Microsoft Security Copilot (GA 2025):** Natural language queries across Entra, Defender, Sentinel, Intune
  - **SailPoint Harbor Pilot (GA March 2025):** AI agents for identity security: natural language workflows, access intelligence
- **Why This Matters:** Transforms how IAM teams interact with identity data: from clicking through dashboards to asking questions in plain language

identitrain

What AI Copilots Can Do for IAM Teams

- **Query Identity Data:** "Which users have access to the finance application but aren't in the finance department?"
- **Investigate Incidents:** "Summarise the risky sign-in activity for user john.smith over the past 7 days"
- **Generate Workflows:** "Create an access review campaign for all users with privileged roles"
- **Explain Policy Decisions:** "Why was this user's access request denied?"
- **Draft Documentation:** "Generate a summary of our current access certification status for the audit committee"
- **Key Principle:** The copilot operates within your existing permissions: it can only access and act on data you're authorised to see

identitrain

The Overpermissioning Problem

- **AI Copilots Surface a Pre-Existing Crisis:** Copilots access everything the user can — including overshared content that was practically hidden before.
- **The Scale: 95% of permissions go unused.** Overshared sites, broad file access, stale group memberships — all become instantly queryable via natural language.
- **Real-World Impact:** US House of Representatives **banned Microsoft Copilot** (March 2024) over data leakage risk. Similar concerns in UK and EU governments.
- **Key Message:** Fix permissions BEFORE deploying copilots. The copilot doesn't create oversharing — it amplifies it. Least privilege and access reviews become *prerequisites*.
- `<<insert: iceberg graphic: the small visible tip above the waterline is labelled "AI Copilot Deployment"; the massive underwater section is labelled with layered text: "Years of Permission Sprawl", "Overshared Content", "Stale Group Memberships", "Dormant Accounts"; the underwater section should be visually dominant to convey the hidden scale of the problem>>`

identitrain

Using General AI Tools Safely in IAM Workflows

- **Beyond Integrated Platforms:** General-purpose LLMs (ChatGPT, Claude, Gemini) as IAM assistants for tasks outside built-in features
- **Recall the Golden Rule (Chapter 2, Slide 2.4):** NEVER input sensitive or confidential data into public AI tools
- **Safe Use Cases (with Anonymisation):**
  - Summarising *sanitised* log snippets (replace ALL identifiers with placeholders)
  - Drafting initial policy sections, compliance narratives, training materials
  - Explaining complex technical clauses or regulatory requirements
  - Brainstorming architectural approaches at a high, abstract level
- **What NOT to Do:** Upload raw logs, paste internal system names, share credentials, input unredacted incident details
- **Human Review is Always Required:** AI drafts are starting points, not final products

identitrain

General AI: Safe Workflows & Limitations

- **The Safe Workflow:**
  1. Use internal tools (SIEM, IGA, UEBA) to isolate relevant data
  2. Extract only the specific snippets needed
  3. **Anonymise rigorously:** replace ALL identifiers: users → USER_A, IPs → IP_1, resources → RESOURCE_X, systems → SYSTEM_1
  4. Prompt the LLM with anonymised content only
  5. **Verify** the output against your actual environment before acting on it
- **Key Limitations:**
  - AI has no understanding of your specific environment, threat model, or business context
  - Output is based on general training patterns; may be incorrect, outdated, or irrelevant to your situation
  - Can never replace expert human analysis; use as a supplement, not a substitute
  - Utility depends entirely on effective anonymisation
- **When in Doubt:** If you cannot fully anonymise the data, do not use a public AI tool. Use your organisation's approved enterprise AI platform instead.

identitrain

The NHI Explosion: Non-Human Identities

- **The Scale:** NHIs outnumber human identities **50:1** in typical enterprises, reaching **92:1+** in large organisations
- **What Counts as NHI:** Service accounts, API keys, OAuth tokens, certificates, bot accounts, RPA identities, CI/CD credentials, cloud workload identities
- **The Governance Gap:**
  - **97% of NHIs are overprivileged**: far more access than needed
  - Only **29% rotate credentials** within recommended timeframes
- **Real Breaches, NHI as Entry Point:**
  - **Microsoft Midnight Blizzard (2023):** Legacy OAuth app, no MFA → leadership email
  - **Snowflake Campaign (2024):** Stolen service credentials → 165+ orgs compromised
  - **Internet Archive (2024):** Exposed GitLab token → full source code + user data
- `<<insert: visual ratio comparison: a single human-figure icon on the left, connected by a "50 : 1" label to fifty small machine/gear icons on the right; a red warning badge on the machine side reads "97% overprivileged"; clean infographic style that makes the scale difference immediately visceral>>`

