

AI Explainability WG

January 2024

January 2025 updates

- No releases in December 2024
- Release 1.31/0.25 planned for 17th January 2025

What's new?

TrustyAI - What's new?

- **TrustyAI operator 1.31.0**

- **LMEval**

- Use only local unitxt catalogs when in offline mode (#386)
 - Allow unitxt code execution when LMEval code execution enabled (#385)
 - Disable telemetry (#383)
 - Remove unneeded metrics fetch
 - Include HF metrics for offline

- **Security**

- Update golang.org/x/net to 0.33.0

Current work

TrustyAI

- Work in progress
 - Merge Guardrails Orchestrator for ODH 2.24
 - Explainers

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