20TH ANNIVERSARY EDITION

WSO2CON ASIA

PLATFORMLESS MODERNIZATION

Zero-Trust Agents: Secure & Govern GenAI Responsibly



Hello!



Arshardh Ifthikar
Technical Lead
WSO2



Ayesha Dissanayaka
Architect & Associate Director
WSO2



The AI-Powered Business Landscape

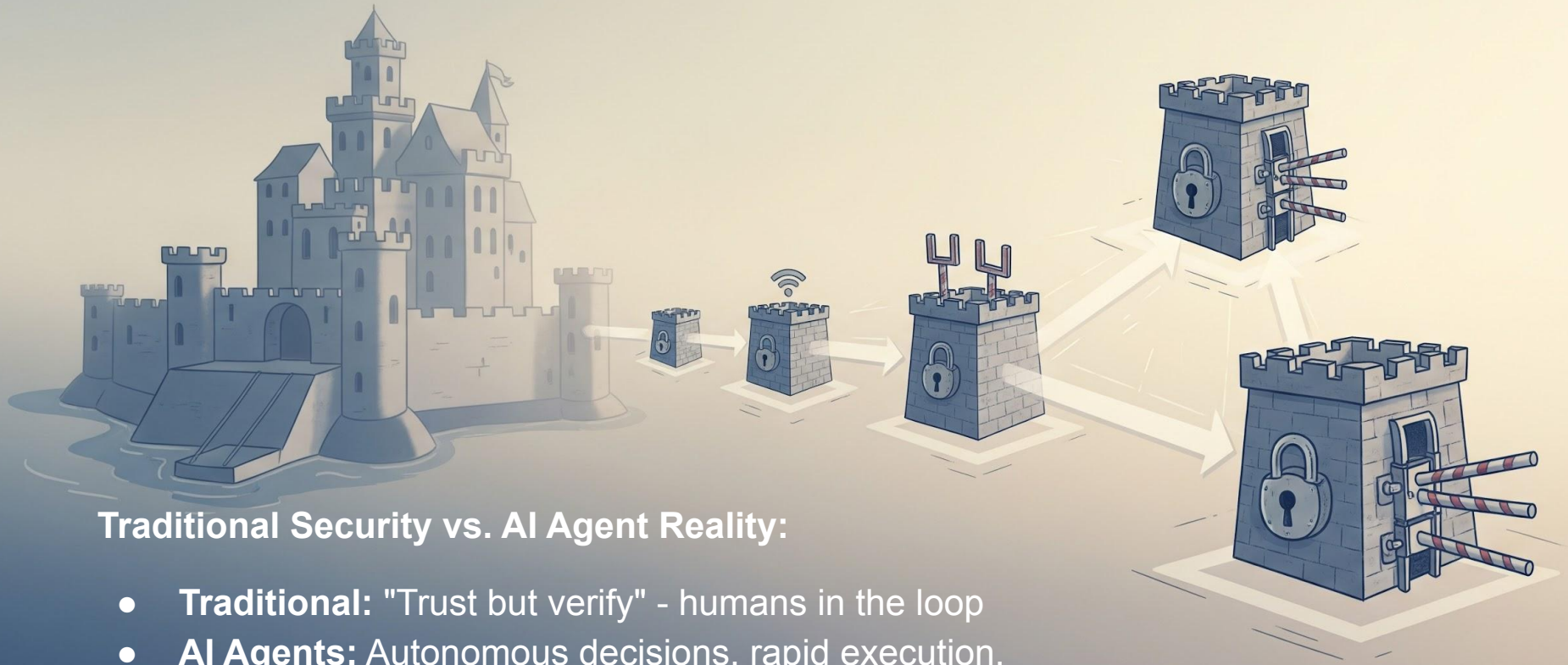
The Value of AI in Modern Business

- **Personalized Service Delivery:** AI-powered assistants provide personalized, 24/7 support.
- **Operational Efficiency:**AI agents can automate complex workflows that can dramatically reduce manual effort and improve response times.
- **Innovation Acceleration:** AI opens doors to new products, services, and business models.

Challenges of Governing and Securing AI in the Enterprise

- Governance
  - **Data Privacy Risks:** Sensitive data may be exposed during AI interactions, and model outputs can be unpredictable or leak proprietary information.
  - **Lack of Trust & Governance:** Hallucinations, bias, and non-deterministic behavior make AI outputs hard to verify, complicating auditability and compliance.
  - **Integration Complexity:** Connecting AI with existing systems, APIs, and workflows requires significant engineering effort and architectural changes.
  - **Cost & Resource Constraints:** Large models demand expensive infrastructure and monitoring to maintain performance and reliability at scale.
- Security
  - **Unauthorized Data Access**: Agent accesses sensitive customer data across all departments
  - **Privilege Escalation**: Marketing chatbot suddenly has admin rights to financial systems
  - **Untracked Actions**: No audit trail of what the agent did or when
  - **Impersonation Attacks**: Malicious actors hijack agent identities
  - **Compliance Violations**: GDPR, HIPAA, SOX violations with massive fines

Zero trust in Agentic AI



Traditional Security vs. AI Agent Reality:

- **Traditional:** "Trust but verify" - humans in the loop
- **AI Agents:** Autonomous decisions, rapid execution, broad system access

Image source: AI generated

Zero-Trust Principles for AI Agents

- Verify Identity
  - Every AI agent must authenticate and maintain identity
  - Continuous validation of agent legitimacy
- Just in time / Just enough access:
  - Agents receive minimum permissions for their specific tasks
  - Dynamic permission adjustment based on context
- Assume Breach
  - Design systems expecting agent compromise
  - Limit blast radius of potential security incidents
- Continuous Monitoring
  - Real-time audit of agent actions and decisions
  - Behavioral analysis for anomaly detection

Gardeo Leisure Digital Services

A platform that helps customers plan and book their ideal trip:

- 🏨 Hotel Discovery
- 📅 Availability Check
- ✅ Hotel Booking
- 🤵 Concierge Allocation

and More

Source code: in Github

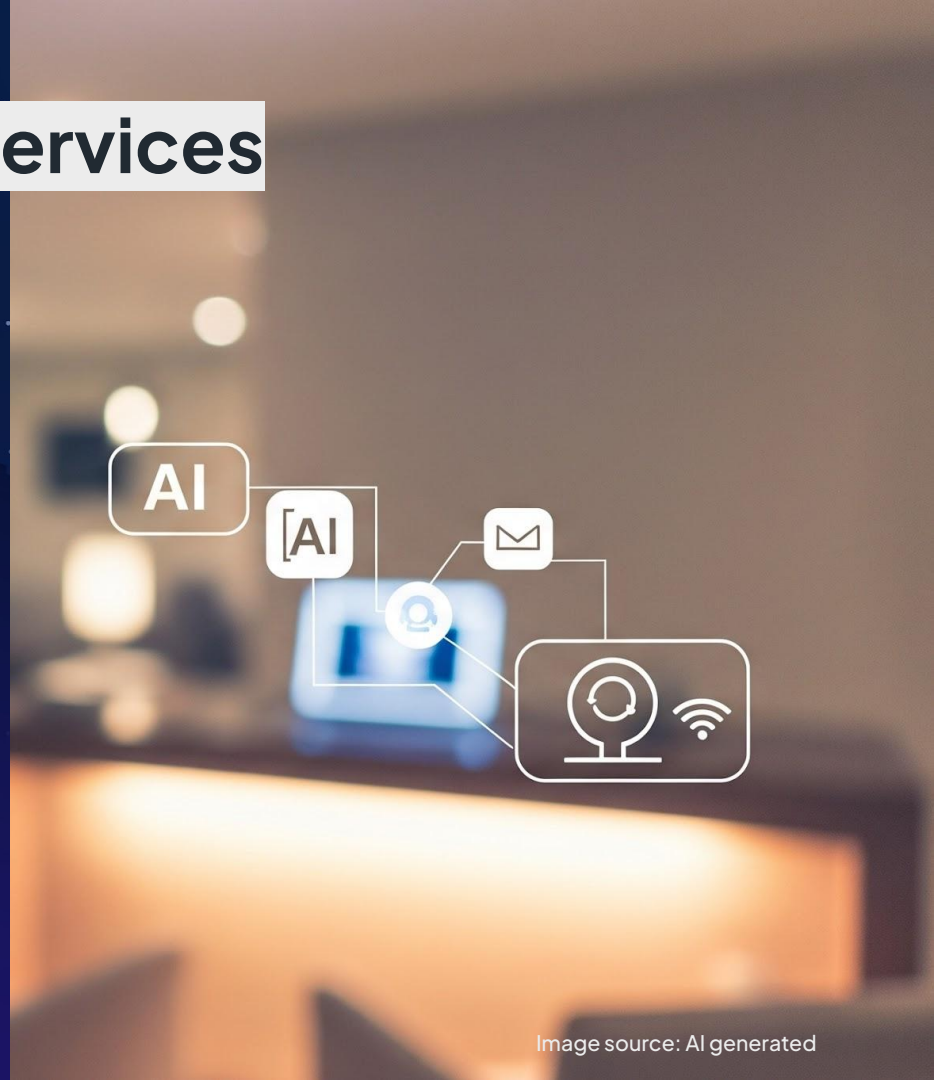
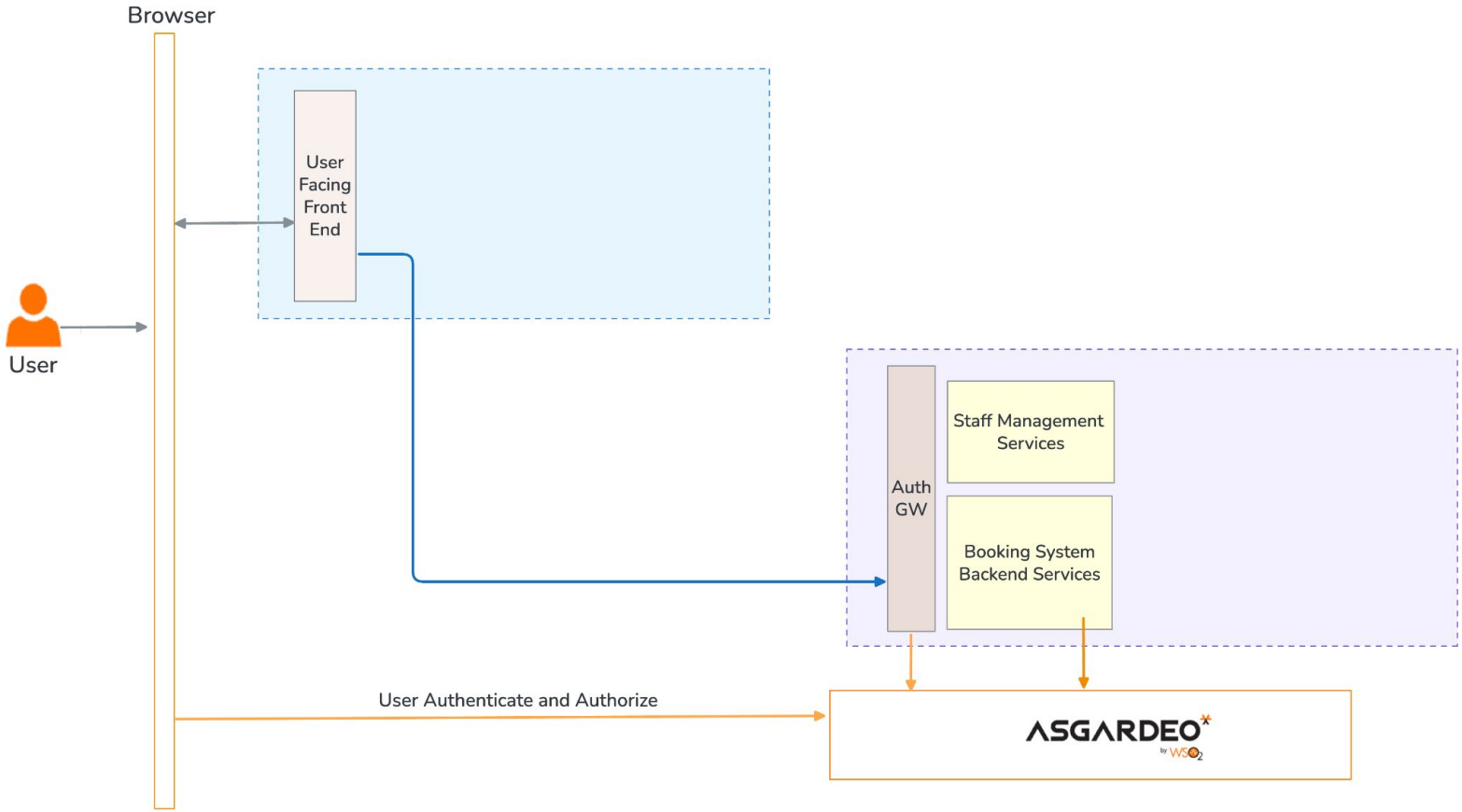