identitrain

AI Agents: A New Identity Type

- **How Agents Differ from Traditional NHIs:**
  - **Autonomous:** Reason, plan, and act — not just execute scripts
  - **Dynamic Scope:** Access needs change per task, not fixed at creation
  - **Delegation Chains:** Agent → User → Organisation: multi-layer accountability
  - **Cross-System:** A single agent may touch dozens of APIs and services
- **The Adoption Reality:**
  - **82–91%** of enterprises deploying or planning AI agents (SailPoint / Okta 2025)
  - Only **44% have governance frameworks**; only **10% have an NHI strategy**
- **The Gap:** Adoption is racing ahead of governance. This is IAM's problem to solve.
- `<<insert: two-column comparison: Left column "Traditional NHI" shows a static service account icon with fixed permission list; Right column "AI Agent" shows a dynamic figure with branching arrows to multiple systems, a "reasoning" brain icon, and a delegation chain; visual emphasis on the contrast between static and dynamic>>`

identitrain

Managing AI Agent Identity

- **Emerging Lifecycle Framework:**
  - **Registration:** Create identity, assign owner, record purpose
  - **Authentication:** OAuth 2.1, federated identity, short-lived tokens. Beyond static credentials
  - **Authorisation:** Fine-grained, least-privilege, with human approval thresholds
  - **Monitoring:** Continuous logging, anomaly detection on agent behaviour
  - **Governance:** Regular reviews, credential rotation at machine speed, decommissioning
- **Emerging Standards:** Model Context Protocol (MCP), OAuth 2.1, OpenID Federation
- **Vendor Solutions Now:** Ping Identity for AI · Entra Agent ID · Auth0 for AI Agents · AWS Bedrock AgentCore · SailPoint Agentic Identity Security

identitrain

IAM Implications of the Agent Economy

- **What IAM Professionals Need to Prepare For:**
  - **Agent Lifecycle Management:** Joiner/mover/leaver processes adapted for agents that can be created and destroyed in seconds
  - **Credential Management at Machine Speed:** Rotation cycles measured in hours, not months. Secretless authentication where possible.
  - **Policy Frameworks for Autonomous Actions:** What can agents do without human approval? What requires a human in the loop? Where are the hard limits?
  - **Accountability and Audit:** Every agent action must be traceable to the agent, its owner, and the user it acts on behalf of
  - **Cross-Platform Governance:** Agents operate across systems; governance can't be siloed per platform
- **The Fundamental Question:** When an AI agent takes an action that causes harm — who is accountable?
- **Start Now:** Even if your organisation isn't deploying AI agents yet, your users probably are. Discovery is the first step.

identitrain

Exercise E10: The Oversharing Audit with AI Analysis

See: Chapter_3.md, Exercise E10

Exercise E11: Design an AI Agent Identity Policy with AI

See: Chapter_3.md, Exercise E11

Shadow AI: The IAM Problem You Already Have

- **The Reality:** Your employees are already using AI tools you haven't approved
- **The Scale:**
  - **80%+** of employees using AI tools at work, most without IT approval
  - **34.8%** entering sensitive data into AI tools (CyberHaven 2025)
  - **Source code with embedded secrets** is the #1 data leak type at **18.7%** of all AI data exposure
- **Real Incidents:**
  - **Samsung (2023, South Korea):** Engineers pasted proprietary source code and confidential meeting notes into ChatGPT, resulting in a company-wide AI ban
  - **CISA (2024, United States):** Employees found using unauthorised AI tools despite federal security mandate
- **International Scope:** This is not a US or tech-sector problem. Shadow AI is global across all industries.

