



# Infosys Responsible AI Toolkit – Telemetry API Usage Instructions

### Contents

About Telemetry Layer	1
Dependencies	2
· Felemetry APIs	
Endpoints usage flow	



# **About Telemetry**

This repository implements a comprehensive telemetry collection and monitoring system designed to provide observability across the entire Responsible AI ecosystem, capturing metrics from moderation APIs, content filtering, bias detection, explainability engines, performance monitoring, and compliance validation systems. The system tracks detailed telemetry from LLM interactions including input prompts, generated responses, safety checks for privacy violations, security risks, hallucination, and biases, while also monitoring explainability features such as Thread of Thought, Chain of Thought, and Chain of Verification. Built with FastAPI and Elasticsearch in a multi-tenant architecture, it provides real-time dashboards, compliance reporting, and performance analytics to ensure comprehensive oversight and continuous improvement of Responsible AI deployments.

Once API swagger page is populated as per instructions given in the github repository Readme file, click on 'try it out' to use required endpoints. Details of endpoints associated with moderation layer are outlined below.

# **Dependencies**

All telemetry endpoints in this repository depend on upstream AI services (such as Moderation, Privacy, Profanity, and Explainability modules) to provide the data they collect. The system requires a properly configured Elasticsearch cluster for data storage and analytics, and Kibana for visualization and monitoring. Telemetry APIs will not function unless these backend services are running and accessible. Ensure all environment variables and service URLs are set as described in the README. Before using the telemetry platform, verify that Elasticsearch and Kibana are operational. For detailed setup instructions, refer to the repository README and the documentation of each dependent AI service.

# Telemetry APIs

Our telemetry system collects monitoring data from multiple AI services and Large Language Models (LLMs) including **GPT-4o-mini**, **GPT-3.5 Turbo**, **AWS Anthropic Bedrock Claude models**, **LLaMA 3-70B**, **Gemini 2.5 Flash**, and **Gemini 2.5 Pro**. Please ensure that the upstream AI services using any of these models are properly configured and actively sending telemetry data to these APIs.

#### 3.1 Privacy Telemetry

Endpoint: /rai/v1/telemetry/privacytelemetryapi

This API collects telemetry data from privacy violation detection services that check for PII (Personally Identifiable Information) and sensitive data in AI interactions.

#### Input:

The telemetry payload includes privacy detection results, confidence scores, detected entity types (SSN, Aadhar, Credit Cards, Phone numbers, etc.), redaction actions taken, and



processing timestamps. The system captures both request and response privacy analysis results along with account and portfolio information for compliance tracking.

**Response:** Returns confirmation of telemetry data ingestion with processed data details, storage status in Elasticsearch, and unique tracking identifier for the telemetry record.

#### 3.2 Moderation Telemetry

Endpoint - /rai/v1/telemetry/moderationtelemetryapi

This API collects comprehensive telemetry from content moderation services including toxicity, bias, prompt injection, and jailbreak detection.

#### Input:

Telemetry payload includes moderation results for multiple checks (toxicity, bias, prompt injection, jailbreak, restricted topics), confidence scores, action taken (PASSED/FLAGGED/BLOCKED), processing times, and detailed violation categories with severity levels.

```
Request body required
Edit Value | Schema
   "uniqueid": "123e4567-e89b-12d3-a456-426614174000",
    "moderationResults": {
      "text": "Which is the biggest country in the world?",
      "promptInjectionCheck": {
        "injectionConfidenceScore": "0.98",
        "injectionThreshold": "0.70",
        "result": "PASSED"
     },
"jailbreakCheck": {
        "jailbreakSimilarityScore": "0.82",
"jailbreakThreshold": "0.6",
        "result": "PASSED"
      "privacyCheck": {
        "entitiesRecognised": [
          "PERSON"
        ],
"entitiesConfiguredToBlock": [
           "ADHAR NUMBER"
```

```
Edit Value | Schema
        customIhemeCheck": {
         "customSimilarityScore": "0.82",
         "themeThreshold": "0.6",
         "result": "PASSED"
       "summary": {
    "status": "REJECTED",
    "reason": [
            "PROMPT-INJECTION",
            "PRIVACY"
         ]
      }
    },
"Moderation layer time": {
       "Time for each individual check": {
         "Privacy Check": "string",
         "Text Quality Check": "string",
         "Toxicity Check": "string",
"Profanity Check": "string",
         "Prompt Injection Check": "string",
"Restricted Tonic Check": "string"
```

```
Protably Check: String,

"Prompt Injection Check": "string",

"Restricted Topic Check": "string",

"Refusal Check": "string",

"Custom Theme Check": "string",

"Time taken by each model": {

"Privacy Check": "string",

"Toxicity Check": "string",

"Prompt Injection Check": "string",

"Restricted Topic Check": "string",

"Custom Theme Check": "string",

"Custom Theme Check": "string",

"Toxicity Check": "string",

"Toxicity Check": "string",

"Custom Theme Check": "string",

"Custom Theme Check": "string",

"Profanity Check": 0,

"Profanity Check": 0,

"Prompt Injection Check": 0,

"Restricted Topic Check": 0,

"Restricted Topic Check": 0,

"Lailbreak Check": 0.
```

```
Edit Value | Schema

"Restricted Topic Check": 0,
    "Refusal Check": 0,
    "Refusal Check": 0,
    "Custom Theme Check": 0
},
    "Total time for moderation Check": "string",
    "Time By Model": 0,
    "Latency By API": 0,
    "Time By Validation": 0,
    "Time Difference mdoel and validity": 0,
    "Time Difference btwn ML and MM": 0
},
    "portfolioName": "string",
    "created": "2025-07-14T17:44:40.174Z",
    "accountName": "string",
    "userid": "string",
    "Source": "chatgpt/gemini",
    "lotNumber": "string"
}
```