Image source: AI generated

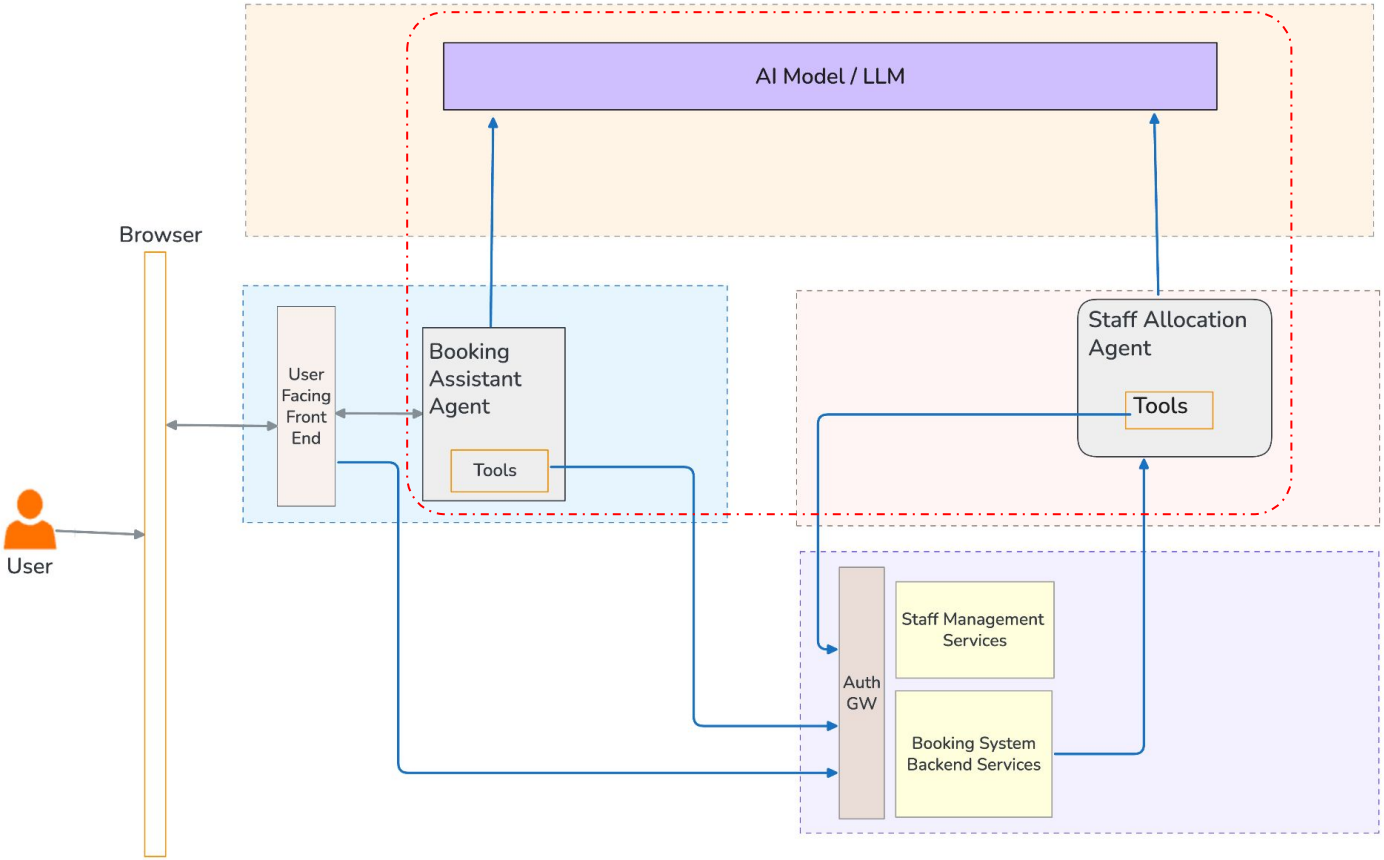
Demo : Gardeo Leisure Platform Without Agentic AI