identitrain

Governing Shadow AI Through IAM

- **IAM's Central Role in Shadow AI Governance:**
  - **Access Controls:** Manage which AI tools employees can access: approve, block, or restrict by role
  - **SSO Integration:** Bring approved AI platforms into your SSO/federation infrastructure for visibility and control
  - **Acceptable Use Policies:** Define what data can and cannot be input to AI tools (reference the Golden Rule from Ch2)
  - **Monitoring & Discovery:** Detect unapproved AI tool usage through network monitoring, CASB, and endpoint telemetry
  - **DLP Integration:** Deploy data loss prevention controls that can detect sensitive data being sent to AI endpoints
- **The Cost of Inaction:** Shadow AI incidents carry an average **$670K cost premium** over standard data breaches
- **Pragmatic Approach:** Don't just ban AI — provide approved alternatives. If you don't give people safe tools, they'll use unsafe ones.

identitrain

Exercise E12: Shadow AI Discovery Challenge with AI

See: Chapter_3.md, Exercise E12

Responsible AI & the Road Ahead

An introduction to artificial intelligence concepts and applications

- Using AI Power Wisely and Safely
- From Real-World Threats to Governance Frameworks
- `<<insert: cover graphic: a shield icon with subtle AI circuit-board patterns overlaid, balanced scales of justice on one side and a forward-pointing arrow on the other; colour palette conveys authority and responsibility; professional, not decorative>>`

identitrain

AI as a Dual-Use Technology

- **The Fundamental Tension:** AI empowers defenders AND attackers.
- **Defenders use AI to:**
  - Detect anomalies and compromised accounts (UEBA, risk scoring)
  - Automate governance and policy analysis
  - Accelerate incident response and investigation
- **Attackers use AI to:**
  - Generate hyper-realistic phishing and deepfakes
  - Create synthetic identities that bypass verification
  - Inject malicious prompts into AI-powered enterprise tools
  - Scale social engineering and credential attacks
- **Key Insight:** Every AI capability we've discussed in Chapter 3 has an offensive mirror image.

identitrain

AI-Powered Social Engineering & Deepfakes

- **The Threat Landscape Has Shifted:** AI makes social engineering more convincing, personalised, and harder to detect.
- **Real Incidents:**
  - **Arup (2024):** Deepfake video call impersonating CFO → **$25.6M** loss
  - **Ferrari (2024):** AI-cloned CEO voice via WhatsApp → detected by personal verification question
  - **FAMOUS CHOLLIMA (2024–2025):** AI-generated personas + deepfake interviews → **320+ companies** infiltrated
  - **GoldPickaxe (2024):** Malware stealing facial biometric data to bypass bank IDV
- **Scale:** Deepfake fraud exceeded **$200M in Q1 2025**. North America 38%, Asia 27%, Europe 21%.
- `<<insert: world map with four incident pins: Hong Kong (Arup $25.6M), Italy (Ferrari voice clone), Southeast Asia (GoldPickaxe), Global spread (FAMOUS CHOLLIMA 320+ companies); each pin has a brief label; a sidebar shows "$200M+ deepfake fraud in Q1 2025"; clean data-visualisation style>>`

identitrain

Prompt Injection & AI System Attacks

- **A New Attack Category for IAM:** Adversaries targeting the AI tools your organisation uses.
- **Prompt Injection:** Hidden instructions in content an AI processes, causing data leaks or unintended actions.
- **Real Examples:**
  - **EchoLeak / CVE-2025-32711 (June 2025):** Zero-click: hidden email prompts caused M365 Copilot to exfiltrate data. **CVSS 9.3.** No user interaction required.
  - **LOLCopilot (Black Hat 2024):** Red-team tool weaponised Copilot to alter banking details and exfiltrate data via prompt injection.
  - **ASCII Smuggling (2024):** Invisible Unicode characters hiding injection payloads — undetectable to humans, executable by AI.
- **IAM Implication:** Copilots operate with the user's permissions. Prompt injection = an attacker with your access level.
- `<<insert: attack flow diagram: Step 1: attacker sends email containing hidden prompt injection payload; Step 2: AI copilot's RAG retrieval ingests the email; Step 3: copilot follows hidden instructions and extracts sensitive data from user's SharePoint/OneDrive/Teams context; Step 4: data exfiltrated to attacker-controlled URL; each step numbered with clear arrows; red highlighting on malicious elements>>`

identitrain

Synthetic Identity Fraud

- **What Is It?** Combining real identity fragments with fabricated data to create a "person" who doesn't exist.
- **Why It Works:** Traditional verification checks individual data points; each may look valid even though the whole identity is fabricated.
- **AI Acceleration:**
  - **ProKYC (2024):** Criminal tool generating synthetic passports + deepfake video for $629/year — bypasses KYC liveness checks
  - Fake IDs and forged documents: **50% of all identity fraud attempts** in 2024
- **Scale:** New account fraud exceeded **$5.3 billion** (2023), rising annually — fastest-growing financial crime category.
- **IAM Impact:** Challenges IDV onboarding, CIAM registration, and any process that assumes documents are trustworthy.

