Securing the Backbone of Generative AI with Idira Blueprint

Table of Contents

Executive Summary

The integration of generative artificial intelligence (GenAI) into the fabric of modern IT infrastructure presents both unparalleled opportunities and significant security challenges. GenAI uses generative models to produce text, images, videos, code, or other forms of data. These models learn the underlying patterns and structures from training data and generate new content based on input, often provided as a natural language prompt. As industries increasingly rely on GenAI for a wide range of applications—from content creation to predictive analytics—the security of these systems becomes paramount.

Idira™ Identity Security Blueprint (Idira Blueprint), by Palo Alto Networks, provides a comprehensive framework to secure GenAI application infrastructure using a holistic approach that includes both human and machine identities. It advocates for a practical, risk-based strategy to enhance GenAI security posture.

By adopting Idira Blueprint, organizations can ensure a comprehensive, risk-based approach to GenAI security that pursues zero standing privileges (ZSPs) to safeguard their AI workloads and services. Follow the guidance in this whitepaper to help your organization secure your GenAI environments against increasing complexity and threats.

What You'll Learn

- **GenAI landscape:** The proliferation of services and identities created by AI applications underscores the critical need for a dynamic identity security strategy.
- **Minimize the attack surface:** An identity security program must advocate for ZSPs, just-in-time (JIT) access, and entitlements provisioning within GenAI environments.
- **Risk-based guidance:** This battle-tested, best practices framework is designed to reduce the risk for all types of identities. It also helps focus on the best approach to securing complex GenAI infrastructures.
- **User experience:** To move fast while remaining secure, GenAI teams need processes that will have minimal impact on their daily workflow.
- **Critical identity security controls:** The cornerstone of a successful cybersecurity strategy is the ability to stop lateral and vertical movement and limit privilege escalation and abuse.
- **Prioritization strategies:** A one-size-fits-all path doesn't exist for adopting identity security controls in your GenAI environments. Multiple approaches are available, each of which considers the existing capabilities, risk-based prioritization, compliance requirements, and cloud adoption approaches.

GenAI and Security

The rise of GenAI technologies has revolutionized how organizations operate, how they transform data into insights, and how they deliver innovative services. However, with the rapid integration of GenAI across industries and the increasing complexity and autonomy of these systems, new vulnerabilities have emerged. Securing your GenAI infrastructure is crucial to maintaining the integrity and effectiveness of these transformative technologies.

GenAI Infrastructure and Idira Blueprint

GenAI is rapidly transitioning from a content creation engine to an autonomous agent that can make decisions and perform actions on an organization's behalf. Even though AI agents are not widely used in major production environments, analysts predict rapid adoption in the near future.

As GenAI technologies evolve and become more intricate—and business drivers force rapid innovation—the responsibility to safeguard these systems increasingly falls on the organizations that deploy them. GenAI systems are highly fluid. They involve complex interactions between datasets, machine learning models, and automated decision-making processes, creating numerous security challenges at every layer of the architecture.

In GenAI environments, the traditional concept of a "security perimeter" dissolves, shifting the focus to the security of the human and machine identities that interact with AI systems. As these technologies gain capabilities and access within organizations, managing the identities and permissions associated with GenAI becomes a formidable challenge. Every day, new AI models, data scientists, automated processes, and interconnected systems expand the scope of access, requiring robust and dynamic identity security measures.

The diversity in authentication (AuthN) and authorization (AuthZ) methods used by various GenAI components makes it essential to centralize and standardize identity security controls. This centralization is vital for maintaining oversight and control over who and what can interact with GenAI systems.

To help address these challenges, we present Idira Blueprint—a comprehensive framework designed to secure identities within any enterprise environment, including GenAI environments. Idira Blueprint builds upon our proven best practices in identity security, framed within this whitepaper to meet the unique demands of GenAI technologies. By developing a common language and unified approach to identity security across different GenAI components, Idira Blueprint aims to help organizations significantly reduce the risks associated with these powerful systems.

Challenges with Identity Across the GenAI Infrastructure

GenAI infrastructure integrates various services and platforms into a complex web of systems, complicating the security landscape. This infrastructure typically includes cloud-based services, on-premises data centers, third parties, and hybrid systems that combine both, leading to a varied and fluid responsibility model and complex identity and access management (IAM) challenges. As a result, organizations are responsible for solving the challenges related to four key areas of identity security:

- **Authentication:** GenAI infrastructure consists of a mix of components that support a wide array of authentication mechanisms for various types of identities. Different identities use different methods for authentication depending on the use case or scenario. There are valid business justifications for using SAML-based username and password or access or API key authentication.
- **Authorization:** Managing permissions across a GenAI landscape involves coordinating access rights in a secure and functional manner. Permissions might be defined and enforced differently in cloud services vs. cloud infrastructure vs. on-premises components, increasing the complexity of authorization management. Organizations must navigate these differences to maintain tight control over who can access what within their GenAI infrastructure.
- **Access:** Methods of access vary from identity to identity and scenario to scenario. Managing and controlling the various access planes for each unique situation is a new problem for many organizations.
- **Audit and monitoring:** GenAI's nondeterministic nature (legitimately performing different actions based on its own judgment) makes effective oversight a challenge. Organizations need enhanced monitoring capabilities and sophisticated logging solutions that can operate across cloud providers, cloud infrastructure, and on-premises systems. Integrating these systems for a unified view of activities and security events is crucial but remains a complex task.

The combination of these challenges makes the GenAI infrastructure a ripe target for malicious actors to take advantage of and certainly ensures they'll find a way to do it. The speed at which organizations are developing GenAI applications to drive increased revenue, operational efficiency, and competitiveness means that security is likely an afterthought. By recognizing and addressing these challenges, organizations can better secure their GenAI infrastructure against the evolving landscape of cyberthreats, ensuring the integrity and reliability of their GenAI applications, all while going to market with speed.

The Emergence of AI Agents

GenAI is quickly evolving from a tool for content creation into an autonomous system that can make decisions and take actions on an organization's behalf. While AI agents are not yet widely implemented in large-scale production settings, experts anticipate rapid adoption soon. This shift will introduce significant security challenges, particularly in managing machine identities—AI agents—that might not always perform predictably.

New Challenges and Security Considerations

The complexity of securing these identities at scale, potentially involving thousands or even millions of agents running concurrently, introduces key security considerations. For example, they include authenticating AI agents to various systems, managing and limiting their access, and ensuring their lifecycle is controlled to prevent rogue agents from retaining unnecessary access.

Figure 1 shows a simplified process that includes multiple agents, executing various roles within the overall system managed by an orchestrator agent. Agentic processes can and will be even more complex, expanding to include various agents performing bigger and smaller specialist roles. By nature, agents tend to act less predictably, making it harder to enforce security controls and ensure they are performing the correct actions. This becomes even more complicated as each AI agent might use a different large language model (LLM) service for optimal performance. The agent could also be provided as a service by an external supplier or vendor.