Gardeo Leisure Platform Without Agentic AI



Introducing Agentic AI into Gardeo Platform

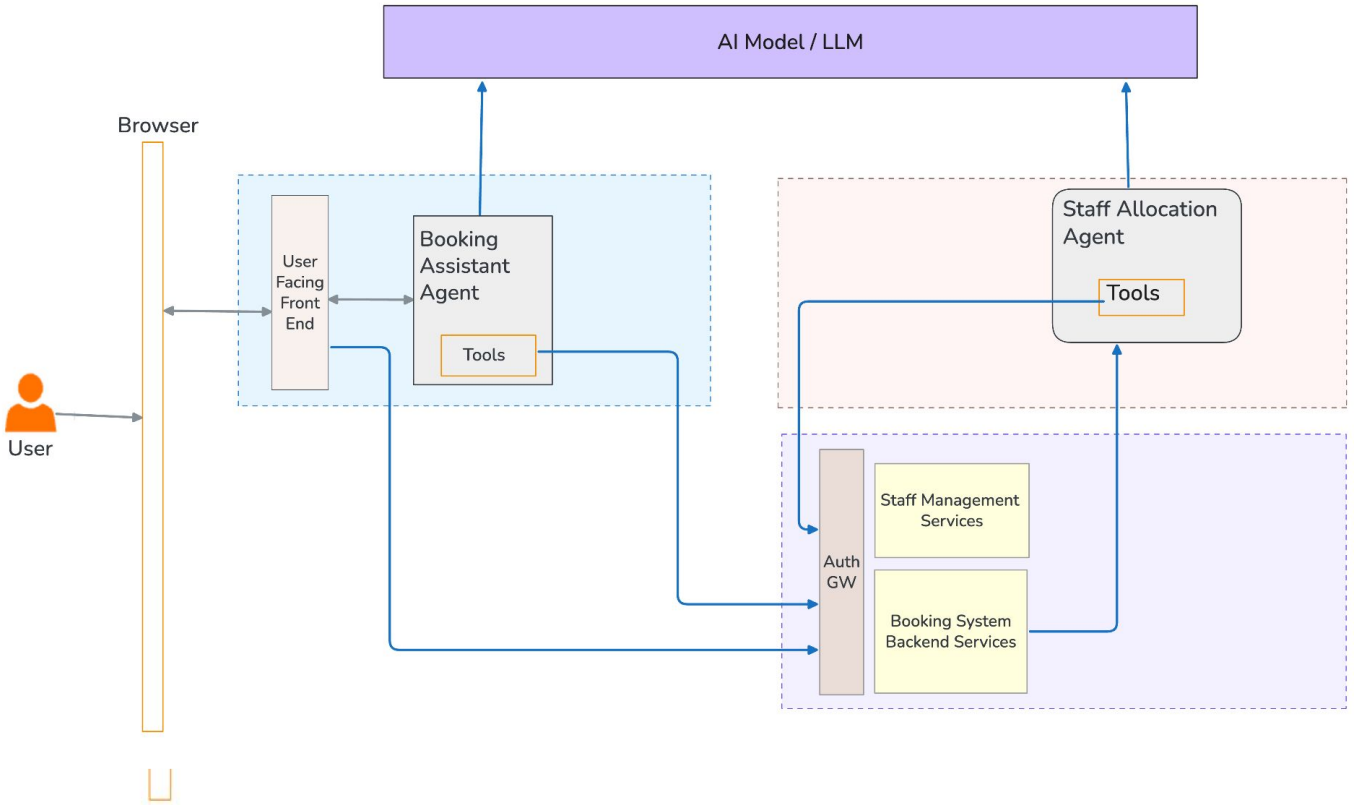
With Agentic AI



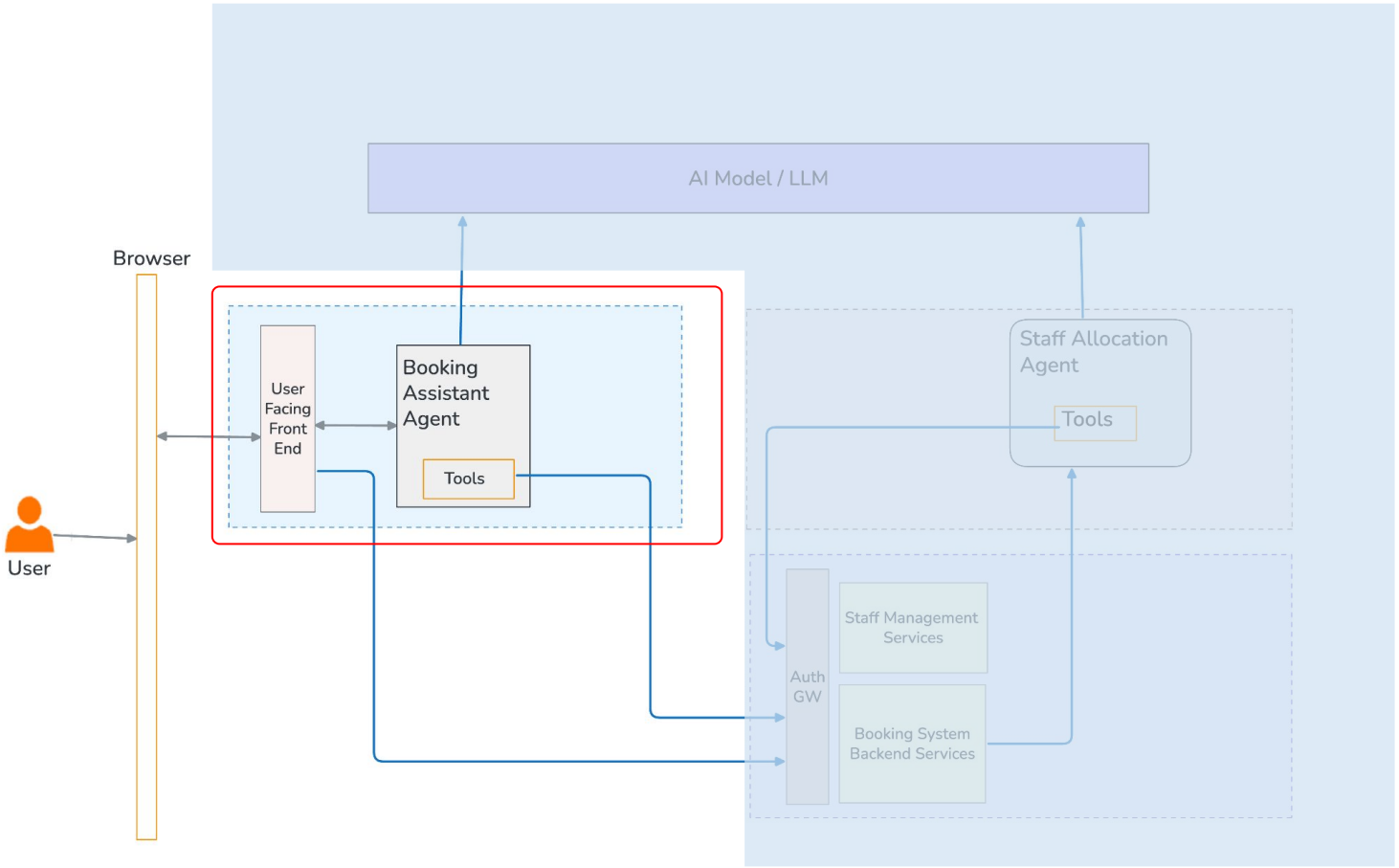
Source code: in Github

Where to apply Zero-Trust

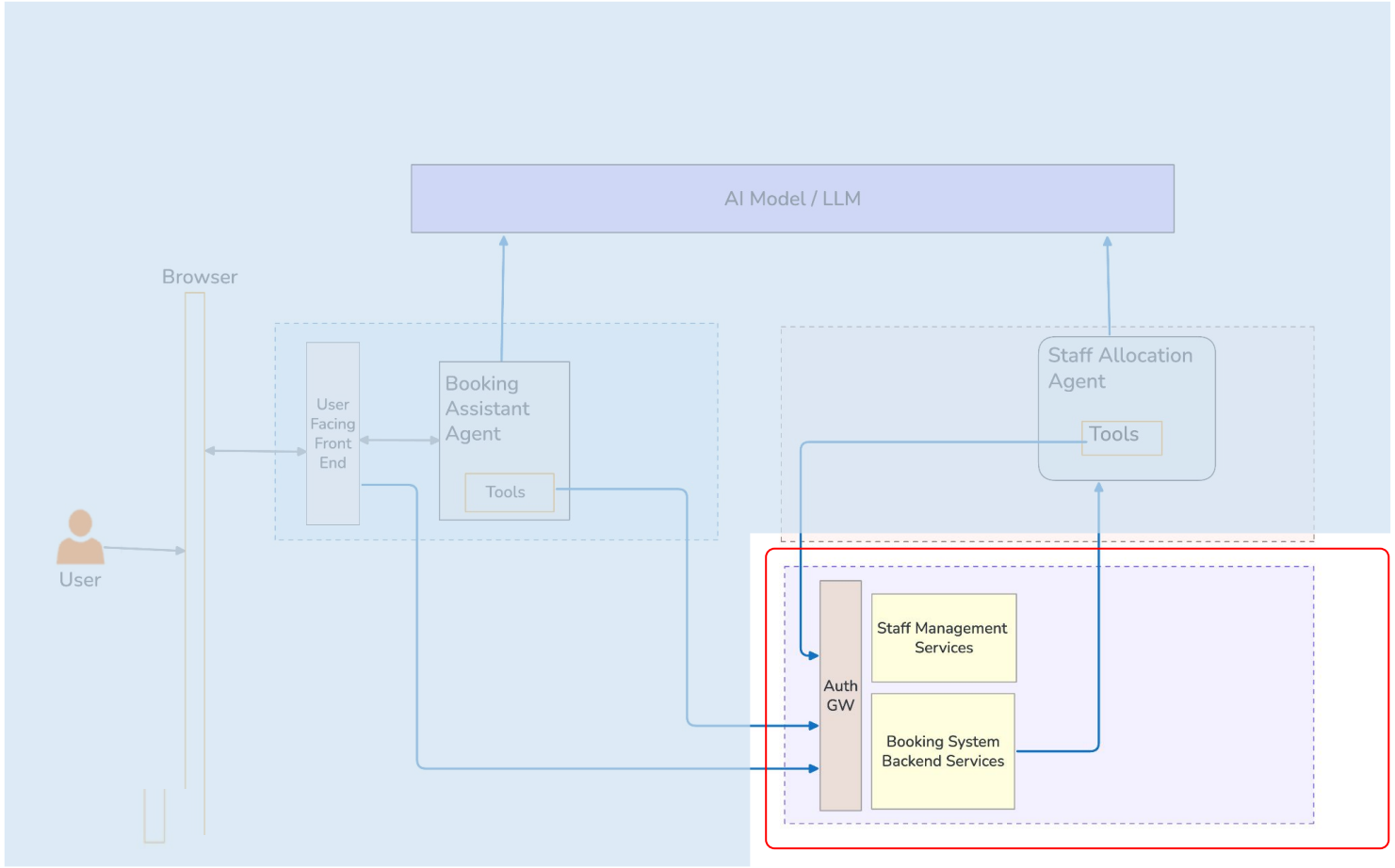
Defining Trust Boundaries



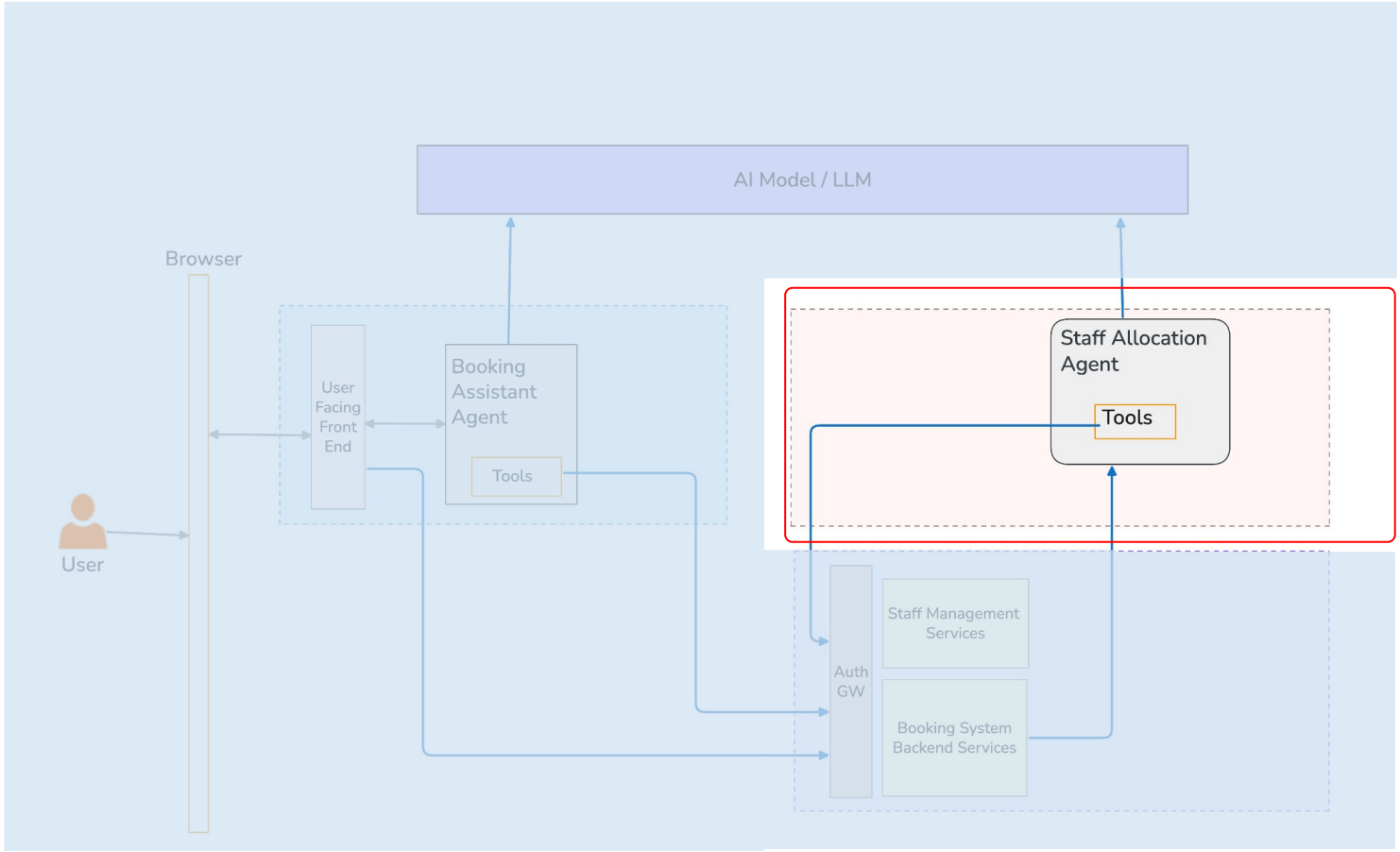
Defining Trust Boundaries



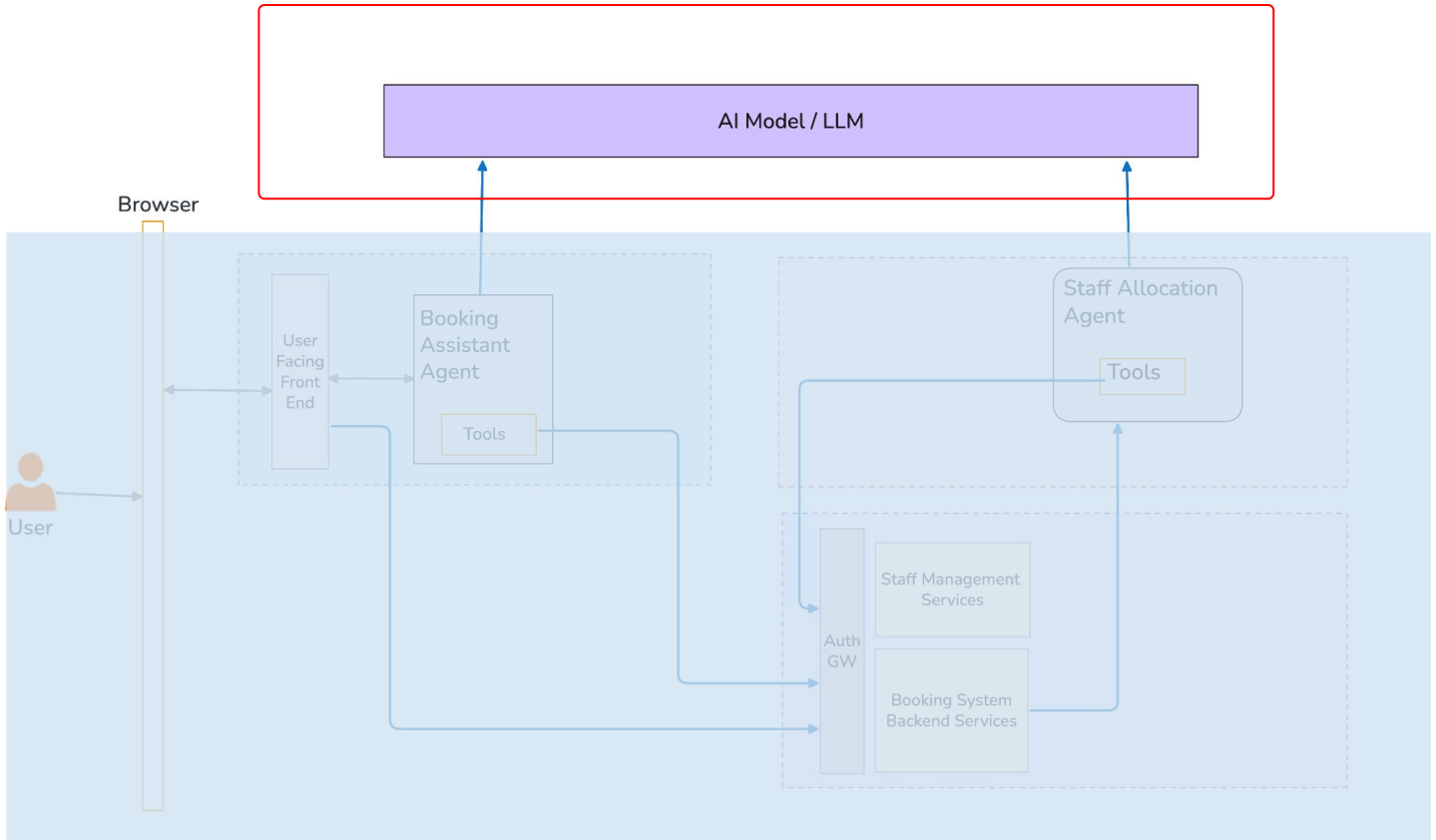
Defining Trust Boundaries



Defining Trust Boundaries



Defining Trust Boundaries



How do we secure agent interactions

- **The Problem:** If an agent calls an API, who made the call? An anonymous service? The end-user?
- **The Solution:** Every AI agent must have its own, unique identity within your system.
- **Why It's Critical:**
  - **Authentication:** The system knows *which* agent is making the request.
  - **Authorization:** You can grant specific permissions to each agent (e.g., the booking agent can *read* room availability but not *change* staff schedules).
  - **Auditing:** You have a clear, immutable record of every action taken by every agent. This is essential for security, compliance, and debugging.