**Response:** Returns acknowledgment of moderation telemetry receipt, Elasticsearch indexing status, and processed moderation results summary with unique record identifier.

```
Response body
         "Time taken by each model": {
            "Privacy Check": "string",
"Toxicity Check": "string",
"Prompt Injection Check": "string",
            "Restricted Topic Check": "string",
            "Jailbreak Check": "string",
"Custom Theme Check": "string"
         },
"Time taken By API": {
    Check": 0,
            "Toxicity Check": 0,
            "Profanity Check": 0,
            "Prompt Injection Check": 0,
            "Restricted Topic Check": 0,
            "Jailbreak Check": 0,
            "Refusal Check": 0,
            "Custom Theme Check": 0
         },
"Total time for moderation Check": "string",
         "Time By Model": 0,
         "Latency By API": 0,
"Time By Validation": 0,
"Time Difference mdoel and validity": 0,
         "Time Diffrence btwn ML and MM": 0
      },
"portfolioName": "string",
"created": "2025-07-15T06:09:53.481000Z",
"accountName": "string",
"string",
```

```
"userid": "string",

"Source": "chatgpt/gemini",

"lotNumber": "string"

}
}
```

#### **Endpoint –** /rai/v1/telemetry/moderationrequesttelemetryapi

This API captures telemetry data for moderation requests before processing, tracking request patterns and input characteristics.

**Input:** Request telemetry includes unique request IDs, input text characteristics, requested moderation checks, user context, timestamp data, and pre-processing metadata for trend analysis and performance monitoring.

```
POST
            /rai/v1/telemetry/moderationrequesttelemetryapi
Parameters
No parameters
Request body required
Edit Value | Schema
   "AccountName": "string",
   "userid": "string",
"PortfolioName": "string",
   "lotNumber": 0,
    "translate": "string",
    "EmojiModeration": "string",
   "Prompt": "string",
   "ModerationChecks": [
      "string"
    "ModerationCheckThresholds": {
     "PromptinjectionThreshold": 0,
     "JailbreakThreshold": 0,
     "PiientitiesConfiguredToBlock": [
        "string"
      "RefusalThreshold": 0,
      "ToxicityThresholds": {
  "ToxicityThreshold": 0,
```

```
"ToxicityThreshold": 0,
 "SevereToxicityThreshold": 0,
 "ObsceneThreshold": 0,
 "ThreatThreshold": 0,
 "InsultThreshold": 0,
 "IdentityAttackThreshold": 0,
 "SexualExplicitThreshold": 0
"ProfanityCountThreshold": 0,
"RestrictedtopicDetails": {
 "RestrictedtopicThreshold": 0,
 "Restrictedtopics": [
   "string"
"CustomTheme": {
 "Themename": "string",
 "Themethresold": 0,
 "ThemeTexts": [
   "string"
```