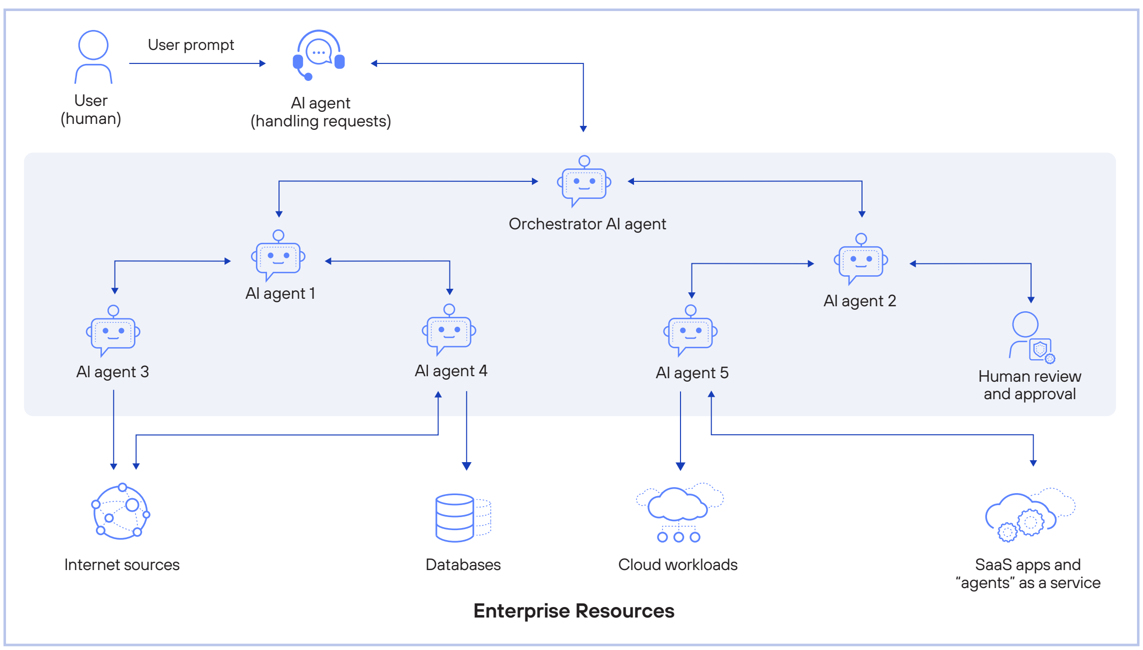

Figure 1. Securing agentic processes requires specific identity security measures

A few identity security requirements become immediately apparent:

- Establish strong authentication between all the entities, including:
  - Between humans and agents.
  - Agents and other agents.
  - Agents and resources.
  - Multifactor authentication (MFA) for humans as they authenticate to run agent processes or provide their input as part of the process.
  - Securing keys, certificates, and secrets used by AI agents and additional services.
- As agents access critical resources across the enterprise, use least-privileged access and JIT access, especially due to their unpredictable behavior and risk of compromise.
- Apply strong audit and traceability across the interactions between different entities and activities that are performed on resources.

In production environments, organizations should develop robust strategies for access management, authentication, and lifecycle management. As this technology continues to evolve, additional insights will shape best practices for enterprise integration. While securing the backend infrastructure is essential for running AI agents on a secure platform, a deeper exploration of this topic lies beyond the scope of this paper.

The Anatomy of GenAI Identities

To better understand the challenges organizations face when securing their GenAI environments, you first need to understand the anatomy of GenAI architecture and the identities that interact with it.

Application Architecture, Infrastructure, and Identities

Generally, a GenAI application architecture includes the following components.

LLM Service

This platform or framework hosts and serves LLMs, including the LLM agent. The LLM is the core machine learning model that processes the input and generates intelligent responses. The LLM agent acts as the intermediary between the application and the LLM. The LLM service as a whole provides AI-driven responses based on user inputs or organizational context. It's often consumed as a native service from the cloud service provider (CSP) like Amazon Web Services (AWS) Bedrock, Azure OpenAI, or Gemini Enterprise Agent Platform.

Application Layer

This main processing layer orchestrates all operations. Its function is to serve as the core logic of the system (e.g., validating requests, calling LLM services, and interacting with external services). It manages integration with other components like the LLM service, Retrieval-Augmented Generation (RAG) database, and server-side services. The application layer often runs as some form of compute, whether in a virtual machine, in a Docker or Kubernetes container, or as a native serverless compute function such as AWS Lambda or Azure Functions.

Server-Side Services

These external services or APIs provide supplementary functionality required by the application that can be invoked by the LLM agent, such as external APIs for third-party data or other backend services for user management, analytics, or other business logic. These service-side services augment application capabilities by handling specific tasks or providing external data. They are often hosted on compute instances like virtual machines, in containers, or as serverless compute functions.

Retrieval-Augmented Generation

This database, also referred to as a RAG vector database, is optimized for storing and retrieving vector representations of data to enhance the LLM's capabilities by integrating organizational knowledge. It supports RAG workflows, where relevant knowledge is retrieved from the database to provide context for the LLM's response. These databases are often consumed from a cloud service provider and include examples like Pinecone, OpenSearch VectorDB, and MongoDB.

Training Datasets

These datasets, also known as fine-tuning datasets, are a storage layer used to improve or fine-tune the LLM. These stored datasets are collected from user interactions or external sources that ML engineers use to fine-tune or train the LLM for better performance on specific tasks. These training datasets are often stored in unstructured data sources like AWS S3 buckets, elastic file system (EFS), or other remote storage services.

External Data Sources

The data is collected or crawled from the internet and used for improving the system. It provides the raw material for training datasets and can include both public data (e.g., web-crawled data) and private organizational data (e.g., internal documentation and customer records). This information is also often stored in unstructured data sources like the training data but is sourced from other places, like OneDrive or the public internet.

Ingress Service

This service is the entry point for incoming requests from users over the internet. It handles request authentication, rate limiting, and routing to the correct service and ensures secure and scalable interaction between users and the system. The ingress service is typically a cloud service provider's native API gateway (e.g., AWS API Gateway and Azure API Management).

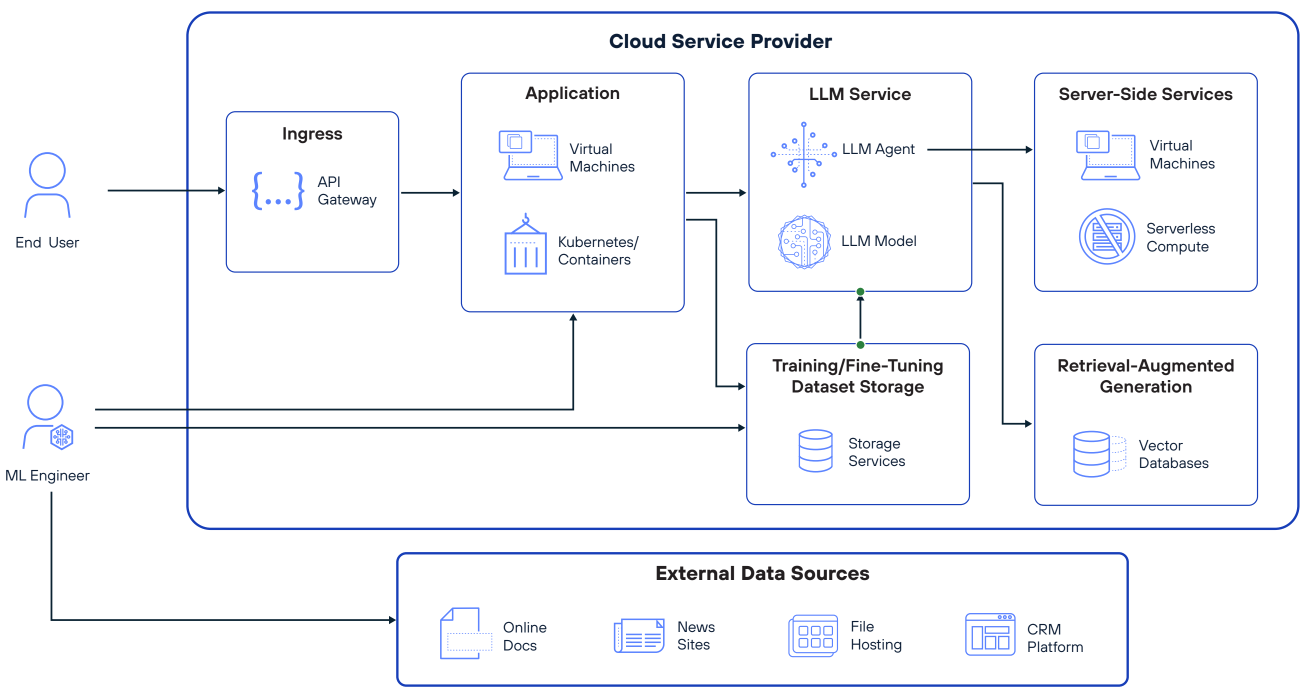

Figure 2. Reference architecture for GenAI applications

Most of these components are commonly deployed or exist natively within cloud service providers. If you're unfamiliar with cloud anatomy, read our whitepaper: *Best Practices for Securing Cloud Identities.*

