TrustyAI community meeting

March 2025

March 2025 updates

- TrustyAI core / service
  - Current: 0.27.0 release
    - https://github.com/trustyai-explainability/trustyai-explainability/releases/tag/v0.27.0
    - quay.io/trustyai/trustyai-service:v0.27.0
- TrustyAI operator
  - Current: 1.34.0 release
    - https://github.com/trustyai-explainability/trustyai-service-operator/releases/tag/v1.34.0
  - quay.io/trustyai/trustyai-service-operator:v1.34.0

What’s new?

TrustyAI - What's new?

- **TrustyAI core/service 0.25.0**
  - Add Trivy image scanning by @ruivieira in #654
  - Clean up spurious merge conflict strings in javadoc by @donaldh in #673
  - Add multi-input support to kserve inference payload parsing by @RobGeada in #661
  - Change fail criteria for security scan by @ruivieira in #674
  - CVE-2025-24970: Override to io.netty:netty-handler@4.1.108.Final by @ruivieira in #672
  - CI
    - Update operators to latest versions by @RobGeada in #662
    - Downgrade ODH to 2.22 by @RobGeada in #664
    - Increase SHAP CF Generator test timeouts by @RobGeada in #668
    - Migrate to Python cluster setup scripts by @RobGeada in #667
    - Update owners and PR template by @RobGeada in #670
    - Clean up artifacts directory by @RobGeada in #669

TrustyAI - What's new?

- **TrustyAI operator 1.32.0 ... 1.34.0**
  - docs: Add CONTRIBUTING.md by @ruivieira in #423
  - fix(RHOAIENG-6720): Remove Kustomize v5 deprecation warnings by @ruivieira in #425
  - feat: Add Trivy scans to images by @ruivieira in #346
  - chore(RHOAIENG-15495): Update main component metadata with correct branches by @ruivieira in #429
  - feat(RHOAIENG-20330): Update GuardrailsOrchestrator unit tests by @artemsa223 in #426
  - fix: Use final Kustomize deployment name for operator image by @ruivieira in #430
  - TrustyAI service
    - Remove generic deployment watcher by @ruivieira in #400
  - LMEval
    - Switch LMES driver to a container port by @yhwang in #409
    - Add support for S3 offline assets by @ruivieira in #399
    - Replace obsolete UNIX_ARTIFACTORIES env var in LMEval by @ruivieira in #415
    - Support custom template and prompt by @yhwang in #404
  - Guardrails
    - Add guardrails orchestrator controller by @christinaexyou in #403
    - Add correct orchestrator container manifest name by @ruivieira in #410
    - Fix orchestrator container manifest by @ruivieira in #411
    - Add guardrails to components by @christinaexyou in #413
    - Get orchestrator image from operator's ConfigMap by @ruivieira in #417
    - Use correct ConfigMap for orchestrator image config by @ruivieira in #420
  - CI
    - Migrate to Python tests, further unify testing between service+operator by @RobGeada in #406
  - Other fixes
    - CVE-2024-28110: Fix cloudevents-sdk target version by @ruivieira in #414
    - Typo in guardrailsorchestrator_controller.go by @RobGeada in #419
    - Clean up module dependencies by @ruivieira in #421

Current work

TrustyAI - Service

- Service Python rewrite
  - https://github.com/trustyai-explainability/trustyai-service

Llama Stack

- Eval
  - Leveraging LMEval / lm-evaluation-harness
- Safety
  - Leveraging Guardrails

Roadmap

TrustyAI 2025 roadmap

Legend
Not started
In progress
Completed

- Llama Stack
  - Eval
  - Guardrails
- KServe explainer integration
- Detoxification fine-tuning
- Saliency Explainers
- Guardrails
  - Orchestrator
- LM-Eval
  - LM-Eval v2 iteration on upstream roadmap (target 6th December)
    - https://github.com/trustyai-explainability/trustyai-service-operator/issues/366

Other topics