identitrain

Attacks Against AI Systems

- **Targeting the AI Itself:** Exploiting how AI models learn and make decisions.
- **Data Poisoning:** Corrupting training data to introduce backdoors or misclassification — especially dangerous for continuously learning UEBA systems.
- **Adversarial Examples:** Inputs that look normal to humans but fool AI — bypassing facial recognition or evading anomaly detection.
- **Model Extraction:** Querying a model to reconstruct it (IP theft) or infer sensitive training data (privacy breach).
- **IAM Implications:** Poisoned UEBA ignoring attacks as "normal"; adversarial inputs bypassing biometrics; compromised risk engines giving false confidence.
- **Key Message:** The AI systems we rely on for security can themselves become attack targets.

identitrain

Defence Implications

- **AI Is Not a Silver Bullet:** Powerful, but has limitations and vulnerabilities, as we've just seen.
- **Defence-in-Depth Still Applies:** Multiple layers; don't over-rely on any single AI control. If the risk engine fails, other controls must catch the threat.
- **Monitor the AI Itself:** Track model performance, investigate anomalies in AI behaviour, maintain non-AI fallback mechanisms for critical decisions.
- **Plan for AI Failures:** Include AI compromise and incorrect results in your incident response playbooks.
- **Update Human Processes:** Train users on deepfake verification and prompt injection. Update helpdesk and approval procedures for a world where voice and video can be faked.
- `<<insert: concentric-ring defence-in-depth diagram: outermost ring "Network Controls", next ring "Endpoint Security", next "Policy & Access Controls", next "AI-Powered Detection", innermost ring "Human Oversight"; emphasis that AI is one layer among many; no single point of failure; label on the side: "If one layer fails, others catch the threat">>`

identitrain

Exercise E13: Deepfake Detection Challenge

See: Chapter_4.md

Algorithmic Bias: Sources & Mechanisms

- **What is Algorithmic Bias?** Systematic errors in AI leading to unfair outcomes for certain groups — often unintentional, always consequential.
- **Source 1: Biased Training Data:**
  - Data reflects historical biases or underrepresents groups — the data IS the worldview the AI learns
  - Example: Facial recognition trained on lighter-skinned faces performs poorly on darker-skinned faces
- **Source 2: Flawed Algorithm Design:**
  - Developer choices about features or proxies introduce bias
  - Example: Using postal code as a risk factor when it correlates with ethnicity
- **Source 3: Feedback Loops:**
  - Biased outputs skew future data collection, reinforcing and amplifying the original bias
- `<<insert: circular feedback loop: four nodes connected by arrows forming a cycle: "Biased Training Data" → "Biased Model" → "Biased Decisions" → "Biased Data Collection" → back to start; each arrow labelled with what happens at that step; red colour intensifying around the cycle to convey reinforcement>>`

identitrain

Bias Impacts in IAM: International Evidence

- **Facial Recognition Bias (Documented):**
  - NIST studies: **10–100x higher false positive rates** for certain demographic groups
  - **8+ wrongful arrests** linked to facial recognition errors (US, UK)
  - Clearview AI fined in **UK, France, Italy, Australia, and Greece**; FTC enforced against Rite Aid (US, 2023)
- **Regulatory Trend:** GDPR enforcement + EU AI Act classifying remote biometric identification as high-risk. Growing global pattern of fines targeting biometric bias.
- **IAM-Specific Risks:**
  - **IDV:** Higher false rejection rates for underrepresented groups → denial of service
  - **Risk Scoring:** Location, device type, or network used as unfair proxies for elevated risk
  - **Behavioural Biometrics / Voice Recognition:** Performance varies by demographics, accent, age, and physical ability

