Hitachi Digital Services

RAI Marketplace Listing

GenAI offering

Hitachi Team
March 2024

Infuse Responsible AI into your GenAI project

Hitachi Digital wants to empower your organization with our ML development framework and internally developed **Responsible AI** toolkit.

Bring GenAI solutions to your users and customers but also ensure that **safety** systems and **guardrails** are in place to **protect** both your organization, employees and customers.

Start with a **4-hour workshop,** our expert-led AI/ML team will meet with you and your IT team to introduce the core concepts of a Responsible AI, **MDLC** and **LLMOps.** The Hitachi team will then take your business problem/opportunity and in less than **6 weeks** integrate our Responsible AI toolkit into your system.

Why Responsible AI?

- Responsible AI is the practice of developing and deploying AI systems in a way that is fair, accountable, transparent, and ethical.

- **GenAI** is extremely powerful but that power needs to be tempered and controlled so that adverse impacts are eliminated or minimized. **Responsible AI** is a key control mechanism that imposes those guardrails.

- Responsible AI is a capability that lets organizations deliver GenAI powered applications in a **reliable** user-centric or customer-centric fashion.

Outcomes

- Integrate Hitachi Digital's Responsible AI solution into your project. Project can be greenfield or brownfield

- LLMOps Dashboards for monitoring your GenAI applications for signals such as toxicity, cost, relevance, etc.

Gen AI Challenges

Gen AI while getting to become widely adopted has its own challenges not limited to the below:

Lack of Transparency : The AI model being non-deterministic is not always reliable, so the response may be difficult to understand and may contain errors.

Lack of Vendor accountability: Despite lack of transparency, the practitioner who is using the AI model is responsible for any damages done.

Privacy Violations: Unregulated data collection and use by AI systems can violate individual privacy rights, leading to legal concerns related to data protection and consent.

Safety and Reliability issues: Prompts and Responses can be Injection or Jailbreak attempts to take hostage of the AI system.

Sensitivity of the Responses : Prompts and Responses can be Toxic or Insensitive targeting a specific ethnicity or group.

Building Fairness into AI: Biases in data, algorithms, responses from AI systems can lead to unfair results.

Use of responsible AI is not limited to single entity but rather **spread across various stakeholders**(organizations, developers, investors, regulators, end users and consumers).

Responsible AI (RAI) is the area that is being dedicated by AI vendors to address the above issues. Practitioners from governments, businesses, universities, and other groups are working together to make sure AI is used safely and fairly. The responsible AI development is imperative in mitigating various legal, ethical, and social risks that could otherwise undermine trust and confidence in AI technologies.

What is Hitachi Digital's AI Compass ?

Hitachi AI Compass is our Responsible AI solution, revolutionizes AI solutions by ensuring safety, trustworthiness, and ethical integrity through its comprehensive suite of microservices, safeguarding against toxicity, bias, and other ethical pitfalls, thus guaranteeing responsible AI implementation for every customer.

Hitachi AI Compass provides a vital collection of microservices acting as a gateway to FM/LLM models, offering essential capabilities to detect and score toxicity, sentiment, refusal, jailbreak, bias, and prompt injection. Unlike industry standard LLMs with limited RAI controls, H-AI Compass is highly flexible, policy-driven, and customizable, becoming indispensable especially for customers utilizing open-source LLMs.

Key Customer Benefits

- Hitachi AI Compass provides tailored Responsible AI controls, allowing flexibility to adjust policy scores according to regional, cultural, and organizational requirements.
- By proactively filtering out invalid calls, Hitachi AI Compass optimizes cost efficiency associated with expensive LLM calls.
- Unlike open-source LLMs, Hitachi AI Compass ensures robust Responsible AI guardrails, enhancing trust and reliability in AI solutions

E3 Model Development Life Cycle (MDLC) Methodology

Hitachi Digital Services

ENVISION

Understand the path and your destination via value-based assessment of your business portfolio for effective AI/ML enablement