**Response:** Returns request telemetry processing confirmation, data validation status, and unique tracking ID for correlation with subsequent moderation results.

```
Response body
    "data": {
       "AccountName": "string",
"userid": "string",
"PortfolioName": "string",
       "Portfollowame";
"lotNumber": 0,
"translate": "string",
"iModeration": "string",
       "EmojiModeration": "si
"Prompt": "string",
"ModerationChecks": [
          "string"
       ],
"ModerationCheckThresholds": {
           "PromptinjectionThreshold": 0,
          "JailbreakThreshold": 0,
          "PiientitiesConfiguredToBlock": [
              "string"
           ],
"RefusalThreshold": 0,
"ToxicityThresholds": {
              "ToxicityThreshold": 0,
"SevereToxicityThreshold": 0,
              "ObsceneThreshold": 0,
              "ThreatThreshold": 0,
              "InsultThreshold": 0,
              "IdentityAttackThreshold": 0,
"SexualExplicitThreshold": 0
```

```
Response body
       ],
"RefusalThreshold": 0,
"Thresholds":
       "ToxicityThresholds": {
         "ToxicityThreshold": 0,
         "SevereToxicityThreshold": 0,
         "ObsceneThreshold": 0,
         "ThreatThreshold": 0,
         "InsultThreshold": 0,
         "IdentityAttackThreshold": 0,
         "SexualExplicitThreshold": 0
       },
"ProfanityCountThreshold": 0,
"profanityCountThreshold": {
         "RestrictedtopicThreshold": 0,
         "Restrictedtopics": [
            "string"
       "ThemeTexts": [
            "string"
```

#### **Endpoint** – /rai/v1/telemetry/coupledmoderationrequesttelemetryapi

This API collects telemetry from coupled moderation workflows that analyze both input prompts and expected outputs together.

#### Input:

Coupled request data includes paired input-output analysis requests, relationship mapping between prompt and response, workflow identifiers, processing pipeline information, and correlation metadata.



```
"ModerationCheckThresholds": {
 "PromptinjectionThreshold": 0,
 "JailbreakThreshold": 0,
 "PiientitiesConfiguredToBlock": [
   "string"
 ],
 "RefusalThreshold": 0,
 "ToxicityThresholds": {
   "ToxicityThreshold": 0,
   "SevereToxicityThreshold": 0,
   "ObsceneThreshold": 0,
   "ThreatThreshold": 0,
   "InsultThreshold": 0,
   "IdentityAttackThreshold": 0,
   "SexualExplicitThreshold": 0
 },
 "ProfanityCountThreshold": 0,
 "RestrictedtopicDetails": {
   "RestrictedtopicThreshold": 0,
 "RestrictedtopicThreshold": 0,
 "Restrictedtopics": [
   "string"
 ]
"CustomTheme": {
 "Themename": "string",
 "Themethresold": 0,
 "ThemeTexts": [
   "string"
 ]
"SmoothL1mThreshold": {
 "input_pertubation": 0,
 "number_of_iteration": 0,
 "SmoothLlmThreshold": 0
```

**Response:** Returns coupled moderation request processing status, workflow correlation ID, and confirmation of paired analysis data storage in telemetry system.

#### Response body "string" ], "RefusalThreshold": 0, "Thresholds": 1 "ToxicityThresholds": { "ToxicityThreshold": 0, "SevereToxicityThreshold": 0, "ObsceneThreshold": 0, "ThreatThreshold": 0, "InsultThreshold": 0, "IdentityAttackThreshold": 0, "SexualExplicitThreshold": 0 "RestrictedtopicDetails": { "RestrictedtopicThreshold": 0, "Restrictedtopics": [ "string" "CustomTheme": { "Themename": "string", "Themethresold": 0, "ThemeTexts": [ "string" SmoothLlmThreshold": { "input\_pertubation": 0,

```
},
    "SmoothLlmThreshold": {
        "input_pertubation": 0,
        "number_of_iteration": 0,
        "SmoothLlmThreshold": 0
     }
    }
}
```

#### 3.5 Moderation Test Telemetry

**Endpoint -** /rai/v1/telemetry/moderationtelemetryapitest

This API collects telemetry data from moderation testing environments for validation and performance benchmarking.

#### Input:

Test telemetry includes test case identifiers, expected vs actual results, performance benchmarks, accuracy metrics, false positive/negative rates, and test environment configuration data.

```
POST /rai/v1/telemetry/moderationtelemetryapitest
```