Identities in the GenAI Ecosystem

Organizations also have identities that interact with various elements of this architecture, including access to the cloud service provider management layer, various cloud-native services, and the hosted infrastructure. These identities are typically:

- **AI/ML engineers** are developers who self-administer various cloud services, develop the various elements of their GenAI applications using cloud-native services, push workloads and GenAI code into the cloud, and access supporting resources. These are the riskiest identities as they often have both privileged access to the infrastructure and services, along with the responsibility to develop the actual code running in the AI system.
- **Cloud operations** have evolved from traditional IT admin roles such as infrastructure operations, networking engineers or database admins into roles like cloud operators, architects, and site reliability engineers. Their role is to support the operational stability of all the cloud services in general.
- **End users** are the workforce identities who interact and interface with the GenAI application over the internet. These are the users for whom the GenAI applications are built.
- **Machine identity workloads** interact with the GenAI infrastructure, including cloud-native services like LLMs, custom GenAI applications, and automation tools that support secure software development and cloud processes. Some machine identities function as AI agents with varying levels of autonomy. Some operate under human supervision, such as flight booking agents. Others, like supply chain agents, act independently without human intervention.

Access Models for the GenAI Infrastructure

All these human and machine identities authenticate to their respective resources and services. They use various methods, including standing federated access via an identity provider (IdP), long-lived freestanding local accounts like user passwords and access or API keys, cloud-native IAM, and private key infrastructure like certificates.

To simplify this complex web of access, we present the access models in three scenarios:

- End-user access to the AI application.
- Developer access for building and maintaining the AI application.
- Machine identities (intra-AI application services) that have access within the whole system.

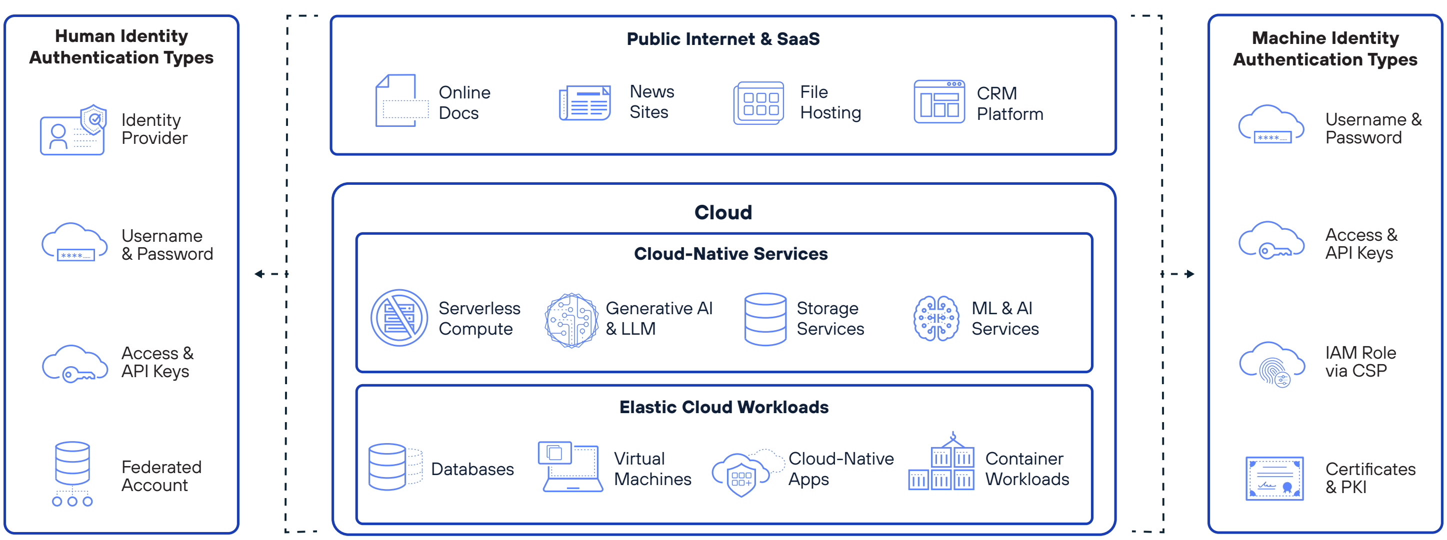

Figure 3. Access models for a GenAI infrastructure

End-User Access to the AI Application by the Ingress Point

This step involves the authentication of users as they attempt to access the AI application via the ingress point. Various mechanisms ensure that only authorized users can interact with the system. These users typically authenticate via SAML with an external identity provider via an API key that's local to the AI service for CLI or automated interaction or via a local username and password.

Since the ingress point is typically a cloud-native API gateway, mutual TLS (mTLS) is often used to mutually authenticate both the ingress service and the application via certificates. IAM role-based access controls (RBAC) are employed in environments like AWS IAM, where the ingress assumes a specific role to communicate securely with the application. API Gateway authentication mechanisms are also crucial to ensure that only authorized backend services handle the requests, maintaining a secure and trusted communication channel.

Access to Build and Maintain an AI Infrastructure and Application

IT team members have different access to specific parts of the AI infrastructure and applications. Except for developer access to coding tools, all the other access types apply to AI/ML engineers, developers, and other IT administrative team members.

Developer and IT Access to Cloud-Native Services

Team members access their cloud service providers and subsequently any component of the AI application that runs on a cloud-native service like the LLM or training dataset. They use a few common methods, including identity providers, access or API keys, and CLI tokens. For web-based access to the cloud provider, SAML-based IdP connections are the default authentication method. Access or API keys are often used for long-lived CLI sessions or automated script-based access. CLI tokens are most often used for CLI-based access on a short-lived basis.

Developer and IT Access to Cloud Workloads

Team members access cloud workloads, such as the server-side virtual machines or application layer. They use the federated access, usernames and passwords, and SSH keys methods.

Developer and IT Access to Data Sources

Team members access the training dataset storage and external data sources by using various authentication mechanisms. Access to training or fine-tuning datasets is typically controlled via an external identity provider with SAML-based authentication or federated access to the data source with a username and password. Access to external data sources is often controlled the same way unless the data is publicly available on the web, in which case no authentication or access control is required.

Developer and IT Access to Pipeline

Team members access the software development pipeline, including tools like GitHub, Chef, Azure DevOps, and Artifactory, by using an external identity provider with SAML-based authentication. Access or API keys are often used for long-lived CLI sessions or automated script-based access.

Developer Access to Coding Tools

AI/ML engineers and developers write and create many aspects of the code that get pushed to the pipeline on their workstations. They run many software development tools locally as a privileged user, like VS Code, Google Colaboratory, and Jupyter Notebook. These IDE tools often sync local code to cloud code repositories like GitHub, which requires using that identity's access and authentication.