identitrain

Mitigating Bias in IAM

- **1. Diverse Training Data:** Audit datasets for demographic representation; source data reflecting your actual user population.
- **2. Bias Testing:** Test across demographic groups **before AND after** deployment. Disaggregate metrics by group. Regular audits, not one-time checks.
- **3. Fairness Metrics:** Track quantitative measures (equal opportunity, demographic parity, calibration). Set thresholds and alert on drift.
- **4. Explainable AI:** If reasoning is visible, bias is easier to detect and fix.
- **5. Human Oversight & Appeals:** Clear, accessible processes to contest AI decisions. Human reviewers who can override. Appeals must be genuinely functional.

identitrain

Explainable AI (XAI) in IAM

- **The Problem:** Many AI systems are "black boxes" — data in, decision out, reasoning opaque.
- **Why Explainability Matters for IAM:**
  - **Trust & Debugging:** Users need to know *why* access was denied; teams need to know what went wrong
  - **Bias Detection:** Visible reasoning exposes reliance on problematic factors
  - **Accountability & Compliance:** EU AI Act and GDPR require transparency for high-risk AI decisions. "The AI decided" is not auditable.
- **How XAI Appears in IAM Tools:**
  - Risk score breakdown: "Unusual Location (0.4), New Device (0.3), Off-Hours (0.3)"
  - Access review: "Revoke — unused 180 days, not held by 95% of peers"
  - Policy trace: "Denied — Condition X failed at rule Y"
- **Goal:** Move from opaque AI pronouncements to transparent, auditable reasoning.
- `<<insert: two-panel comparison: Left panel "Black Box AI": opaque box with a question mark, input goes in, unexplained score comes out; Right panel "Explainable AI": transparent box showing contributing factors (Unusual Location 40%, New Device 30%, Off-Hours 30%) with weights flowing to a labelled decision; the contrast should be visually stark>>`