Giving Agent an Identity

Asgardeo Platform Capabilities

1 **Register & Manage Agents**

- Unique identities with metadata
- Ownership and responsibility chains
- Lifecycle Management

2 **Issue Credentials**

- Agent Friendly Credentials
  - MTLS
  - Pvt Key JWT
- Agent-specific authentication mechanisms

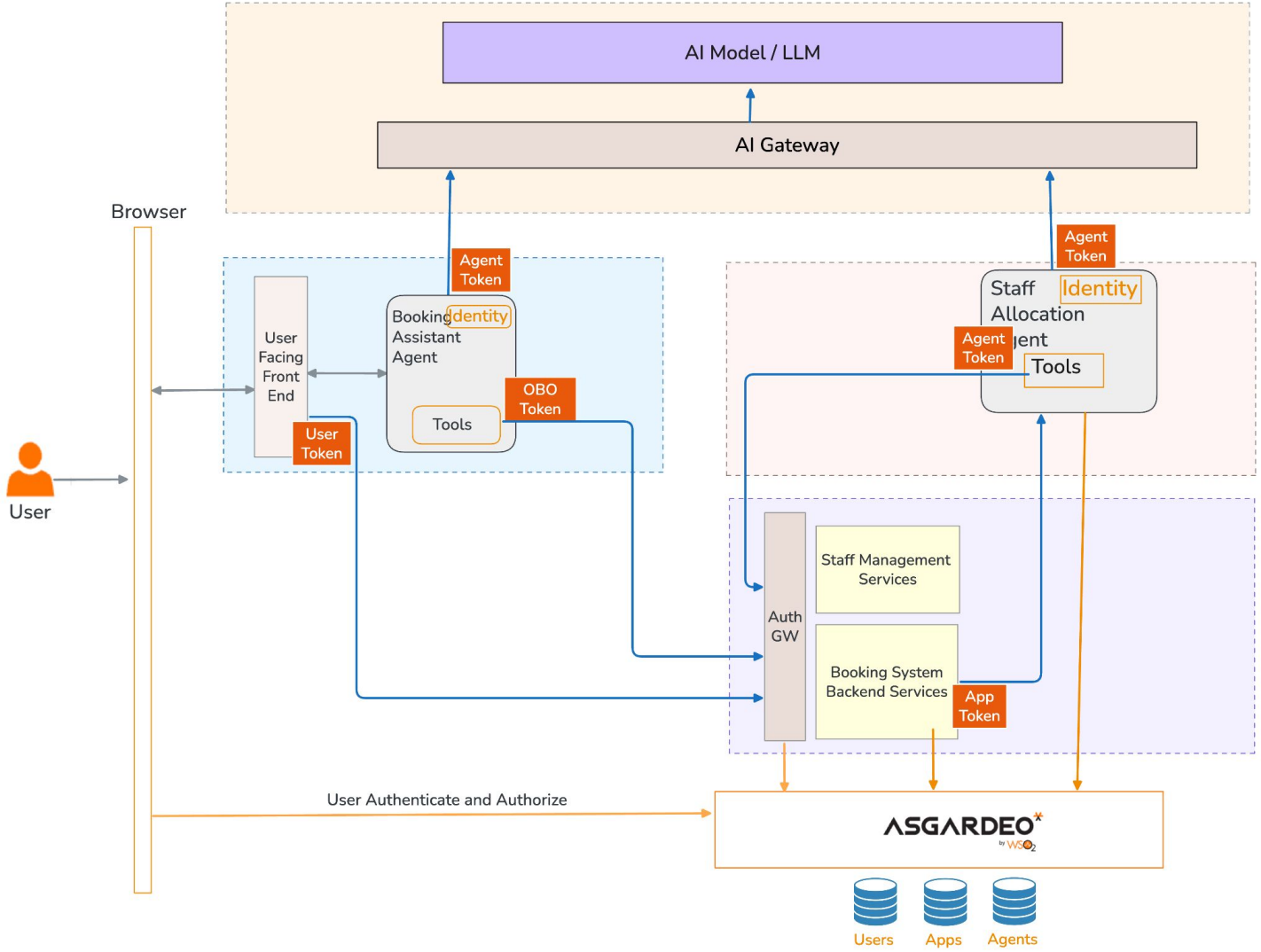
3 **Apply Access controls**

- Zero standing privileges
- RBAC, ABAC and more
- Secure Delegation when acting on-behalf

4 **Enable Audit & Monitoring**

- Complete action traceability
- Monitor specific agent activities

Demo : AI Agents Securely in Action


AI Model / LLM
AI Gateway
Browser
Agent Token
User Facing Front End
Booking Assistant Agent
Identity
OBO Token
Tools
User Token
User
Agent Token
Staff Allocation Agent
Identity
Agent Token
Tools
Auth GW
Staff Management Services
Booking System Backend Services
App Token
User Authenticate and Authorize
ASGARDEO
by WSO2
Users
Apps
Agents

User Direct Booking

```
- GET /api/hotels/4
api/hotels/4 HTTP/1.1" 200 OK
- GET /api/hotels/4/reviews - limit: 5, rating: None
api/hotels/4/reviews?limit=5 HTTP/1.1" 200 OK
- GET /api/hotels/4

api/hotels/4 HTTP/1.1" 200 OK
- GET /api/hotels/4/reviews - limit: 5, rating: None
api/hotels/4/reviews?limit=5 HTTP/1.1" 200 OK
- POST /api/bookings | sub: 591c2a1c-2d08-4b1f-b8ff-cbfa46a0061e | act: None | booking
- Token claims - sub: 591c2a1c-2d08-4b1f-b8ff-cbfa46a0061e, act: sub=None
- Act claim details - type: <class 'app.dependencies.Actor'>, value: sub=None
- Act.sub: None
```

Agent OBO User

```
e9bba747-e91a-4a77-a1ef-1be7dba84bec/bookings HTTP/1.1" 200 OK
/users/e9bba747-e91a-4a77-a1ef-1be7dba84bec/bookings | sub: e9bba747-e91a-4a77-a1ef-1be7dba84bec | act
e9bba747-e91a-4a77-a1ef-1be7dba84bec/bookings HTTP/1.1" 200 OK
i/hotels/search - location: colombo
s/search HTTP/1.1" 200 OK
i/bookings | sub: 591c2a1c-2d08-4b1f-b8ff-cbfa46a0061e | act: 3556c8fe-3989-429a-8ba1-3e66cf58a7ea | b
aims - sub: 591c2a1c-2d08-4b1f-b8ff-cbfa46a0061e, act: sub='3556c8fe-3989-429a-8ba1-3e66cf58a7ea'
m details - type: <class 'app.dependencies.Actor'>, value: sub='3556c8fe-3989-429a-8ba1-3e66cf58a7ea'
 3556c8fe-3989-429a-8ba1-3e66cf58a7ea
oking detected - agent_id: 3556c8fe-3989-429a-8ba1-3e66cf58a7ea, user_id: 591c2a1c-2d08-4b1f-b8ff-cbfa
```

Agent Direct Staff Allocation

```
46 - app.main - INFO - GET /api/admin/staff/available | sub: ef7c14c9-23b9-49a4-8008-7a734daeec
46 - app.main - INFO - Found 1 available staff members for hotel 3
113.36:41696 - "GET /api/admin/staff/available?hotel_id=3 HTTP/1.1" 200 OK
47 - app.main - INFO - PATCH /api/admin/bookings/3 | sub: ef7c14c9-23b9-49a4-8008-7a734daeec7d
47 - app.main - INFO - Assigned staff 5 to booking 3
113.36:41730 - "PATCH /api/admin/bookings/3 HTTP/1.1" 200 OK
48 - app.main - INFO - GET /api/users/591c2a1c-2d08-4b1f-b8ff-cbfa46a0061e/bookings | sub: 591c
```

Benefits & Outcomes

Security Benefits:

- **Zero Standing Privileges** – Agents get only what they need, when they need it
- **Complete Audit Trail** – Every action traced to responsible agent and authorizing user
- **Risk Mitigation** – Real-time threat detection and automatic response
- **Compliance Assurance** – Automated regulatory requirement adherence