Intra-AI Infrastructure and Machine Identity Access

Application to LLM Service Authentication

This component deals with how the application workload authenticates to the LLM services. Authentication methods include the use of API keys, specific to the LLM service in the cloud provider (AWS access key, Azure application key, and Google API key), ensuring that only authorized applications can invoke the LLM. In some cases, organizations might choose to use IAM role-based access to allow the application to assume roles that authenticate them to access LLM services securely. This approach safeguards the interaction against unauthorized access and ensures that AI-driven processes are secure from external threats.

Application to Server-Side Services Authentication

Authentication between the application and various server-side services ensures that interactions are secure and trusted. The method of authentication used by the application to the server-side services will vary depending on the server-side service itself. Access to a virtual machine is often secured using passwords, SSH keys, or certificates. Authentication to containers is often controlled via mTLS (certificates). Serverless compute functions rely on cloud provider authentication mechanisms, such as an AWS access key, Azure application key, Google API key, and IAM role assumption.

Application to RAG Vector Database Authentication

Authentication between the application and the RAV vector database can be established using several methods, often controlled using passwords for service accounts, API keys, or even IAM role assumption.

Application to Training or Fine-Tuning Dataset Storage Authentication

Securing the application's access to datasets used for training or fine-tuning AI models is often managed through access or API keys. This access can also be secured by IAM role-based access, where the application uses the roles defined in the cloud provider's IAM system to access storage solutions like AWS S3 or Azure Blob Storage. Security teams often control access to other storage services via username and password-based credentials.

LLM Service to RAV Vector Database Authentication

The LLM service must securely access the RAG vector database to retrieve or input data necessary for AI model operations. The default approach often is when the LLM service and the RAV vector database are on the same cloud platform and then leverage native cloud authentication mechanisms like IAM roles.

Identity Security in the GenAI Architecture

To best understand which security controls to apply in which circumstance, organizations also need to understand how bad actors attack the underlying services and infrastructure that the GenAI applications are deployed on. Idira Blueprint outlines the common attack path malicious actors (internal and external) take to compromise identities and execute their endgame.

Understanding the Identity Attack Chain

First, identities are compromised by using such techniques as social engineering, MFA bypass, credential theft, or cookie hijacking. From there, bad actors use the identity's access to move around laterally, looking for more access to power and move vertically when they can, eventually escalating and abusing the privileges they've obtained. Malicious actors will often target AI/ML engineers, developers, cloud engineers, and cloud operations, as these users have high levels of privilege. They will also target the workloads and services in the cloud, as their privileges are often overprovisioned.

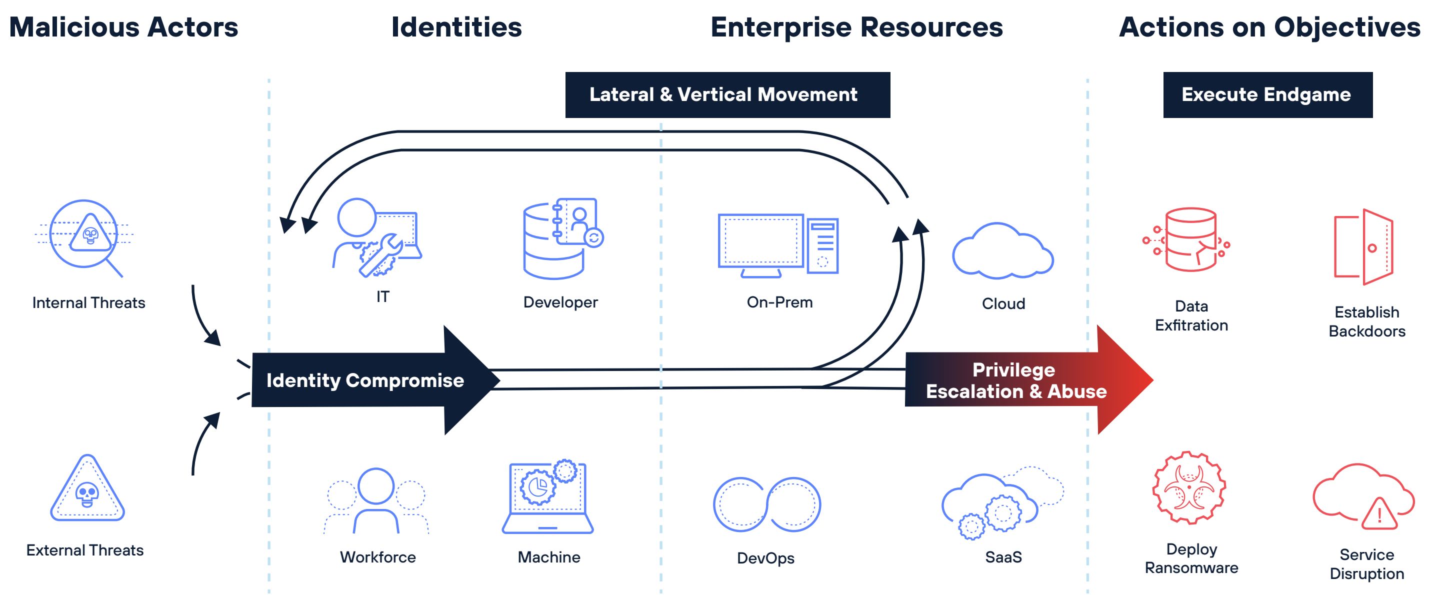

Figure 4. Identity attack chain

For example, a malicious actor might target an AI/ML engineer via a phishing campaign and compromise their standard workforce credentials. Having compromised their workstation, they can also steal any session cookies the developer has to cloud service providers or to the source code repositories where they store code for the LLM. Because that developer has been granted standing access with administrative privileges, that malicious actor can immediately begin to abuse those privileges and perform actions like manipulating the LLM or exfiltrating data.

Identity Security Controls

To holistically address these challenges, organizations need a combination of effective intelligent privilege controls and prioritization. Think of it like securing a building with many doors. If you only lock the front door and keep all the side doors unlocked, your organization is left vulnerable. By adhering to the following recommendations, you'll gain control and visibility into the human-interactive and machine identities with access to your cloud service provider environments and workloads.

Mitigate Risks Associated with the Identity Attack Chain

Organizations should seek to implement security controls that mitigate the risks associated with the identity attack chain. This includes preventing identity compromise and credential theft, stopping lateral and vertical movement, and limiting privilege escalation and abuse. These risks can be mitigated using a combination of zero standing access and secure standing privilege access, secrets management, certificate management, least privilege, and identity governance (lifecycle management and compliance) controls.

Federated Access to the CSP, Its Services, and Workloads

For all federated access to the CSP, its native services (such as the LLM service, API gateway, and S3 buckets) and its workloads (VMs and databases), organizations should strive to achieve ZSPs. They achieve ZSPs by using controls like JIT elevation of access, JIT assignment of entitlements, and limited time-bound durations for the access. By following this approach, organizations reduce the risk of both identity compromise and lateral movement from a targeted identity.

Additional defense-in-depth layers, like session protection, recording, audit, and identity threat detection and response (ITDR), help further deter bad actors. This fundamental shift moves away from traditional freestanding federated access via SSO and standing entitlements. Organizations should strive for all interactive access to align with a ZSP approach.

Nonfederated Access

For nonfederated access, organizations should strive for the objective of achieving secure standing privileges. Use such controls as credential vaulting, password, key management and rotation, complex password policy, MFA, session isolation, session monitoring, audit, and threat detection and response. Apply these controls to any nonfederated, freestanding access directory of user passwords and access or application keys such as shared or emergency access accounts. Overall, strive to minimize the number of freestanding credentials to reduce the attack surface.

Certificate Lifecycle Management

This process entails issuing, provisioning, discovering, renewing, and revoking digital certificates, such as TLS certificates. Ensure that PKI-based authentication of machine identity access stays both secure and operationally functional.