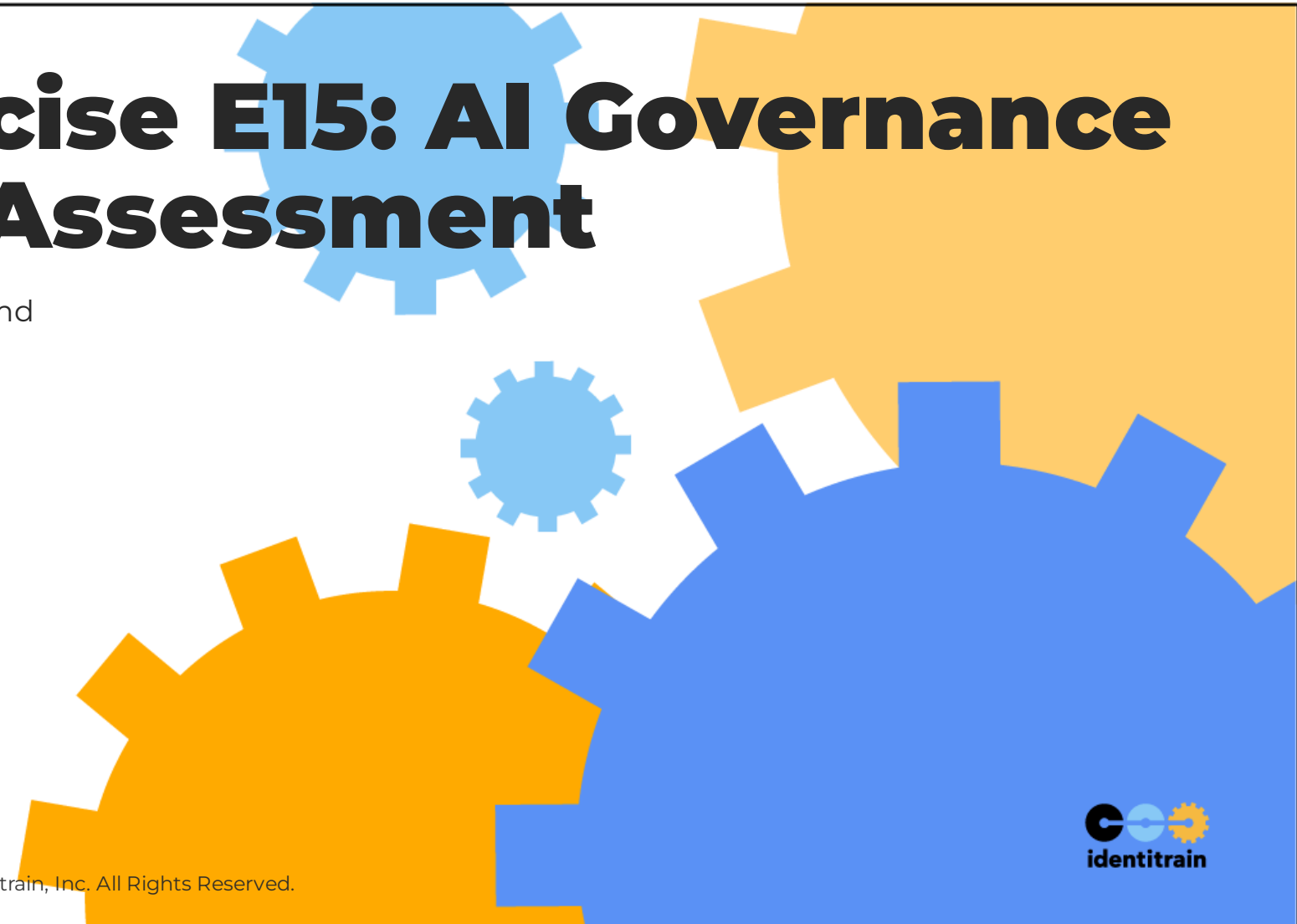
identitrain

Exercise E14: Responsible AI Scenarios with AI Analysis

See: Chapter_4.md

Exercise E15: AI Governance Gap Assessment

See: Chapter_4.md



identitrain

Data Privacy Risks in AI for IAM

- **AI Thrives on Data:** ML models require sensitive IAM data: user profiles, behaviour logs, biometric templates, policy configs.
- **Key Risks:**
  - **Unauthorised Access:** Breaches of AI systems or data pipelines
  - **Data Misuse:** Repurposing beyond governance/consent scope
  - **Inference Attacks:** Probing models to extract sensitive training data
  - **Public Tool Risks:** The golden rule still applies: **never input sensitive data into public AI tools**
  - **Over-Collection / Lack of Transparency:** Gathering more than necessary; users unaware their data feeds AI
- **Key Message:** Securing the AI = securing the data. The AI *is* the data.

identitrain

Data Privacy: Best Practices for AI in IAM

- **Data Minimisation:** Collect only what the AI strictly needs, not "everything available."
- **Anonymisation:** Remove or replace direct identifiers wherever feasible, especially for model training.
- **Secure Infrastructure:** Protect AI models, pipelines, and inference systems with the same controls as critical data stores.
- **Vendor Due Diligence:** Vet data handling, residency, and model-training data usage *before* procurement.
- **Access Control for AI Systems:** Treat models, APIs, and management interfaces as sensitive systems.
- **Transparency & Consent:** Inform users specifically how their data feeds AI. "We use AI" is not sufficient.
- **Retention Policies:** Define retention for AI training data and inference logs. Ensure deletion actually works.

identitrain

The EU AI Act: What IAM Teams Must Know

- **Status: Active law with obligations already in force.**
- **Key Timeline:**
  - **Feb 2025:** Prohibited practices enforceable (social scoring, certain biometric surveillance)
  - **Aug 2025:** General-purpose AI model obligations active
  - **Aug 2026: High-risk AI system obligations**: data quality, transparency, human oversight, conformity assessments
- **What's "High-Risk" for IAM?** (Annex III)
  - Remote biometric identification · AI in employment decisions · AI in access to essential services
- **Fines:** Up to **€35M or 7% of global turnover**
- **Extraterritorial:** Applies to any organisation serving the EU market
- `<<insert: horizontal timeline spanning Feb 2025 to Aug 2026: three milestone markers: "Feb 2025: Prohibited Practices" (in force), "Aug 2025: GPAI Obligations" (in force), "Aug 2026: High-Risk Obligations" (upcoming, prominently highlighted in red/amber); a "YOU ARE HERE" indicator at current date; the Aug 2026 deadline should visually dominate>>`

