

AI Explainability WG

November 2024

November 2024 updates

- TrustyAI core / service
 - Current: 0.23.0 release
 - <https://github.com/trustyai-explainability/trustyai-explainability/releases/tag/v0.23.0>
 - quay.io/trustyai/trustyai-service:v0.23.0
- TrustyAI operator
 - Current: 1.29.0 release
 - <https://github.com/trustyai-explainability/trustyai-service-operator/releases/tag/v1.29.0>
 - quay.io/trustyai/trustyai-service-operator:v1.29.0

What's new?

TrustyAI - What's new?

- **TrustyAI core/service 0.23.0**
 - Change KServe cloudevent responses to byte (#657)
 - Override Quarkus JUnit XStream to 1.4.21 (#655)
 - **CI/testing**
 - Add download check for all manifests
 - Install kserve prereqs, clean up logging

TrustyAI - What's new?

- [TrustyAI operator 1.29.0](#)
 - Revert metrics service target port to https (#355)
 - Remove TAS-only patch from rhoai overlay (#352)
 - **LMEval**
 - Merge LM-Eval dev branch (#337)
 - Add overlay for Kueue support (#364)
 - Add Watsonx LLM support to LMEval Job image (#360)
 - Send Job logs to Pod StdOu (#358)
 - Add offline support for LMEvalJobs (#351)
 - Change batchSize to string in order to support "auto" and "auto:N" (# 350)
 - Add PVC management
 - Checkout current LMEH release from midstream (#341)
 - Support more customization for the job pod of the LMES (#340)
 - Update GH Action build and push comment, add LMH images to manifests (#338)
 - Add initial Kueue integration (#347)
 - **TrustyAI service**
 - Re-enable KServe Raw support (#361)
 - Ensure proper finalizer order and KServe logger removal (#357)
 - **CI/testing**
 - Allow smoke test to pass (#356)
 - Sync logging + makefile with service (#329)

Current work

TrustyAI - LM-Eval

- Available on ODH 2.21
 - Documentation available at <https://trustyai-explainability.github.io/trustyai-site/main/lm-eval-tutorial.html>
- Kueue support as an overlay
 - This requires Kueue installed and manual changes to Kueue's configuration
 - Aim at removing these limitations in ODH 2.22 (6th December)

TrustyAI - Guardrails

- Work ongoing in
 - Orchestrator support KServe serverless
 - TLS configuration for Serverless
 - gRPC service addressing

Roadmap

TrustyAI 2024 roadmap

- KServe explainer integration
- Detoxification fine-tuning
- Saliency Explainers
- Guardrails
- LM-Eval

- LM-Eval v2 iteration on upstream roadmap (target 6th December)

- <https://github.com/trustyai-explainability/trustyai-service-operator/issues/366>

Legend

Not started

In progress

Completed

Other topics