Least Privilege

Least privilege controls include two key concepts: the move to on-demand or as-needed privilege and limiting an identity's permissions to only those needed to perform its function or responsibilities. Moving privilege to an on-demand or as-needed basis, instead of granting always-on access, is a key mechanism to minimize the risk of standing privileges. Privileges are not authorized on the target resources at any point until they are needed. Reducing standing privileges provides immediate risk reduction while organizations refine and limit the permissions and privileges required through privilege analysis and policy creation or modification.

Food for Thought

Holistic GenAI infrastructure security encompasses both the objectives of achieving ZSPs for federated access and secure standing access for nonfederated access. While the emphasis is on ZSP for humans, all CSPs and workload components of a GenAI architecture have built-in local admin-type credentials that also require protection. Furthermore, machine identities with admin access to your GenAI environment require secrets management controls.

Your cloud isn't secure until all objectives are accomplished.

Identity Governance

Identity governance controls include:

- **Lifecycle management** enforces the process of granting authorized identities access to the resources they need (via the appropriate control plane) at the time of hire or inception and revoking access when those identities no longer require them—your traditional Joiner-Mover-Leaver process.
- **Compliance controls** enforce the periodic certification and attestation that identities still require access to the things they currently can access and revoke that access if no longer required.

All Identities

All identities, regardless of access method or whether they're human or machine, should be assigned only the minimum necessary permissions for their job function—inline with the least privilege methodology. This greatly reduces the risk of lateral movement, privilege escalation, and abuse. Similarly, all identities should be subject to identity lifecycle governance to ensure that identities are granted correct access, at the right time, and have it revoked when it is no longer required.

When all these controls are put together properly, you can develop an effective access model with JIT access and ZSPs at the center.

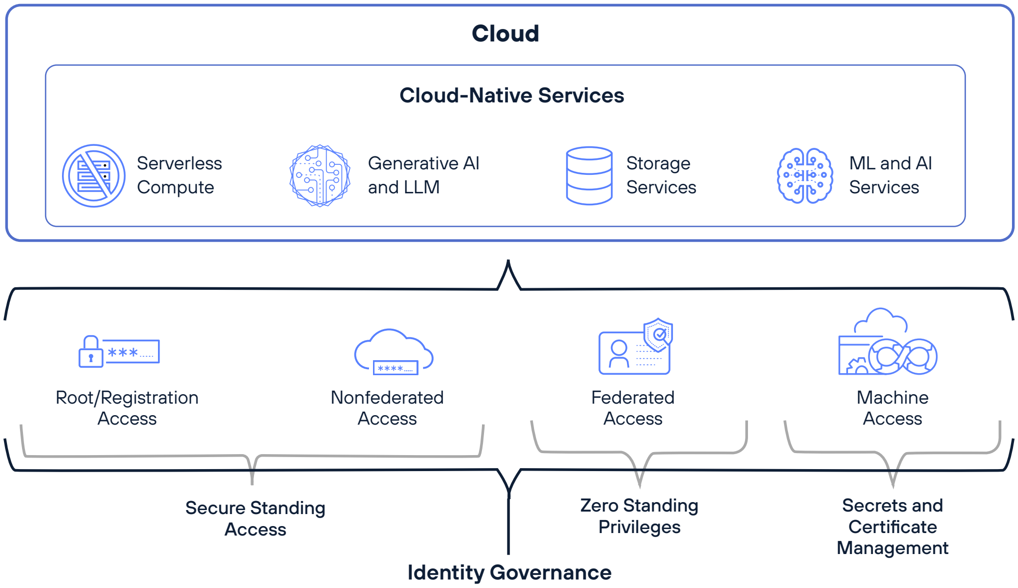

Figure 5. Alignment of access methods and identity security controls

Alignment to Well-Architected Frameworks

The unpredictable behavior of GenAI necessitates advanced oversight mechanisms that span across on-premises environments, cloud infrastructure, and multiple cloud providers. The well-architected frameworks from major cloud service providers outline the guidelines that help organizations build secure, high-performing, and resilient cloud environments by focusing on effective design principles and practices, including those related to identity and access management. These frameworks support any application type or service deployed within the cloud and are applicable to GenAI environments.

We've consolidated and highlighted the seven common principles that Amazon, Google, and Microsoft recommend to be compliant with their frameworks. Secure your cloud environments by aligning your identity security posture and security controls with the following principles:

- **Principle of least privilege:** Assign the minimum necessary permissions to users, processes, and systems to perform their tasks, reducing the risk of unauthorized access. Even when providing access with ZSPs, no user should have unnecessary permissions for the job at hand.
- **Authentication and authorization:** Implement strong authentication mechanisms, like MFA, and ensure proper authorization controls to manage access to cloud resources effectively.
- **Centralized identity management:** Use a centralized identity management system for user authentication and authorization, facilitating better control and monitoring of access across your cloud environment.
- **Credential management:** Regularly rotate and manage credentials securely, including for machine identities like service accounts. Avoid hard-coding credentials and use IAM roles whenever possible.
- **Audit trails and monitoring:** Implement comprehensive logging and monitoring for all identity-related events, enabling timely detection and response to security incidents. Incorporate cloud log solutions and cloud monitoring capabilities.
- **Automated compliance checks:** Employ automated tools and processes to regularly assess and ensure compliance with security best practices, IAM policies, and configurations.
- **Secure DevOps practices:** Integrate security measures into the DevOps pipeline, ensuring that identity and access controls are considered and tested throughout the development lifecycle.

Idira Blueprint weaves these common principles into the fabric of identity security and the recommended controls. An integrated identity security strategy that is aligned with these well-architected guidelines is a critical element of defense against attacks in today's threat landscape. Keep these principles in mind as you develop your own prioritization approach and GenAI security strategy.

Alignment to Secure AI Principles

In addition to well-architected frameworks, Amazon, Google, and Microsoft outline principles and guidelines that help organizations build secure GenAI environments. We've identified the following overarching philosophies for each CSP to secure the GenAI infrastructure.

These common principles of security by design also align with our recommendations, ensuring a foundational approach to identity security within AI and cloud infrastructures. Promoting a robust identity security strategy is crucial for mitigating advanced threats in today's digital landscape. We encourage you to embed these security principles in your GenAI security strategies and identity management systems, bolstering your overall defense mechanism from initial design stages to production deployment.

Amazon Web Services

AWS champions a principle of comprehensive security that is embedded throughout the AI lifecycle. They focus on creating a secure environment from the ground up, ensuring that every stage of the AI process is protected. This approach is characterized by a commitment to robust security practices that safeguard data integrity, protect against unauthorized access, and maintain user confidentiality. AWS advocates for proactive security measures, which include automating security tasks to minimize risks and enhance efficiency, supporting the development of safe and reliable AI solutions.

Microsoft Azure

Microsoft emphasizes integrating security from the earliest stages of AI development, adhering to the principle of security by design. They encourage developers to consider security implications as fundamental components of the AI design and development process. This holistic approach ensures that security considerations are not afterthoughts but are integral to the development pipeline. Microsoft's strategy aims to foster a culture of secure innovation, where security and compliance are seamlessly integrated into the fabric of AI solutions, supporting a trustworthy and resilient AI ecosystem.

Google Cloud

Google promotes a security-centric approach to AI with its Secure AI Framework (SAIF), focusing on the development and deployment of AI in a responsible and secure manner. The core principles of the Google Cloud framework include transparency, accountability, and ethical use of AI. Google stresses the importance of understanding and mitigating AI-specific risks, advocating for a security strategy that encompasses the entire AI system, including data protection, model security, and operational integrity. Their approach is designed to empower organizations to build AI solutions that are both innovative and aligned with best security practices and responsible AI usage.