EVALUATE

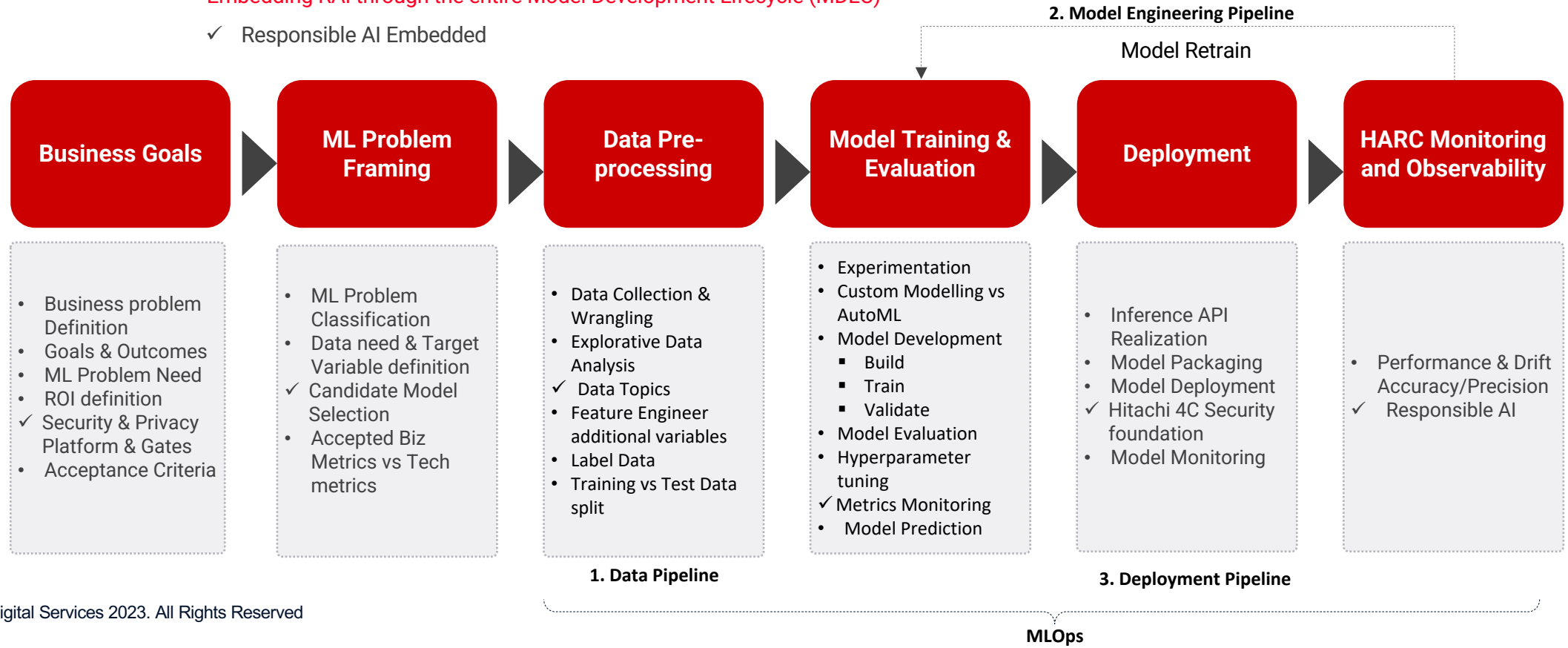
Measure how fast you can navigate by understanding the breadth of change. This starts with effective ML problem framing and proof-of-value elicitation through a quick Pilot

EXECUTE

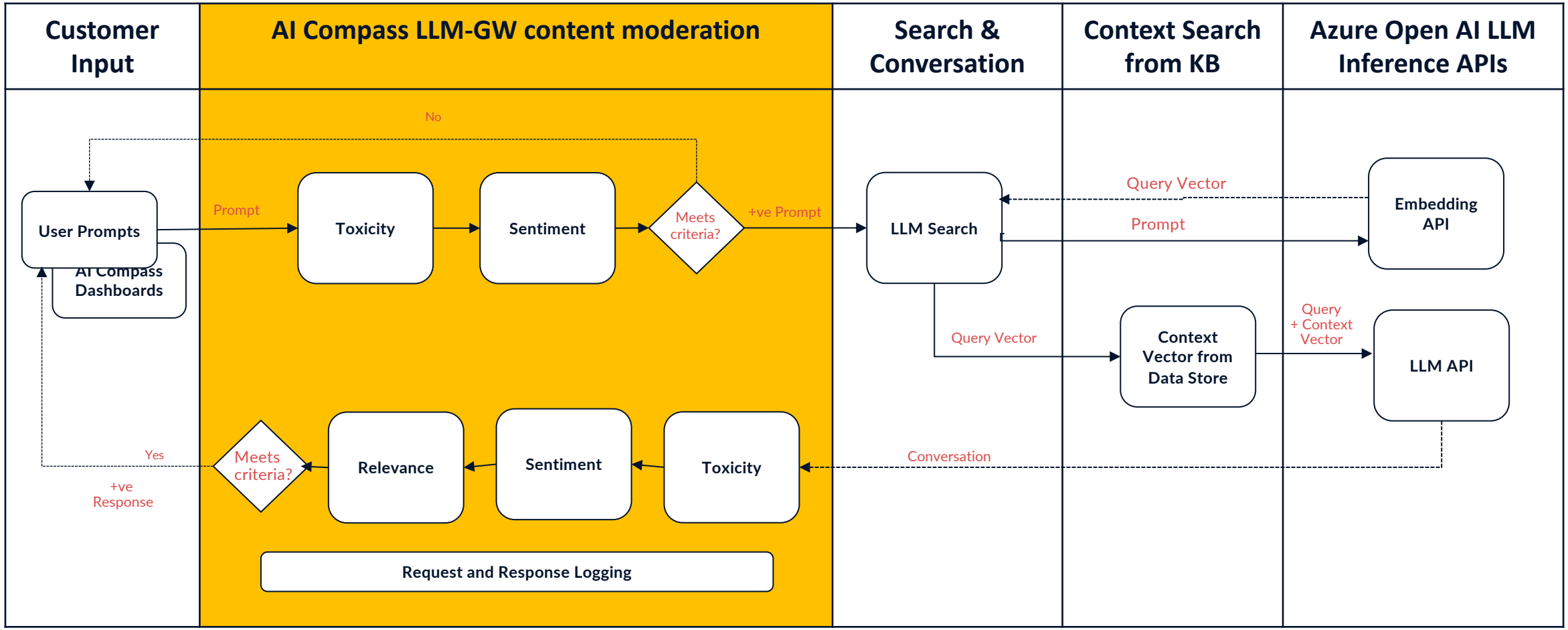
Optimization & enhancement of KPIs through automation of model training, evaluation, and secure deployment, all underpinned by continuous performance monitoring & Responsible AI compliance