identitrain

Global Regulatory Landscape

- **Data Protection Laws (Apply to AI Data):**
  - **GDPR** (EU), including automated decision-making rights (Art. 22)
  - **UK GDPR** · **Brazil LGPD** · **China PIPL** · **India DPDPA** · **Singapore PDPA**
- **AI-Specific Frameworks:**
  - **NIST AI RMF** (US): voluntary risk management + AI 600-1 generative AI profile
  - **ISO/IEC 42001**: certifiable AI management systems standard
  - **Colorado AI Act** (US): state law for high-risk AI, effective **June 30, 2026**
- **Key Message:** Multi-jurisdictional compliance is the reality. Multiple overlapping frameworks apply simultaneously.

identitrain

AI Governance for IAM Teams: A Practical Framework

- **Why?** Ad hoc AI adoption creates invisible risk. Governance makes it visible.
- **Eight Core Components:**
  1. **AI Inventory & Classification** — know what AI you have, classify by risk
  2. **Risk Assessment** — bias, privacy, security, operational
  3. **Data Governance** — minimisation, quality, retention, consent
  4. **Vendor Assessment** — training data, bias testing, data handling
  5. **Human Oversight Design** — review points, override authority, appeals
  6. **Bias Auditing** — regular disaggregated testing; external audits for high-risk
  7. **AI Incident Response** — playbooks for bias, poisoning, prompt injection
  8. **Continuous Monitoring** — performance, fairness metrics, drift detection
- **Aligns to:** NIST AI RMF (Govern, Map, Measure, Manage) and ISO 42001
- `<<insert: circular governance framework diagram: eight segments arranged like a wheel: AI Inventory, Risk Assessment, Data Governance, Vendor Assessment, Human Oversight, Bias Auditing, Incident Response, Continuous Monitoring; NIST AI RMF and ISO 42001 logos sit as foundation blocks beneath the wheel>>`

identitrain

Exercise E16: Regulation Mapping with AI Research

See: Chapter_4.md

identitrain

Future: Deeper Automation in Identity Lifecycle

- **Beyond Current Automation:** Moving from AI-assisted to AI-orchestrated operations.
- **Intelligent Identity Lifecycle:**
  - **Joiners:** AI predicting appropriate access from role, team, projects, and peer comparison — not static templates
  - **Movers:** Proactive recommendations when organisational changes are detected, before access is requested
  - **Leavers:** Deprovisioning triggered by correlated signals (HR changes, inactivity), not just termination date
- **Advanced Role Engineering:** AI designing and maintaining least-privilege roles over time, with continuous cleanup based on actual usage.
- **Self-Tuning Systems:** AI remediating risky configurations or policy drift — requires robust governance with human approval gates.

identitrain

Future: Continuous Authentication & Zero Trust++

- **Zero Trust Principle:** Never trust, always verify. Context-based access decisions.
- **AI Enhancement:** Evolves Zero Trust from point-in-time checks to continuous, dynamic verification.
- **Continuous Authentication:**
  - AI monitors behaviour, device posture, and session context **throughout** a session, not just at login
  - Mid-session risk triggers real-time actions: step-up auth, access restriction, or session termination
- **Hyper-Personalised Access:**
  - AI adjusting **fine-grained data visibility** based on real-time risk — full access when low-risk, redacted views when elevated
- **Dependencies:** Requires sophisticated UEBA, endpoint integration, AI risk modelling, and strong privacy governance.
- `<<insert: horizontal session timeline: left edge "Login (Initial Auth)", then monitoring checkpoints along the session bar where risk signals appear (device change, network switch, unusual data access); at each signal point an adaptive response branches off (step-up auth, restrict scope, terminate session); colour gradient from green (low risk) to red (high risk) along the bar>>`

identitrain

The Evolving Role of the IAM Professional

- **AI Transforms the Role. It Doesn't Replace It.**
- **Emerging Skill Requirements:**
  - **AI Literacy & Prompt Engineering:** Understanding AI + communicating with it effectively
  - **AI Governance:** Managing risk, bias, compliance, and oversight for AI systems
  - **Vendor Evaluation:** Distinguishing real capability from marketing
  - **NHI Management & Data Fluency:** Governing agent identities; understanding AI data and privacy implications
- **Career Opportunity:**
  - AI/ML: **top-5 in-demand skill** for cybersecurity professionals (ISC2 2025)
  - **56% wage premium** for AI skills | **70%** pursuing AI qualifications
- **The Human-AI Collaboration Model:** AI handles scale, speed, pattern matching. Humans provide strategy, judgement, ethics, accountability.