Prioritization Strategies

While all identities with access to the GenAI application infrastructure are considered privileged—everything from read-only permissions to full administrative access—you have to prioritize them little by little. There are many identities and security controls. Organizations must have a strategy to deploy security controls effectively and efficiently.

A central tenet of our prioritization logic is balancing risk and effort. But how do we define risk? Each organization may include other factors in its definition of risk, such as data classification or sensitivity. Through a common lens, risk is a combination of three factors:

- **Level of privilege** refers to the type of privilege that's been granted to the identity, ranging from read-only access to the ability to modify other identity's permissions and access to full administrative control.
- **Scope of influence** (also called the blast radius) refers to the amount or percentage of systems and resources an identity can access. It can range from access to a single cloud-native service to multiple services with access to elastic workloads to full access to every resource and service.
- **Ease of compromise** refers to how easy or challenging it is for a malicious actor to compromise the access, including the technical vulnerabilities that exist and the level of controls applied to protect the identity.

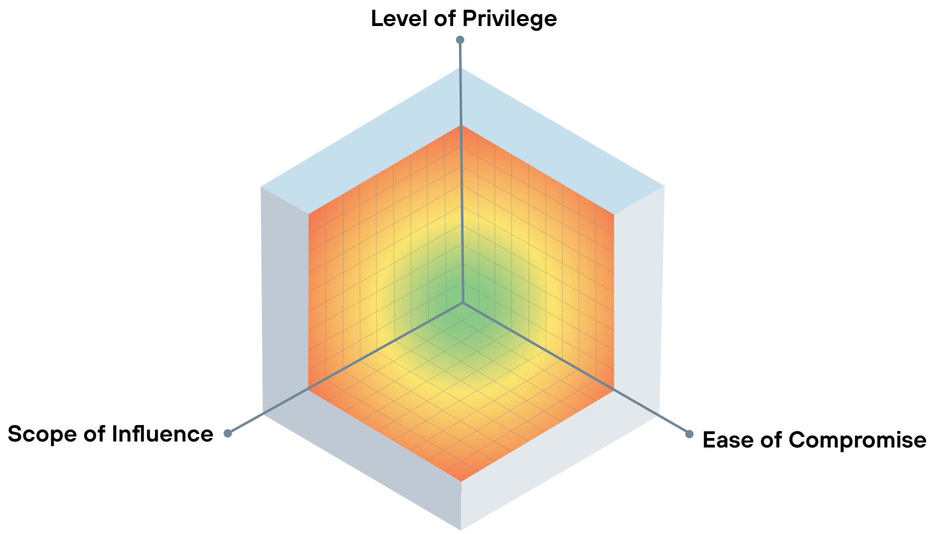

Figure 6. Three elements of identity risk

Idira Blueprint endorses two approaches based on real-world experience. Both are risk-based and maximize impact for the level of effort. One focuses on prioritizing security controls and services you might need to secure the cloud as a whole. The other strategy prioritizes the identity or role that needs to be secured based on their privileges, population, and level of risk.

Security Control-Based Prioritization

In this prioritization method, organizations focus on one security control family at a time. Not every organization has solutions or services to implement all the security controls for each identity. Organizations prioritize by the type of security control they want to apply, which typically correlates directly to a service or solution and the corresponding risk impact and level of effort required to mitigate the risk.

Above all else, always secure your root and registration accounts to your cloud service providers first. Even a simple, temporary one-time passcode (TOTP) MFA application, like what's available in your typical mobile authenticator application, will be valuable here. You don't need to implement full privileged access management (PAM) controls to begin the process of mitigating the risk of these sign-up accounts.

Controls That Support Progress Toward ZSPs

The first control family to implement should support progress toward ZSPs. This includes functionality like role-based federated JIT access, ZSPs, MFA, session protection, session recording, and audit. Privileges should be granted on demand or as-needed to start and, over time, should be refined with the least privileged enforcement. Controls enabling ZSP are considered to have the highest priority because they cover the largest swath of human access by your AI/ML engineers, developers, and IT staff. Implement these controls early to avoid the sprawl of freestanding access that can quickly accumulate as cloud footprints grow.

Controls That Implement ZSP

These controls include credential vaulting, password, and key management and rotation, complex password policy, MFA, session isolation, session monitoring, and audit. Here, return to the root and registration account passwords and access keys first, and then move into the freestanding access (passwords and keys) for those same identities listed before. Work toward minimizing freestanding access at all costs, but when required or necessary, controls like credential management are critical. Strive to refine the privileges for nonbreakglass emergency accounts over time, focusing on well-known roles first and enforcing least privilege.

Secrets Management Controls

These controls include secrets vaulting, secrets rotation, complex secret value policy, removal of hard-coded secrets from applications, dynamic secrets, and JIT secret delivery to those apps. Secrets management controls build on the foundation of PAM and require an additional discovery and prioritization effort, which is why they follow PAM controls. Strive to refine the privileges for machine workloads over time to enforce least privilege. Apply these controls to any machine passwords and keys that workloads, scripts, or services consume so you can mitigate the risk of identity compromise and privilege abuse.

Certificate Lifecycle Management Controls

These controls consist of issuing, provisioning, discovering, renewing, and revoking digital certificates for various machine identities and machine-to-machine communication. Work toward managing all certificates related to the GenAI application and infrastructure to ensure the PKI-based authentication of machine identity access stays both secure and functionally operative.

Identity Governance Controls

Identity governance consists of lifecycle management and compliance mechanisms. Lifecycle management is the process of granting authorized users access to the resources they need (via the appropriate control plane) while hiring and revoking their access when they no longer require it—the traditional Joiner-Mover-Leaver process. Roll out these controls to the explicitly defined roles first, like the engineers, developers, and IT admin staff.

Controls Rollout

Lifecycle management requires the explicit definition of roles to be effective, so organizations should focus on the well-known roles first. Identity compliance is all about periodically certifying and attesting that users still require access to the things they currently can access and if not, revoking that access. Roll out identity compliance controls simultaneously across all human users. This comes last in the list of controls because users must first have access to resources via control planes to conduct compliance campaigns against them.

Zero Standing Privileges

- Least privilege role-based access
- Zero standing privileges
- Federated just-in-time access
- Adaptive MFA
- Session protection and audit

Secure Standing Privileges

- Credential vaulting
- Rotation and isolation
- MFA
- Session isolation, monitoring, and audit
- Protection for shared and breakglass emergency accounts

Secrets Management

- Secrets vaulting
- Rotation and complex secret value policy
- Removal of hard-coded secrets
- Dynamic secrets and just-in-time secret delivery

Certificate Management

- Certificate discovery
- Issuance of digital certificates for machine identities
- Provisioning certificates
- Renewal processes
- Revocation of certificates

Identity Governance

- Joiner-Mover-Leaver processes
- Grant authorized identity access at the time of hire
- Revoke access at time of change or departure
- Certification and attestation processes for privileged access

Figure 7. Security control prioritization

Identity- and Persona-Based Prioritization

In this prioritization method, the organization can apply multiple security control families simultaneously to a persona or type of access. Not every organization has that capability, but the benefit is that this approach is exclusively risk-based. Even if your organization does not have that capability, keep this information in mind as you build your GenAI security strategy.

This prioritization method assumes the organization is applying all security controls to each type of identity at the same time, leaving no gap or hole in any of these types of access. This is the theoretical ideal scenario in the event that you have the capability to deliver comprehensive identity security controls simultaneously.