#### **Parameters**

No parameters

#### Request body required

```
Edit Value | Schema
```

```
"ADHAR_NUMBER"
],
"result": "PASSED"
},
"profanityCheck": {
   "profaneWordsIdentified": [
       "bullshit"
],
   "profaneWordsthreshold": "2",
   "result": "PASSED"
},
"toxicityCheck": {
   "toxicityScore": [
      {
       "toxicity": "0.85"
      }
],
"toxicitythreshold": "0.55",
   "result": "PASSED"
},
```

```
"themeThreshold": "0.6",
    "result": "PASSED"
  "summary": {
    "status": "REJECTED",
    "reason": [
      "PROMPT-INJECTION",
      "PRIVACY"
   1
 }
"Moderation layer time": {
 "Time for each individual check": {
   "Privacy Check": "string",
    "Text Quality Check": "string",
    "Toxicity Check": "string",
    "Profanity Check": "string",
    "Prompt Injection Check": "string",
    "Restricted Topic Check": "string",
    "Jailbreak Check": "string",
  "Refusal Check": "string",
  "Custom Theme Check": "string"
"Time taken by each model": {
  "Privacy Check": "string",
"Toxicity Check": "string",
"Prompt Injection Check": "string",
  "Restricted Topic Check": "string",
  "Jailbreak Check": "string",
  "Custom Theme Check": "string"
},
"Time taken By API": {
  "Toxicity Check": 0,
  "Profanity Check": 0,
  "Prompt Injection Check": 0,
  "Restricted Topic Check": 0,
  "Jailbreak Check": 0,
  "Refusal Check": 0,
  "Custom Theme Check": 0
},
  "Total time for moderation Check": "string",
  "Time By Model": 0,
  "Latency By API": 0,
  "Time By Validation": 0,
  "Time Difference mdoel and validity": 0,
  "Time Diffrence btwn ML and MM": 0
},
"portfolioName": "string",
"created": "2025-07-15T06:18:42.673Z",
"accountName": "string",
"userid": "string",
"Source": "chatgpt/gemini",
"lotNumber": "string"
```

**Response:** Returns test telemetry ingestion confirmation, test results validation status, benchmark data processing acknowledgment, and unique test execution tracking ID.

```
Response body
    "data": {
      "uniqueid": "123e4567-e89b-12d3-a456-426614174000",
      "moderationResults": {
         "text": "Which is the biggest country in the world?",
         "promptInjectionCheck": {
    "injectionConfidenceScore": "0.98",
           "injectionThreshold": "0.70",
           "result": "PASSED"
        },
"jailbreakCheck": {
  "jailbreakSimilarityScore": "0.82",
  "jailbreakThreshold": "0.6",
  "result": "PASSED"
,
        },
"privacyCheck": {
            "entitiesRecognised": [
           ],
"entitiesConfiguredToBlock": [
              "ADHAR_NUMBER"
           ],
"result": "PASSED"
        },
"profanityCheck": {
            "profaneWordsIdentified": [
"bullshit"
```

```
Response body

"RefusalThreshold": "0.6",
"result": "PASSED"
},
"customThemeCheck": {
  "customSimilarityScore": "0.82",
  "themeThreshold": "0.6",
  "result": "PASSED"
},
"summary": {
  "status": "REJECTED",
  "reason": [
    "PROMPT-INJECTION",
    "PRIVACY"
]
},
"Moderation layer time": {
  "Time for each individual check": {
  "Privacy Check": "string",
   "Text Quality Check": "string",
   "Toxicity Check": "string",
   "Profanity Check": "string",
   "Profanity Check": "string",
   "Restricted Topic Check": "string",
   "Restricted Topic Check": "string",
   "Restricted Topic Check": "string",
   "Restricted Topic Check": "string",
   "Custom Theme Check": "string",
   "Custom Theme Check": "string"
},

"Custom Theme Check": "string"
},
```

```
"Time taken by each model": {
    "Privacy Check": "string",
    "Toxicity Check": "string",
    "Prompt Injection Check": "string",
    "Restricted Topic Check": "string",
    "Jailbreak Check": "string",
    "Custom Theme Check": "string"
},

"Time taken By API": {
    "Toxicity Check": 0,
    "Profanity Check": 0,
    "Prompt Injection Check": 0,
    "Restricted Topic Check": 0,
    "Restricted Topic Check": 0,
    "Sailbreak Check": 0,
    "Custom Theme Check": 0,
    "Custom Theme Check": 0,
    "Total time for moderation Check": "string",
    "Time By Model": 0,
    "Latency By API": 0,
    "Time Difference mdoel and validity": 0,
    "Time Difference btwn ML and MM": 0
},
    "portfolioName": "string",
    "created": "2025-07-15T06:18:42.6730002",
    "accountName": "string",
    "userid": "string",
```

```
"userid": "string",
    "Source": "chatgpt/gemini",
    "lotNumber": "string"
}
}
```

#### 3.6 Coupled Moderation Telemetry

**Endpoint** – /rai/v1/telemetry/coupledmoderationtelemetryapi

This API captures telemetry from integrated moderation workflows that process complete request-response cycles.