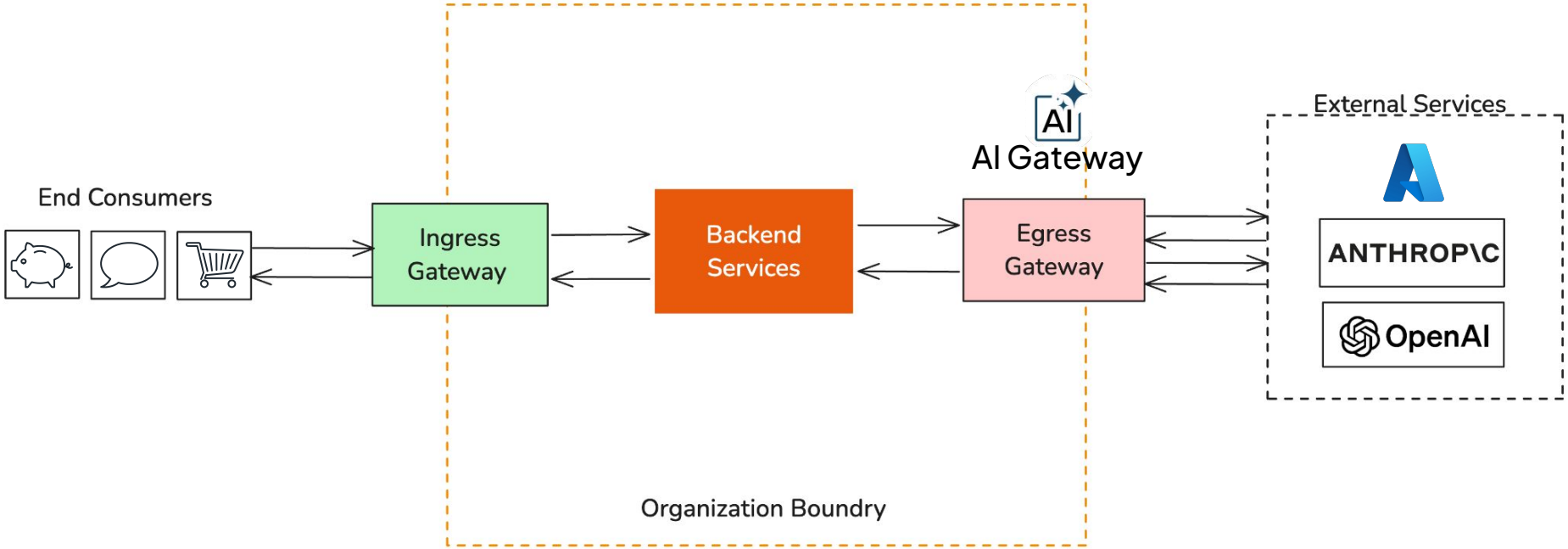
Benefits & Outcomes

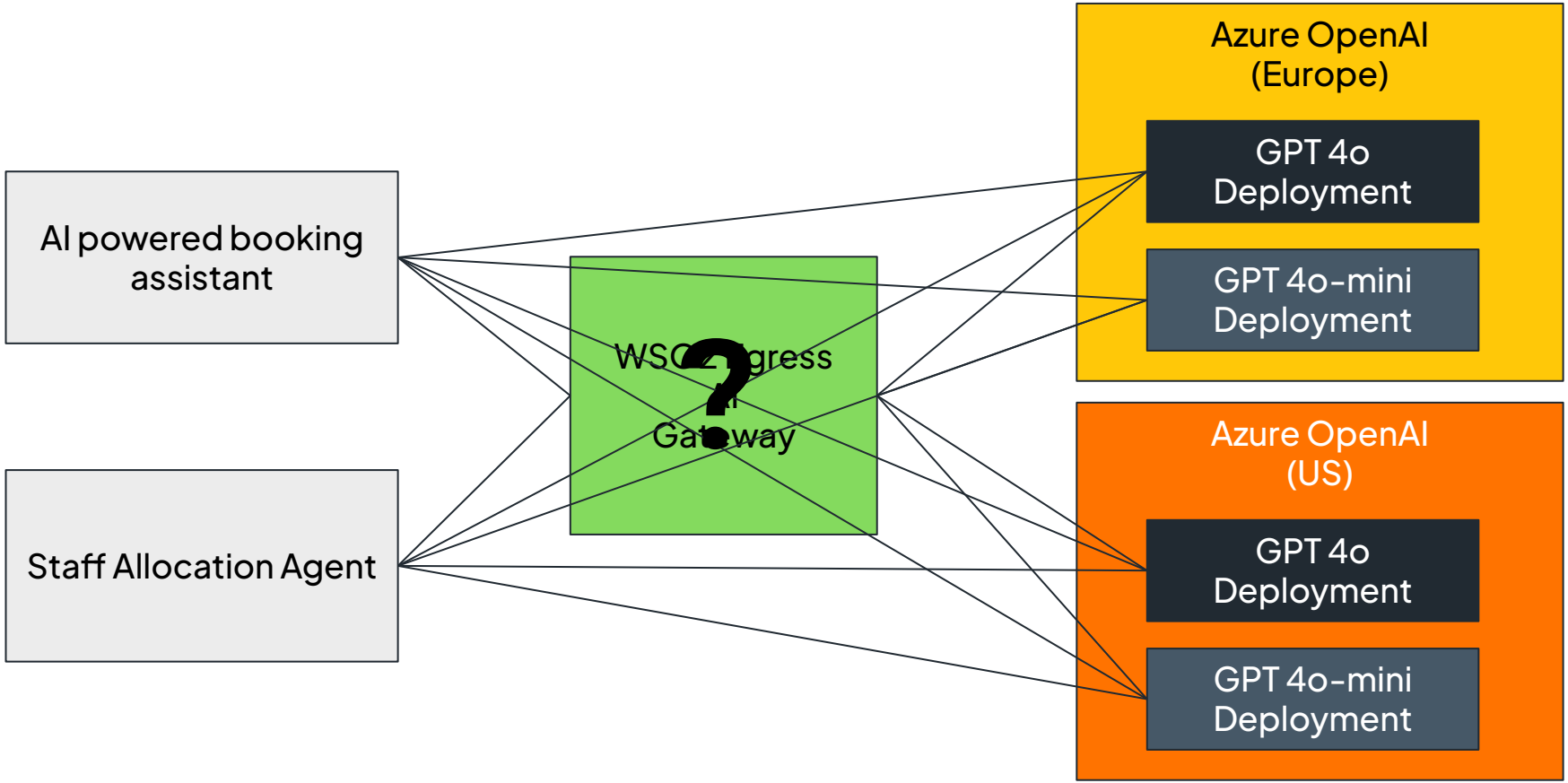
Business Benefits:

- **User Trust –** Transparent control over AI agent actions
- **Operational Efficiency** – Secure automation without security bottlenecks
- **Scalability –** Handle thousands of agents without manual overhead
- **Innovation Enablement** – Deploy new AI capabilities with confidence

Implementing Governance for Safe & Proper Use of AI

WSO2 AI Gateway


End Consumers
Ingress Gateway
Backend Services
AI
AI Gateway
Egress Gateway
External Services
ANTHROP\C
OpenAI
Organization Boundry


AI powered booking assistant
Staff Allocation Agent
?
Azure OpenAI (Europe)
GPT 4o Deployment
GPT 4o-mini Deployment
Azure OpenAI (US)
GPT 4o Deployment
GPT 4o-mini Deployment

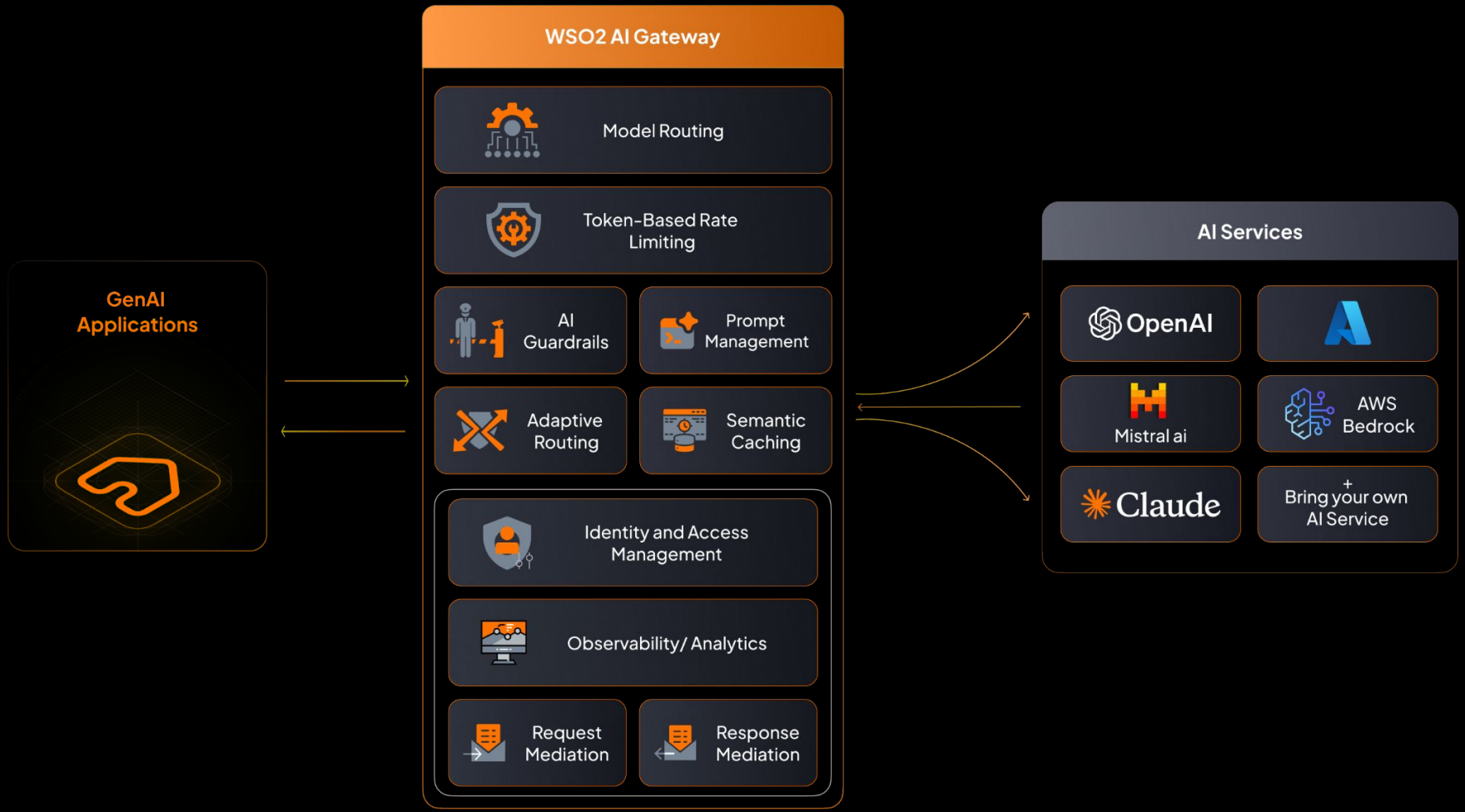
Purpose

What's happening today:

- Developers are rapidly adopting GenAI Services(OpenAI, Claude, Gemini)
- Each app ends up hardcoding its own AI integration
- Security, cost, and governance become chaotic at scale

What an AI Gateway does:

- Acts as a secure and centralized access point to AI models
- Standardizes how apps in your organization talk to AI services
- Adds control, observability, and flexibility


GenAI Applications
WSO2 AI Gateway
Model Routing
Token-Based Rate Limiting
AI Guardrails
Prompt Management
Adaptive Routing
Semantic Caching
Identity and Access Management
Observability/ Analytics
Request Mediation
Response Mediation
AI Services
OpenAI
Mistral ai
AWS Bedrock
Claude
+ Bring your own AI Service

Common usage patterns

Org-wide APIs (Admin-defined)

- Example: A company-wide summarization or classification API
- Enforced prompts, access control, rate limits
- Used by multiple teams/apps with consistency

App-specific APIs (Dev-defined)

- Example: A chatbot app using a custom prompt chain
- Allows app developers to define and use custom logic
- Still goes through the gateway for logging and safety



Performance & Resource Optimization

Token Based Rate Limiting

By enforcing rate limits, you can:

- Prevent unexpected cost spikes from excessive AI API usage.
- Optimize performance by ensuring fair resource distribution.
- Protect AI backends from overuse and service degradation.

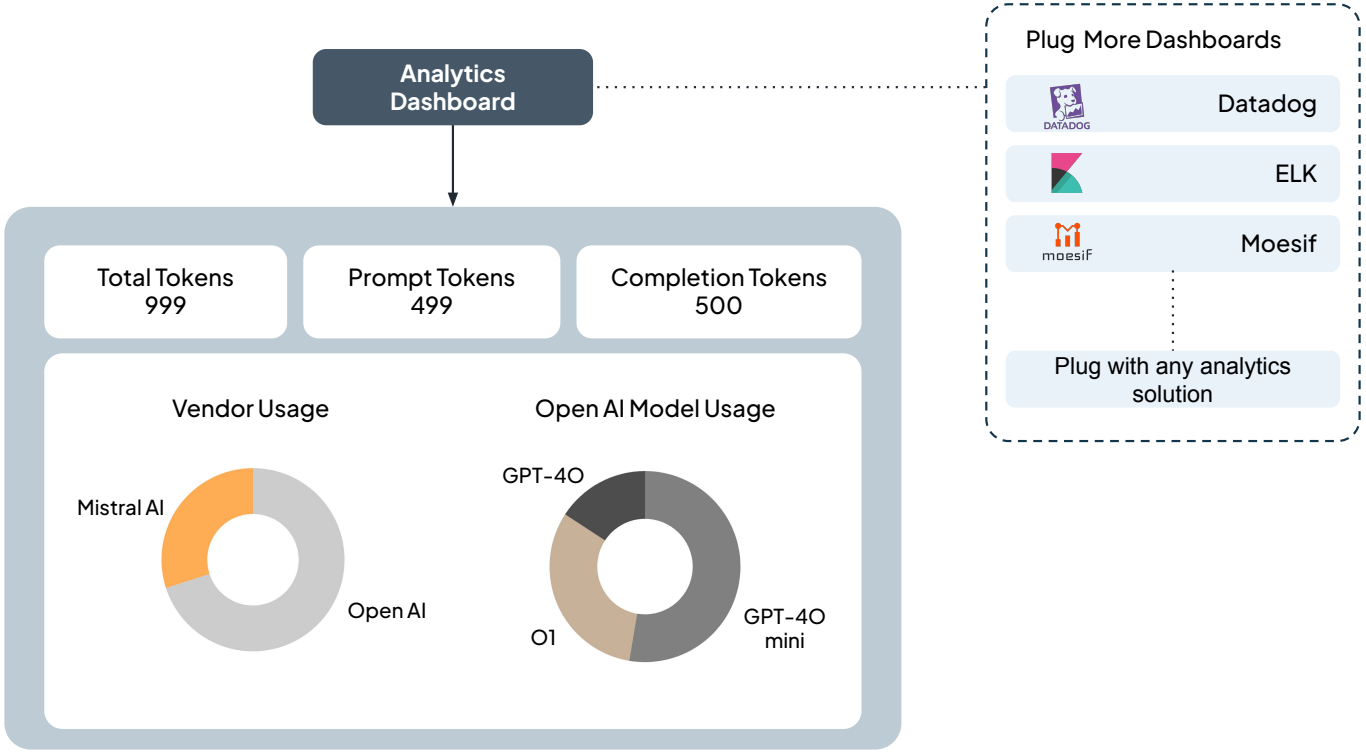
We Support:

- Real-time token consumption tracking per user/team/API key
- Configurable limits: requests/min, tokens/hour, daily quotas
- Burst allowance for peak usage scenarios

Semantic Response Cache

- **Faster responses** – Cached results return in milliseconds vs. seconds for model inference
- **Cost reduction** – Eliminates redundant token usage for similar queries
- **Improved user experience** – Consistent responses for equivalent questions across different phrasings

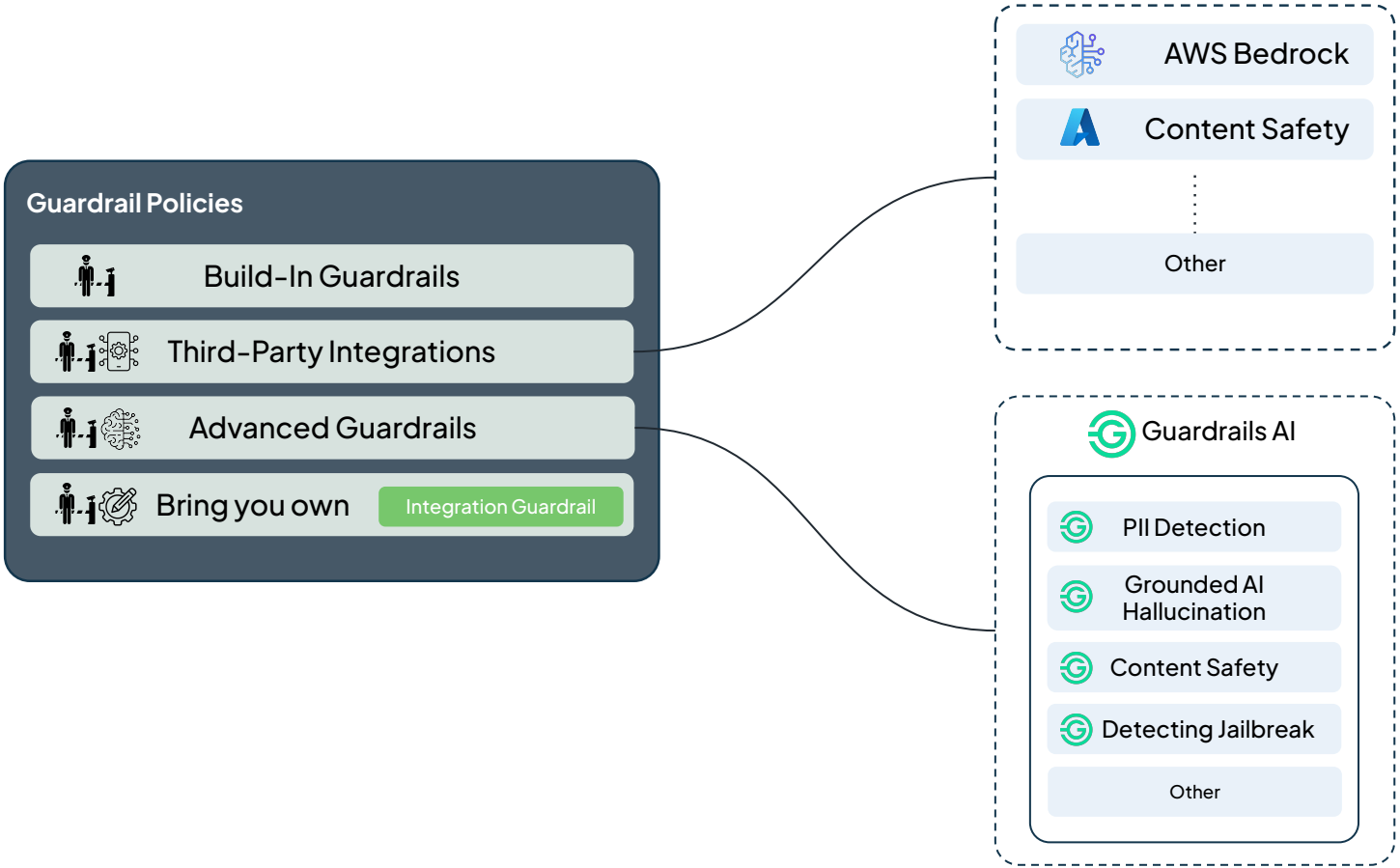
AI Gateway Analytics





AI Guardrails

Guardrail	Sample Use Case
Semantic Prompt Guard	Stops prompts like “Write my homework” in a student assistant app.
Regex based PII Masking	Masks credit card numbers in user prompts for a payment chatbot.
Word Count and Sentence Count	Limits AI replies to 50 words in a quick-answer mobile app
JSON Schema Validation	Validates API responses for correct format in an e-commerce platform
Regex Validation	Verifies user-entered email addresses in a registration form
URL Validation	Ensures links in AI responses resolve via DNS for a news aggregator app.
Content Length	Caps user inputs at 500 characters in a chat AI to prevent spam
Grounded AI Hallucination	Prevents AI from making up facts in product descriptions.
Content Safety	Filters hate speech in comments generated by an AI writing assistant.
PII Detection and Masking	Detects and hides phone numbers in support chatbot inputs.
Jailbreak detection	Stops prompts like “Ignore all rules” in customer service bots.


Guardrail Policies
Build-In Guardrails
Third-Party Integrations
Advanced Guardrails
Bring you own
Integration Guardrail
AWS Bedrock
Content Safety
Other
Guardrails AI
PII Detection
Grounded AI Hallucination
Content Safety
Detecting Jailbreak
Other



Adaptive Routing

Adaptive Routing

Policy	Sample Use Case
Model Round Robin Policy	Evenly distributes API requests across three AI models to balance load in a chatbot platform
Model Weighted Round Robin Policy	Routes 70% of requests to a high-capacity AI model and 30% to a smaller model for cost efficiency
Model Failover Policy	Switches to a backup AI model when the primary model hits a 100 requests/second rate limit in a real-time translation app