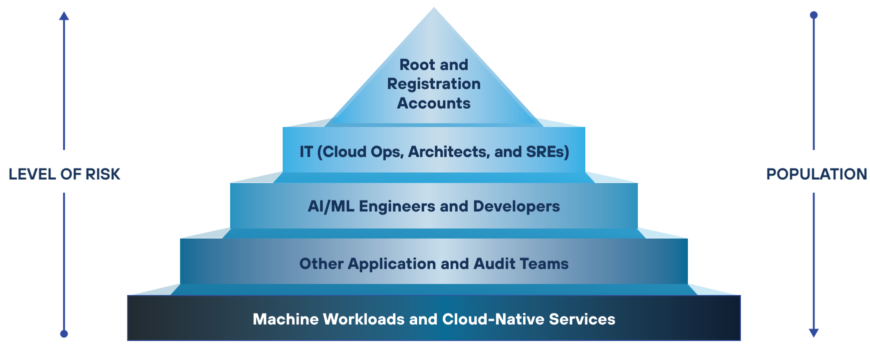

Figure 8. Identity or persona risk prioritization pyramid

Root and Registration Accounts

Always start with these accounts first. Each one represents the email address used to register for the cloud service provider account. Only one of these accounts exists for each CSP account you have, so the population is small. These accounts can do anything on your cloud service and should have a minimum of vaulting and MFA applied. Ideally, however, they should have additional secure standing privileges. In AWS, this is your root, organizational root, or member root user. In Azure, this is the global administrator you signed up with. In Google Cloud, this is the owner you signed up with. In the case of Azure and Google, you can likely identify this account by looking for a shared corporate email address (e.g., LoBCloudProvider@example.com).

IT Administrator Personas

These personas are typically internal roles such as cloud operations, cloud architects, and site reliability engineers. They are usually granted full-fledged administrative access across the cloud service provider accounts. These identities are the highest risk because they have explicit access to control all aspects of IAM and permissions, meaning they can create new identities, modify permissions, and manipulate all aspects of the CSP and its resources. They have the ultimate privileged access to affect every service and resource within the CSP account. These users should have ZSP access to the CSPs and should be additionally protected with least privilege and identity governance controls. Also, fewer IT-specific staff support these cloud provider accounts than the developers who build and operate in the cloud.

AI/ML Engineers and Developers

These staff members have privileged access to various services and resources within the CSPs, like serverless compute, databases, or secret vaults. Unlike IT admins, these identities are not guaranteed power across the cloud service provider accounts, but they can be all-powerful for the services they access or administer and sometimes are granted full administrative access. These identities should have a combination of secure standing and operational ZSP access with careful consideration given to least privilege and identity governance.

Other Application and Audit Teams

Following developers, apply the same controls to other application and audit teams and users with fewer privileges like read-only access to various services.

Machine Workloads, Cloud-Native Services, and Other Application Identities

Lastly, prioritize machine workloads, cloud-native services, and other application identities. As AI-driven systems continue to evolve, their presence within cloud environments will increase and some might hold high-privilege roles, such as administrative or service-level permissions. This growing population of machine identities—particularly AI agents—introduces new security risks that must be proactively managed. To mitigate these risks, implement a combination of secrets management, certificate management, and least privilege controls.

Prioritize What Makes Sense for Your Organization

Every organization has unique prioritization needs. Looking at this list, you might see areas that you want to implement first before the others. In that case, choose the order that makes the most sense for your organization. Your priority is whatever is driving your initiatives and is important to your organization's goals. Consider this prioritization guidance and rationale so you understand the tradeoffs and make the most informed decision for your organization.

Table 1. Identity or Persona Prioritization and Control

Identity/Controls	Zero Standing Privileges	Secure Standing Privileges	Secrets and Certificate Management	Identity Governance
Root and Registration Accounts		X		X
IT (Cloud Ops, Cloud Architects, and SREs)	X	X		X
AI/ML Engineers and Developers	X	X		X
Other Application and Audit Teams	X	X		X
Machine Workloads and Cloud-Native Services			X	X

Alignment to GenAI Adoption Strategies

Many organizations are just starting to explore the use of custom, in-house developed GenAI applications. They have a critical and opportune moment to cement their GenAI initiatives with the secure by design principle. This philosophy entails embedding security deeply within the architecture of enterprise resources and the processes that create and access them.

Traditionally, many organizations deploy resources first and then layer security measures on top, often resulting in a reactive approach to identity security. However, with the accelerating pace of AI innovation, this method leaves a widening gap in identity security that can be exploited by malicious actors. Because these technologies can rapidly scale and evolve, embedding security from the start mitigates risks and aligns with best practices for future-proofing investments in AI.

By using automation and orchestration services, like Terraform or Ansible, organizations can automatically integrate robust security controls as they provision new resources. This includes the onboarding of new cloud service provider accounts to an identity security control plane that enforces ZSPs and secrets

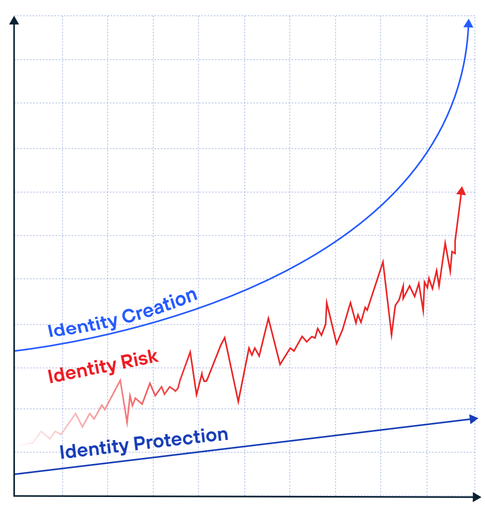

Figure 9. Widening gap of identity risk

management, which facilitates secure human access and effective machine identity security for cloud-native workloads. By integrating security at the inception of resource creation, organizations can ensure that every asset is secure from the outset, significantly narrowing the window of opportunity for breaches.