identitrain

Human Oversight: The Non-Negotiable

- **AI Augments, Doesn't Replace Judgement:** AI lacks genuine understanding, common sense, and ethical reasoning.
- **Humans Remain Essential For:**
  - **Strategy & Policy:** Defining risk tolerance, ethical boundaries, security goals
  - **Edge Cases & Ambiguity:** Novel threats and situations that don't fit patterns
  - **Critical Decisions:** High-impact events requiring accountability — blocking users, reporting breaches, revoking privileged access
  - **Ethics & Accountability:** "The AI decided" is not an acceptable answer to regulators, auditors, or affected individuals
- **The Design Principle:** Every AI system in IAM must answer: "When does a human review this? Who can override? How does someone appeal?"
- **Not anti-AI — pro-resilience.** The strongest IAM programmes combine AI capability with robust human oversight.
- `<<insert: graphic showing two figures side by side: an AI icon (circuit brain) labelled "Scale · Speed · Pattern Matching" and a human figure labelled "Strategy · Judgement · Ethics · Accountability"; bidirectional arrows between them; a shared output labelled "Resilient IAM"; conveys collaboration not replacement>>`

identitrain

Exercise E17: Capstone: AI-in-IAM Action Plan

See: Chapter_4.md

Course Summary: What We Covered

An introduction to artificial intelligence concepts and applications

- Three Days, Four Chapters, One Mission: Making AI Practical for IAM
- `<<insert: four connected nodes in a horizontal learning-journey arc, each node labelled with a chapter theme (AI Foundations → Working with AI → AI in IAM Today → Responsible AI & Road Ahead); each node has a small representative icon; connecting arrows show progression; clean, professional infographic style>>`

Key Takeaways

- **AI is integral to modern IAM** — not a future possibility, but today's reality across every major platform.
- **Understanding core concepts is essential** — not to become AI engineers, but to evaluate tools, assess vendor claims, and make informed decisions.
- **Prompt engineering is a practical, immediately applicable skill**, and the golden rule of data privacy is non-negotiable when using AI tools.
- **Real threats demand real defences:** deepfakes, prompt injection, and synthetic identities are documented, quantified, and growing.
- **Responsible AI principles are non-negotiable:** bias testing, privacy governance, regulatory compliance, and explainability aren't optional extras.
- **AI augments human capabilities; it doesn't replace judgement:** strategy, ethics, accountability, and edge-case handling remain human responsibilities.
- **The IAM profession is evolving, not disappearing:** AI literacy, governance skills, and NHI management are career-defining competencies with a 56% wage premium.
- **Continuous learning is essential;** the AI landscape changes faster than any course can capture. Stay curious, stay critical, stay informed.

identitrain

Next Steps & Resources

- **Continue Learning:**
  - **Experiment (Safely!):** Practice prompt techniques on non-sensitive tasks. Explore AI features in your IAM platforms.
  - **Use Your Action Plan:** Execute the commitments from Exercise E17: this week, this quarter, this year.
  - **Stay Current:** The landscape changes monthly.
- **Key Resources:**
  - **Frameworks:** NIST AI RMF | ISO/IEC 42001 | EU AI Act full text
  - **Vendor Docs:** Entra, Ping, Okta, SailPoint, AWS AI feature documentation
  - **Community:** RSA Conference | Identiverse | Gartner IAM Summit | ISC2 | ISACA
  - **Research:** NIST FRVT | MITRE ATLAS | OWASP Top 10 for LLM Applications
- **Apply Responsibly:**
  - Prioritise data privacy. Be critical of AI outputs. Advocate for governance and human oversight.

identitrain

Thank You & Q&A

- **Thank You!**
- The future of AI in IAM is being built right now — and you're part of it.
- Questions?
- `<<insert: clean centred "Thank You" typographic graphic with the course title "AI for IAM Professionals" below it; subtle background motif blending a neural-network pattern with a keyhole/shield identity icon; same colour palette as the cover slide>>`

identitrain