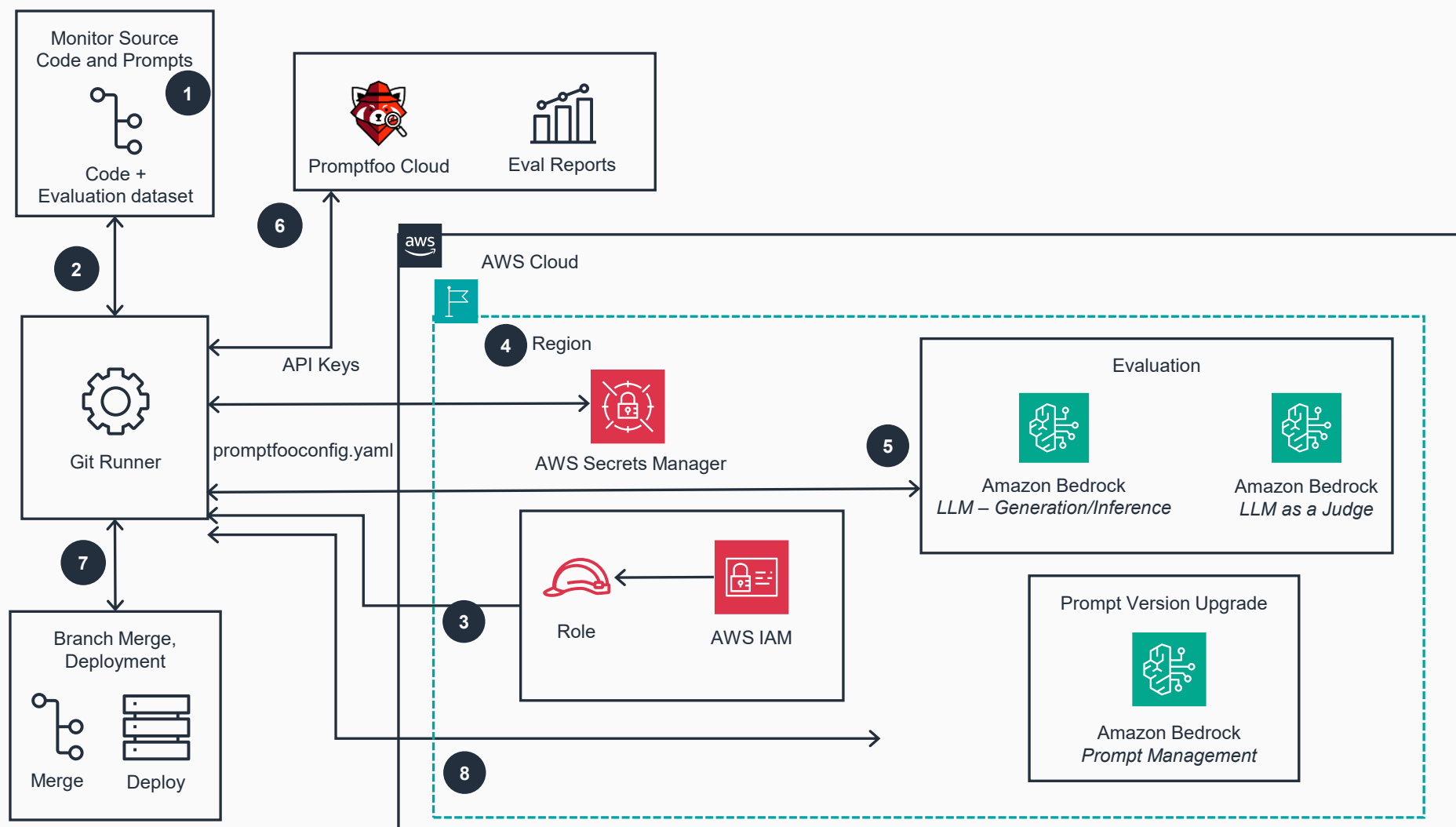


Guidance for Evaluating Generative AI Applications using OSS on AWS

Test Driven Development – Promptfoo Evaluation | Cloud

This architecture diagram illustrates a CI/CD pipeline where source code changes trigger automated prompt evaluations using Promptfoo Cloud and Amazon Bedrock Foundation Models, leading to automated deployment of approved prompts and code. This slide shows steps 1-7.



1. Maintain prompts and code in source repository. GitHub Actions workflow automatically detects changes and pull requests made by developers and triggers the evaluation pipeline.
2. During the next stage in the pipeline, Prompt Evaluation, the Git Runner process establishes an authorized connection with Promptfoo and executes the evaluation API with needed test cases.
3. **AWS Identity and Access Management (AWS IAM)** provides role-based access control as the authorization mechanism for all stages within the pipeline, ensuring secure execution of each step.
4. Git Runner retrieves the Promptfoo API key from **AWS Secrets Manager**.
5. Promptfoo uses a YAML configuration file to define evaluation criteria, test cases, datasets, and custom scoring functions. The open source CLI library orchestrates prompt evaluation based on this configuration and generates results. Tests can run using foundation models available through **Amazon Bedrock** to evaluate multiple prompts across multiple models, with Bedrock models also serving as judges to create evaluation scores.
6. Optionally, the promptfoo library can share the evaluation results with the Promptfoo Cloud offering for dashboard visualization.
7. Review the evaluation report and approve the merging of the pull request to continue the CI/CD pipeline.



Guidance for Evaluating Generative AI Applications using OSS on AWS

Test Driven Development – Promptfoo Evaluation | Cloud

This architecture diagram illustrates a CI/CD pipeline where source code changes trigger automated prompt evaluations using Promptfoo Cloud and Amazon Bedrock Foundation Models, leading to automated deployment of approved prompts and code. This slide shows steps 1-7.

8

The Deploy workflow updates the version of the prompts, managed in **Amazon Bedrock** prompt management, that passed the evaluation available for next deployments.