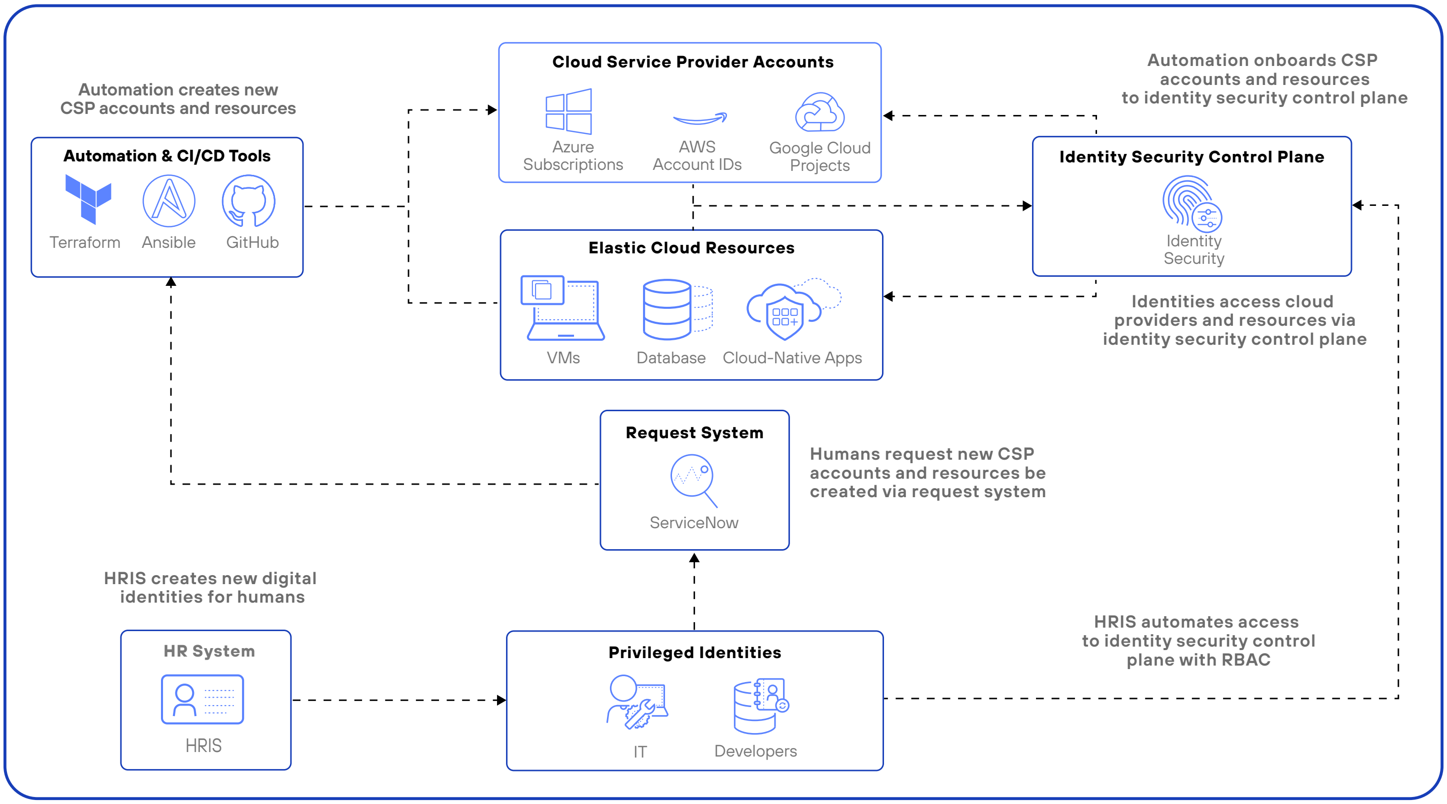

Figure 10. Secure by design in the cloud

Adopting a secure by design approach enables organizations to maintain a proactive security posture, significantly reducing technical and security debt over time. For enterprises at the cusp of designing their GenAI systems, this approach is a security strategy and a fundamental business imperative. It ensures that as the organization scales and diversifies its technological capabilities, the foundational security infrastructure evolves concurrently, enabling a secure, resilient, and compliant operational environment.

This proactive stance on security is essential in an era where the pace of technological change outstrips the speed of traditional reactive security measures, making secure by design a critical strategy for the modern digital enterprise.

Accenture and Palo Alto Networks: Secure AI

As AI continues to transform industries, organizations must prioritize security to safeguard their AI-driven innovations. In this rapidly evolving landscape, we see our clients needing robust security measures to address challenges such as unauthorized access, data integrity, and adversarial threats.

Securing AI in the Era of Reinvention

The Palo Alto Networks product portfolio, combined with Accenture's expertise in AI security, provides a robust foundation for securing AI models, data, and infrastructure against emerging risks.

- **IAM:** Organizations must ensure that only authorized users and systems can interact with sensitive AI assets. IAM solutions help minimize risks associated with unauthorized access, privilege escalation, data poisoning, model manipulation, and adversarial attacks.
- **API security:** AI models often expose APIs for interaction and integration. It has become increasingly essential to safeguard these endpoints by enforcing authentication, access controls, and real-time anomaly detection to prevent unauthorized usage and data exfiltration.
- **Runtime monitoring and threat detection:** Continuous and proactive monitoring of AI workloads is critical for detecting deviations in model behavior that might indicate adversarial attacks or system compromise.
- **Credential and secrets management:** As AI applications increasingly depend on machine-to-machine communication, securing API keys, certificates, and credentials is critical to preventing unauthorized access and maintaining system integrity.

By integrating these security solutions, organizations can begin their journey to establish a resilient AI security framework that aligns with best practices for protecting AI models, data, and infrastructure.

Securing AI Agents: The Future of Autonomous Security

With the rise of AI-driven automation, enterprises are increasingly adopting AI agents to drive efficiency and innovation. Scaling AI agents, however, introduces unique security challenges. Organizations need solutions that ensure trust, governance, and compliance as AI agents become embedded in business workflows. In response, we have developed pioneering security solutions tailored for securing the agentic architecture.

Imperatives and Innovative Solutions for Securing AI Agents

The growing deployment of AI agents requires a security-first approach. Along with Accenture, we've outlined four key imperatives we have jointly developed innovative solutions for:

- **Exponential agent growth**
  - Billions of AI agents will be deployed in the coming years, demanding scalable IAM solutions.
  - Each AI agent requires a distinct identity, permission set, and access pattern to prevent unauthorized actions and ensure compliance.
- **A nonnegotiable security-first architecture**
  - Secure-by-design principles must be embedded from day one to instill trust in AI-driven enterprise environments.
  - Zero trust frameworks become critical as AI agents gain autonomous decision-making capabilities, ensuring every access request is continuously verified.
- **Critical importance of unified governance**
  - AI agent governance must be centralized, providing a single-pane-of-glass view into agent activity, permissions, and interactions.

    - Complex agent-to-agent and agent-to-human interactions require real-time monitoring, access control, and auditability.
    - Regulatory compliance and accountability mechanisms must be enforced to track AI agent actions and ensure ethical AI usage.
- **Forthcoming regulatory requirements**
    - Emerging AI regulations will necessitate robust security controls for autonomous agents, pushing organizations to adopt stringent governance frameworks.
    - Proactive compliance with AI security mandates will future-proof enterprises, reducing legal and operational risks associated with AI agent deployment.

Driving AI Security Leadership Together

In this era of AI reinvention, organizations require forward-thinking security solutions to enable innovation with confidence. By combining our industry-leading security capabilities with Accenture's deep expertise in AI strategy and implementation, we empower enterprises to embrace AI transformation securely. Our partnership ensures that businesses can deploy AI systems and agents at scale while maintaining security, trust, and compliance at every step.

Concluding Insights

- **Inclusive security strategies:** GenAI application security encompasses developers, IT teams, third parties, and various machine identities. This necessitates strategies that address the diverse needs and access patterns of these personas and roles.
- **Efficiency with Idira Identity Security Blueprint:** Idira Blueprint can help organizations more quickly and effectively secure privileged entities within their GenAI environments.
- **Risk-based approach:** The most pragmatic strategy for securing identities should focus first on areas with the most significant impact based on risk.
- **Secure by design:** Our guidance follows major cloud providers' architectural best practices and regulatory frameworks, ensuring it works well with existing cloud setups and meets compliance needs.

Next Steps for Securing Your GenAI Infrastructure

Organizations today must lead their GenAI initiatives with a secure by design approach, protecting all of their infrastructure and identities from the start. While CSPs might seem vast and GenAI architectures might seem complicated, Idira Blueprint provides prioritized best practice guidance that can help your organization develop an informed, risk-based plan for success.

See how Idira Blueprint can help your organization build the right framework to secure its cloud identities. Request a demo.

About Palo Alto Networks

Palo Alto Networks (NASDAQ: PANW), the global AI cybersecurity leader, protects our digital way of life with a comprehensive portfolio of cybersecurity solutions and platforms across Network, Cloud, Security Operations, AI, and Identity. Trusted by more than 70,000 customers and powered by Unit 42® threat intelligence, our AI-driven platforms eliminate complexity, empowering enterprises to modernize with confidence and securing the speed of innovation. Explore the future of security at www.paloaltonetworks.com.

3000 Tannery Way
Santa Clara, CA 95054

Main: +1.408.753.4000
Sales: +1.866.320.4788
Support: +1.866.898.9087

www.paloaltonetworks.com

idira_wp_securing-the-backbone-of-gen-ai-with-idira-blueprint_050726