Embedding RAI through the entire Model Development Lifecycle (MDLC)

✓ Responsible AI Embedded



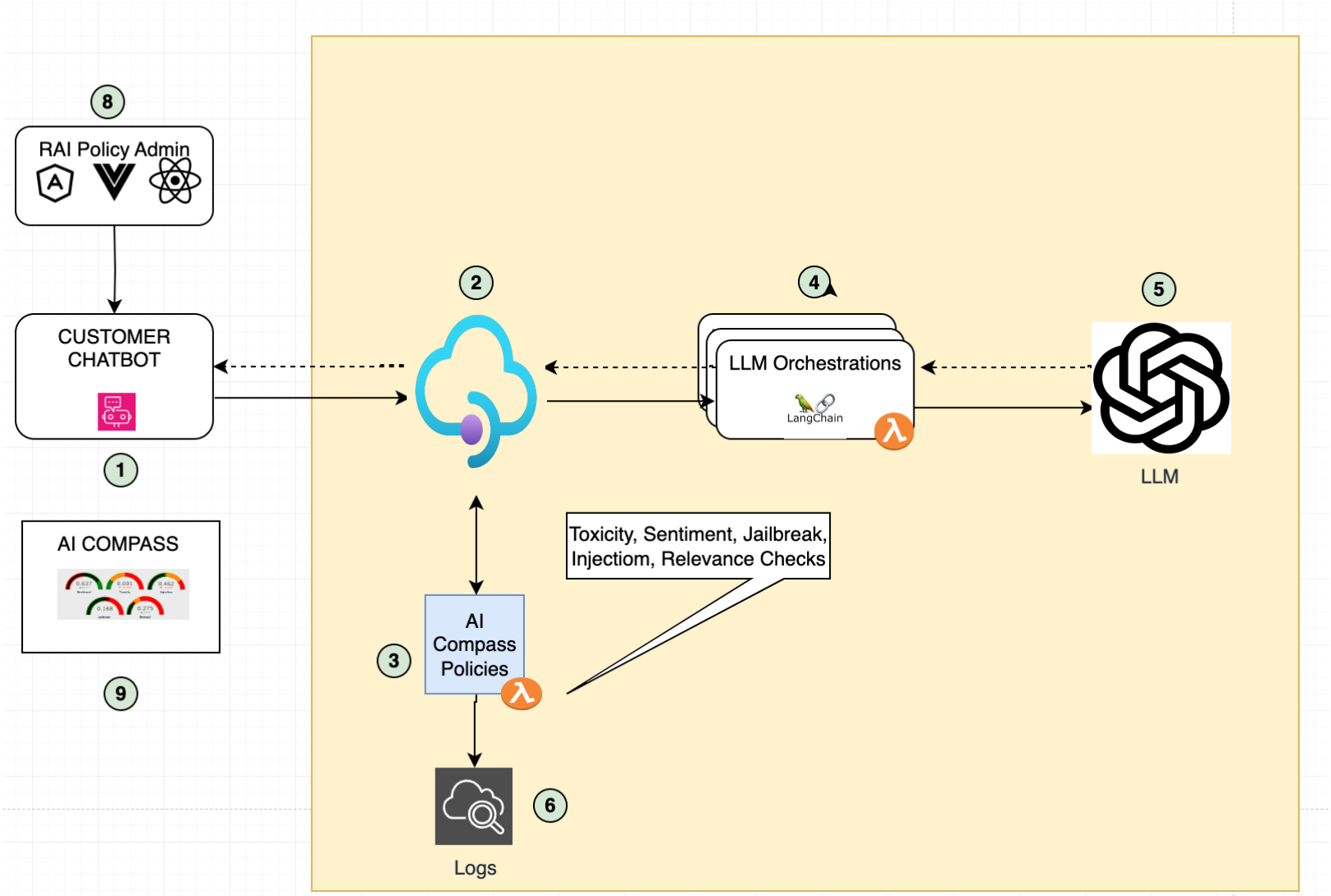
AI Compass Content Moderation



Solution Options:

1. AI Compass LLM-GW as a Mediator
2. AI Compass LLM-GW an extension policy to API GW

AI Compass – High Level Architecture

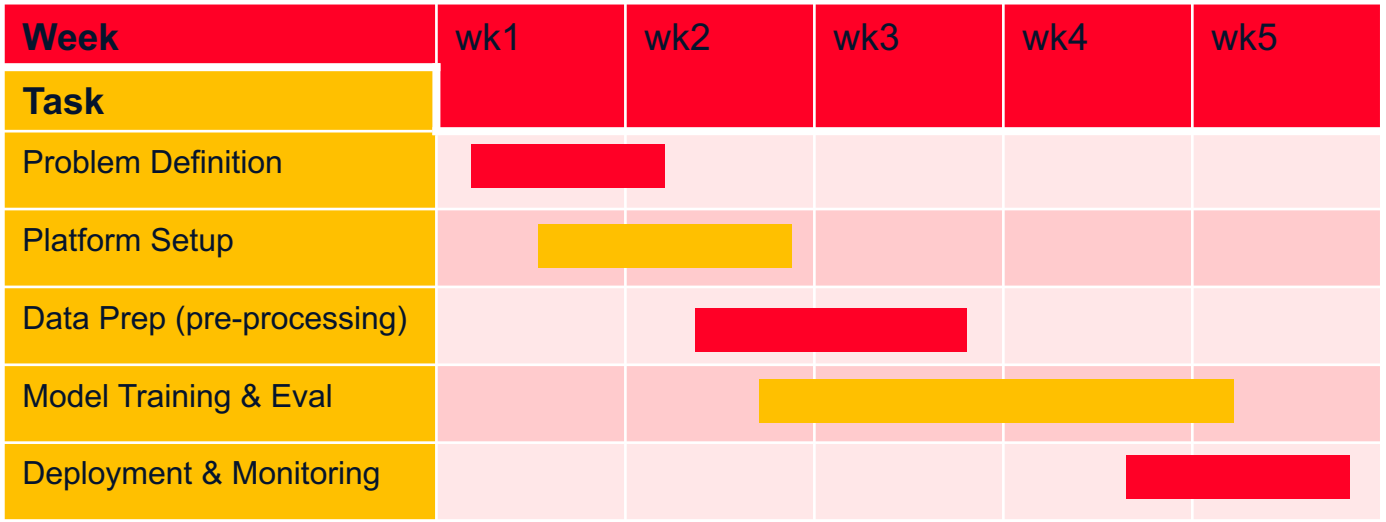


AI Compass is a collection of microservices with 3rd party OSS S/W like NLTK and Hugging Face libraries which provide RAI capabilities.

Annotated Flow:

1. User sends NL query to LLM App
2. SPA connects to a published API protected by API gateway (Azure APIM)
3. AI Compass LLM Gateway and safety/policy are applied as extension to API GW Policy. Request is then forwarded to LLM Orchestrator
4. LLM Orchestrator orchestrates the calls for Query and Completion context for RAG
5. LLM takes as input sanitized query and returns a response
6. Response is then processed by API Gateway based on RAI policies
7. All requests and responses will be logged to CloudWatch
8. AI Compass Dashboard allows AI Ops team to monitor LLMOps metrics
9. AI Compass provides dashboards to view the displays

Responsible AI in less than 6 weeks



Deliverables

1. NFR Requirements including Security, Perf and Compliance Needs
2. Acceptance criteria from RAI perspective
3. Sample Q&A Prompts and Responses
4. RAI Acceptable Policies tuned for Penske
5. Updated Platform for RAI and Security Architecture
6. Data Bias Detected for fine-tune modelling problems
7. AI Compass LLMGW platform deployment
8. RAI System Testing
9. AI Compass Logs
10. AI Compass Dashboards

Hitachi's AI Compass Dashboard