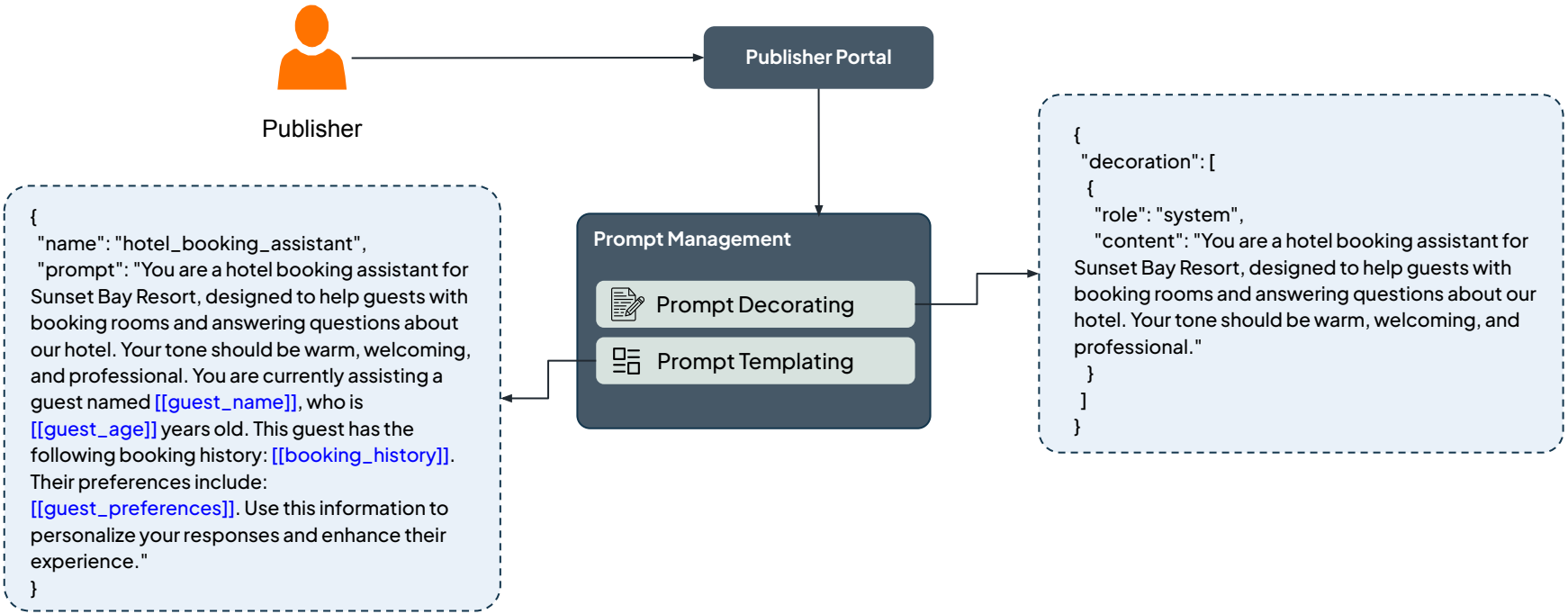


Prompt Management

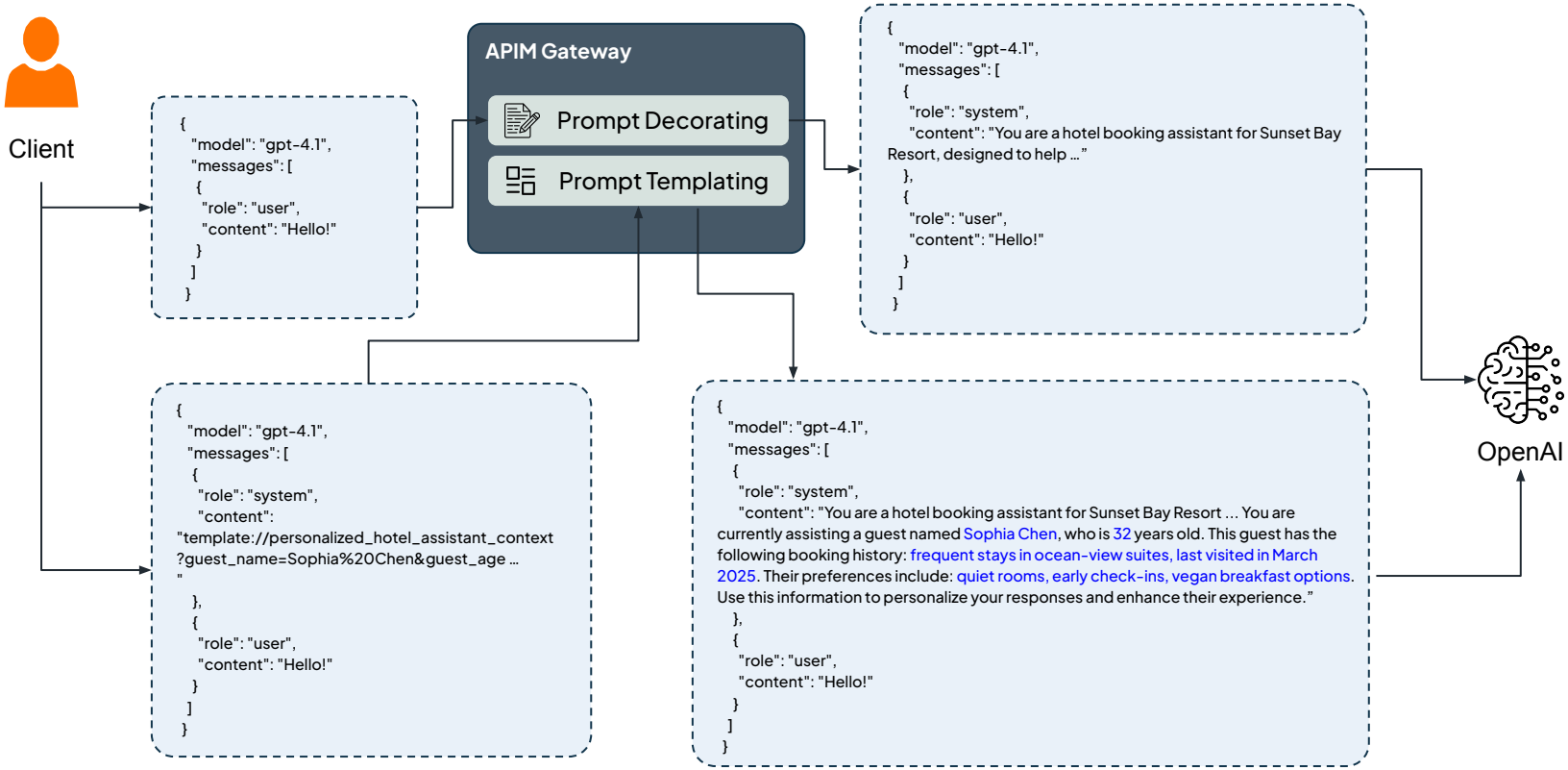
Prompt Management

Feature	Sample Use Case
Prompt Templating	Uses a standard template to insert user queries into a chatbot, ensuring consistent AI responses
Prompt Decorating	Adds role instructions (e.g., “Act as a teacher”) to user prompts for tailored AI tutoring output

Publisher: Manage Prompt Decorations and Templates

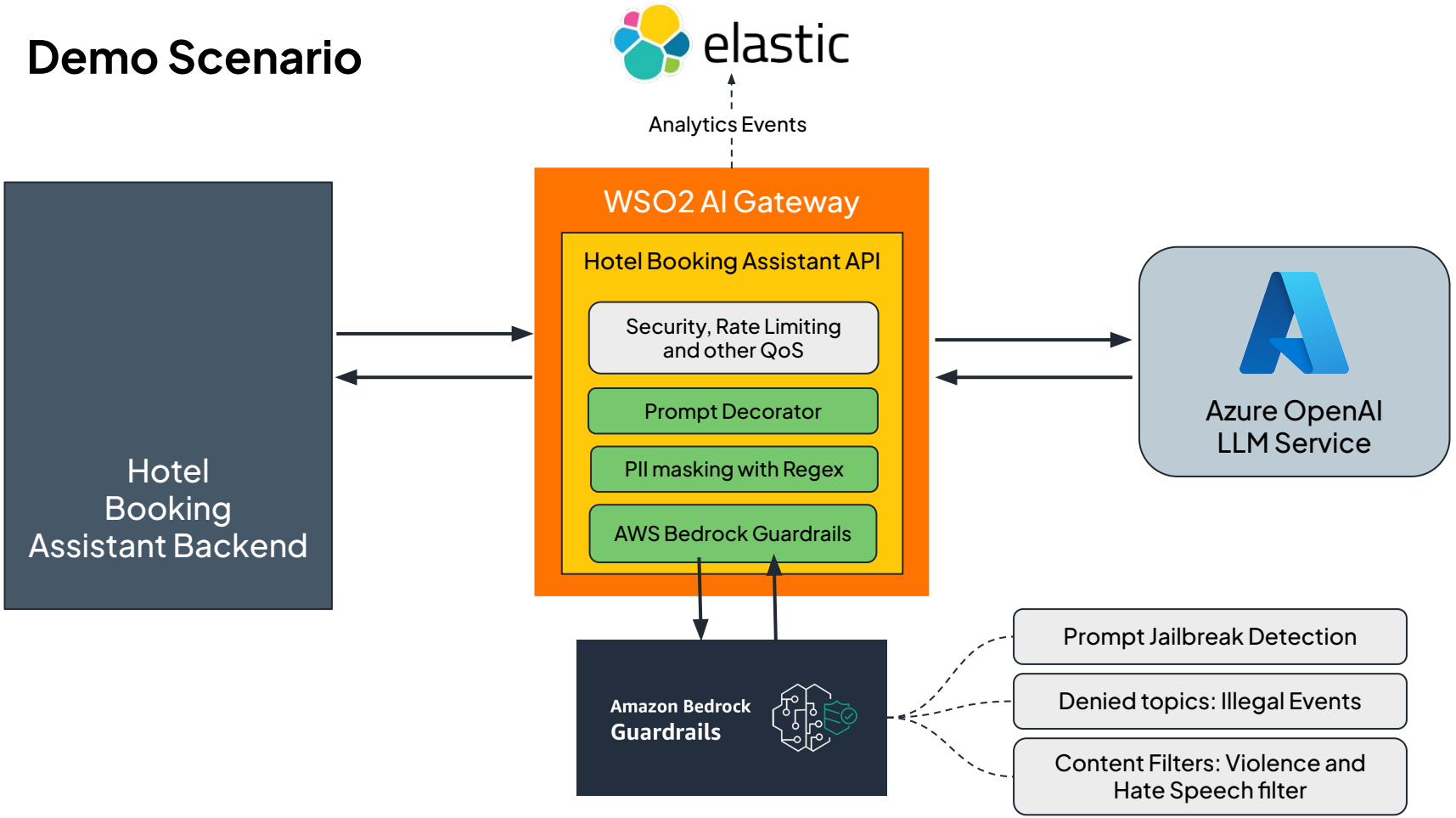


Client: Consume AI APIs

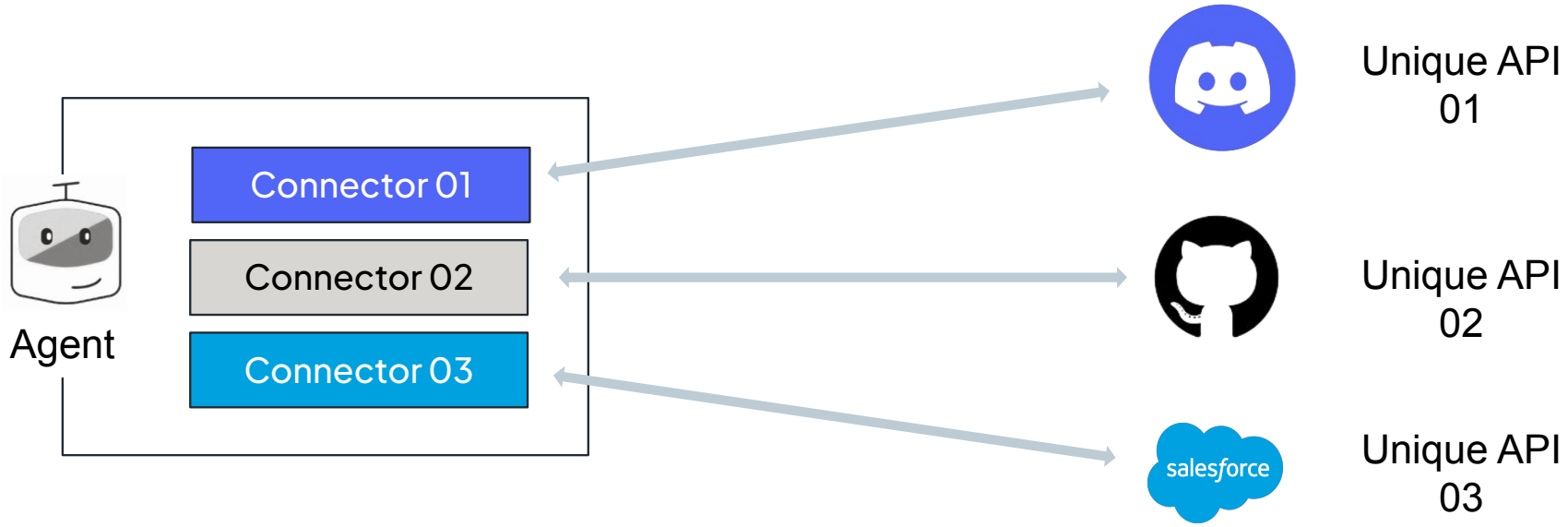


Demo #2

Demo Scenario

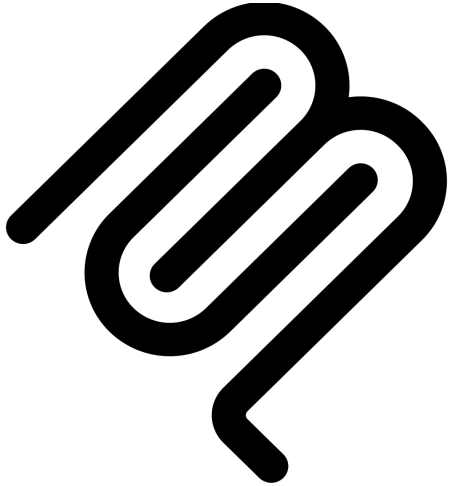


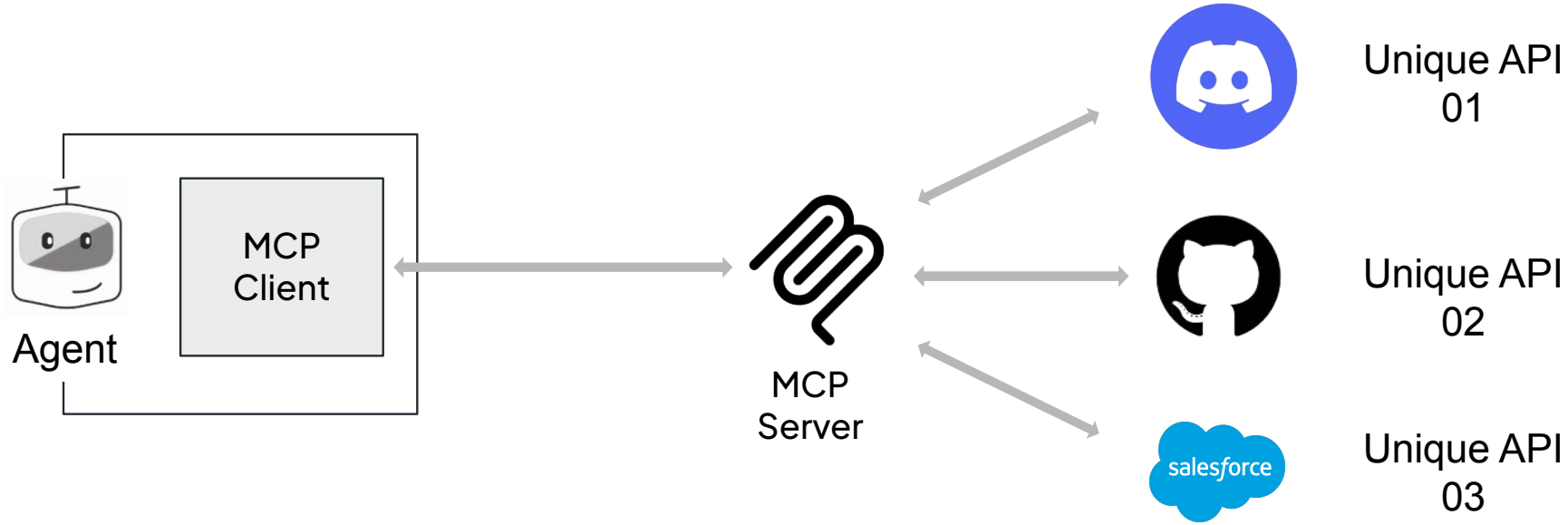
Making Business Resources AI Consumable


Agent
Connector 01
Connector 02
Connector 03
Unique API 01
Unique API 02
Unique API 03
salesforce

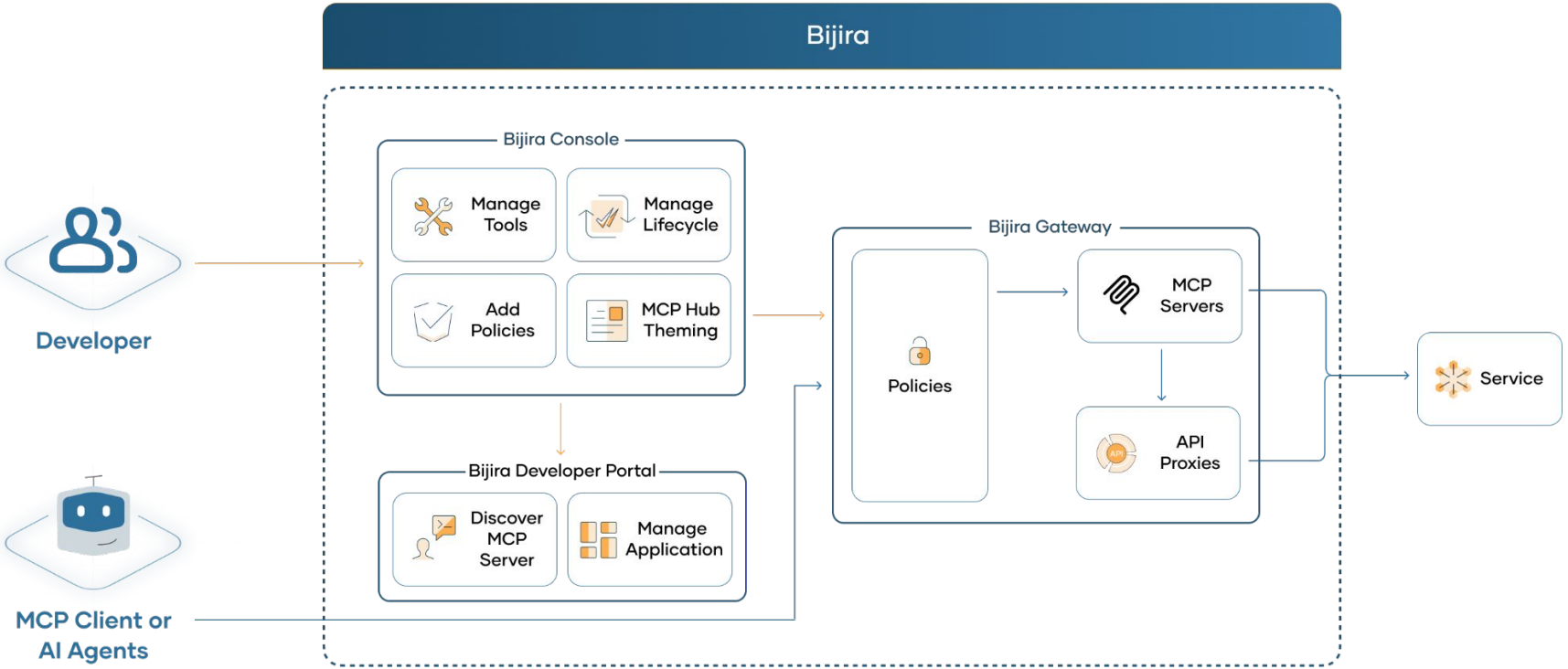
Bridging the AI-Enterprise Gap: Enter the Model Context Protocol

- **Discoverability:** AI systems can automatically learn tools, reducing manual configuration.
- **Flexibility:** Organizations can add MCP server capabilities without updating AI clients, enabling rapid iteration.
- **Consistency:** A single interface standard reduces complexity across AI implementations.



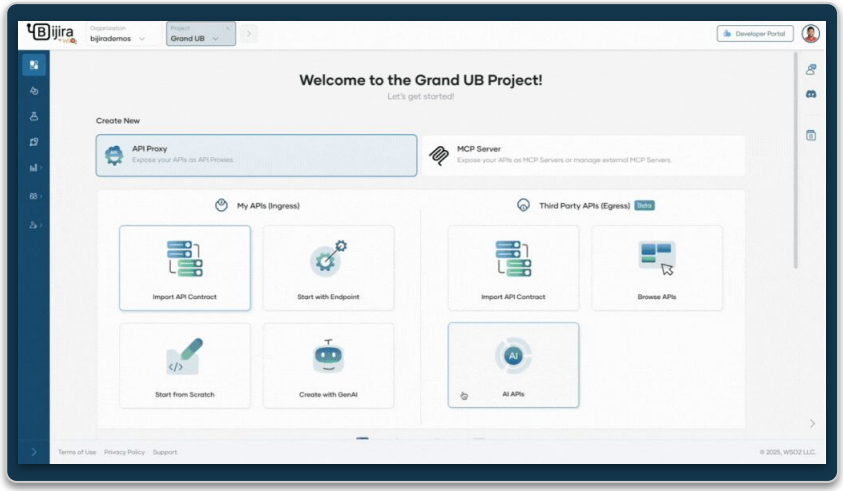

MCP
Client
Agent
MCP
Server
Unique API
01
Unique API
02
Unique API
03
salesforce

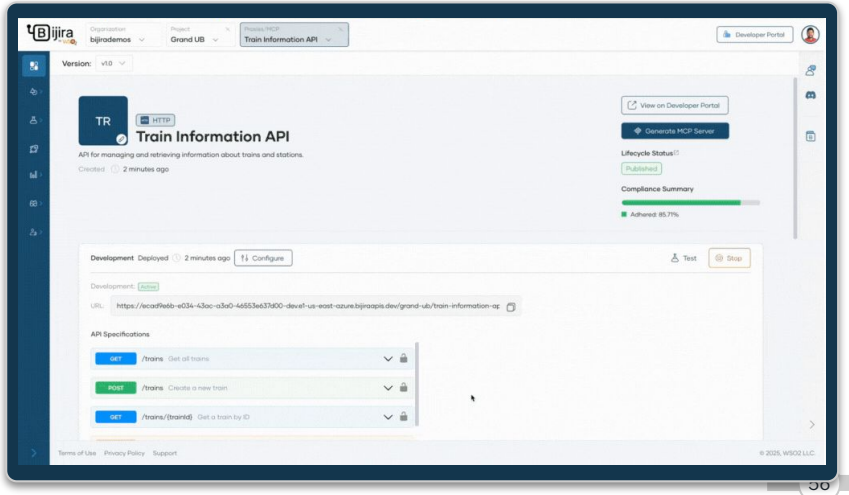
Connecting APIs to AI Agents: MCP Server Management



Expose APIs as MCP Server Tools

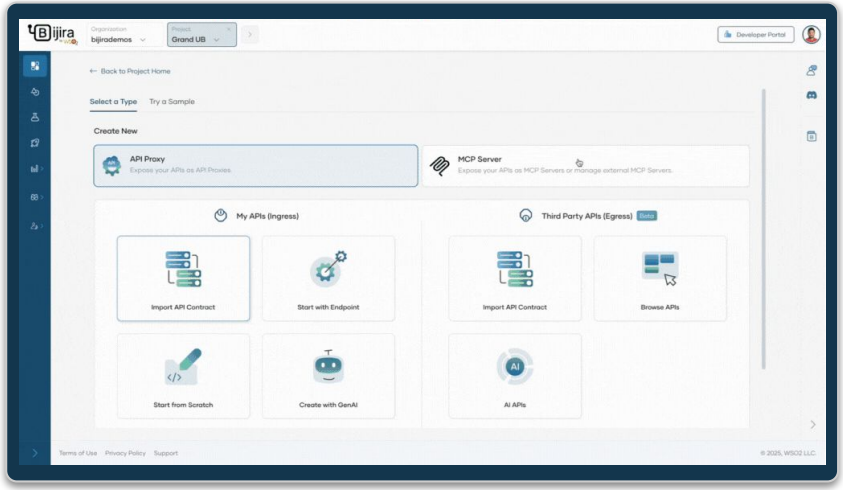
- Transform APIs or API Proxies as Tools
  - Turn existing APIs into reusable, callable functions for AI agents.
- Add Structured Metadata
  - Include descriptions and other metadata
- Apply Policies
  - Enforce security, rate limiting, and other governance policies on tools.
- Publish MCP Servers to the MCP Hub
  - Make your APIs discoverable as tools for AI agent developers.





Govern Access to Remote MCP Servers

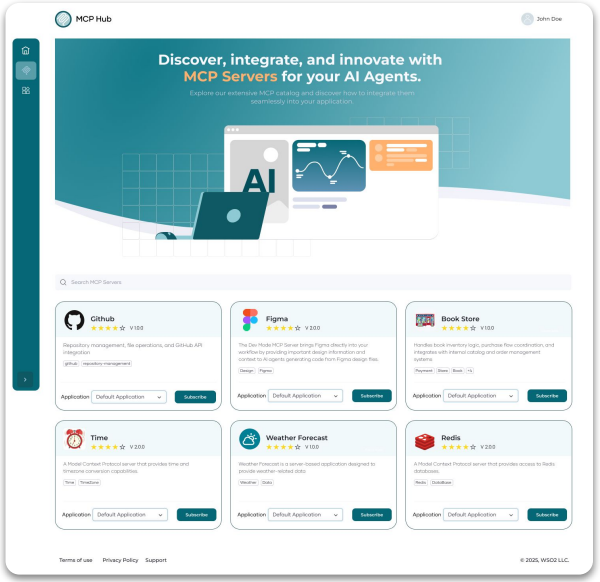
- Bring External Tools to Your Control Plane
  - Connect AI-ready tools hosted outside your environment.
- Register with Metadata
  - Provide tool descriptions, endpoints, and capabilities to enable AI agent usage.
- Policy Enforcement
  - Apply centralized security, rate limiting, and access controls
- Discover via MCP Hub
  - Expose remote MCP servers through the hub for AI agent developers

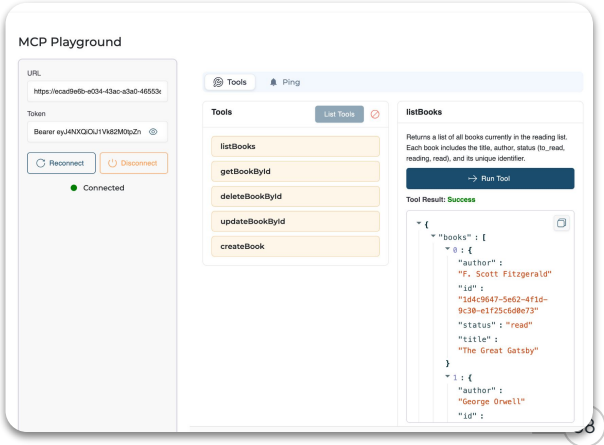


MCP Hub — A Dedicated Discovery Portal for AI Agents

A curated and clutter-free marketplace within Bijira's developer portal, purpose-built for discovering MCP Servers.

- **Focused Experience:** Only MCP tools, no REST/GraphQL noise
- **Agent-Friendly Docs:** Clear capabilities, parameters & usage
- **Custom Theming:** Tailor the look & feel of your hub
- **Built-In Playground:** Try out MCP server functionality instantly





Demo #3

Question Time!




Thank you!
20TH ANNIVERSARY EDITION
WSO2CONASIA
PLATFORMLESS MODERNIZATION