#### Input:

Comprehensive moderation results including input analysis, output analysis, correlation findings, workflow execution times, decision reasoning, and complete moderation layer performance metrics.



```
"result": "PASSED"
"jailbreakCheck": {
  "jailbreakSimilarityScore": "0.82",
  "jailbreakThreshold": "0.6",
  "result": "PASSED"
"privacyCheck": {
  "entitiesRecognised": [
    "LOCATION"
  "entitiesConfiguredToBlock": [
    "ADHAR_NUMBER"
  ],
"result": "PASSED"
},
"profanityCheck": {
  "profaneWordsIdentified": [
    "bullshit"
  ],
```

```
"profaneWordsthreshold": "2",
   "result": "PASSED"
 },
 "toxicityCheck": {
   "toxicityScore": [
     {
       "toxicScore": [
           "metricName": "toxicity",
"metricScore": 0.0006679997895844281
           "metricName": "severe_toxicity",
           "metricScore": 0.0000013805756680085324
           "metricName": "obscene",
           "metricScore": 0.00003030665720871184
         "metricName": "threat",
         "metricScore": 0.000021546358766499907
         "metricName": "insult",
         "metricScore": 0.00013677669630851597
       },
         "metricName": "identity_attack",
          "metricScore": 0.00017613476666156203
       },
         "metricName": "sexual_explicit",
"metricScore": 0.000014234438822313678
     ]
   }
 "toxicitythreshold": "0.55",
 "result": "PASSED",
  "toxicityTypesConfiguredToBlock": [
    "toxicity",
    "severe_toxicity",
    "obscene",
    "threat",
    "insult",
    "identity_attack",
    "sexual_explicit"
  "toxicityTypesRecognised": [
    "string"
},
 'restrictedtopic": {
  "topicScores": [
       "Explosives": "0.85"
  "topicThreshold": "0.65",
```

```
"result": "PASSED",
  "topicTypesConfiguredToBlock": [
    "Terrorism",
"Explosives"
  ],
  "topicTypesRecognised": []
"textGrade": "Grade 12-13"
"refusalCheck": {
  "refusalSimilarityScore": "0.82",
  "RefusalThreshold": "0.6",
  "result": "PASSED"
},
"customThemeCheck": {
  "customSimilarityScore": "0.82",
  "themeThreshold": "0.6",
  "result": "PASSED"
 "randomNoiseCheck": {
   "smoothLlmScore": "0.82",
    "smoothLlmThreshold": "0.6",
    "result": "PASSED"
  "advancedJailbreakCheck": {
    "text": "NON ADVERSARIAL",
"result": "PASSED"
 },
  "summary": {
    "status": "PASSED",
    "reason": []
 }
responseModeration": {
  "generatedText": "Russia is the biggest country by area.",
  "privacyCheck": {
    "entitiesRecognised": [
      "LOCATION"
    ],
```

```
"entitiesConfiguredToBlock": [
    "ADHAR_NUMBER"
  ],
  "result": "PASSED"
},
"profanityCheck": {
  "profaneWordsIdentified": [
    "bullshit"
  ],
  "profaneWordsthreshold": "2",
  "result": "PASSED"
"toxicityCheck": {
  "toxicityScore": [
      "toxicScore": [
          "metricName": "toxicity",
          "metricScore": 0.0006679997895844281
        },
```

```
"metricName": "severe_toxicity",
   "metricScore": 0.0000013805756680085324
},
{
   "metricName": "obscene",
   "metricScore": 0.00003030665720871184
},
{
   "metricName": "threat",
   "metricScore": 0.000021546358766499907
},
{
   "metricName": "insult",
   "metricScore": 0.00013677669630851597
},
{
   "metricName": "identity_attack",
   "metricScore": 0.00017613476666156203
},
{
```

```
"metricName": "sexual_explicit",
        "metricScore": 0.000014234438822313678
  }
],
"toxicitythreshold": "0.55",
"result": "PASSED",
"toxicityTypesConfiguredToBlock": [
  "toxicity",
  "severe_toxicity",
  "obscene",
  "threat",
  "insult",
  "identity_attack",
  "sexual_explicit"
],
"toxicityTypesRecognised": [
  "string"
```

```
"restrictedtopic": {
  "topicScores": [
      "Explosives": "0.85"
 ],
  "topicThreshold": "0.65",
  "result": "PASSED",
  "topicTypesConfiguredToBlock": [
   "Terrorism",
   "Explosives"
  ],
 "topicTypesRecognised": []
"textQuality": {
  "readabilityScore": "80",
  "textGrade": "Grade 12-13"
"textRelevanceCheck": {
 "PromptResponseSimilarityScore": "0.82"
```

```
"refusalCheck": {
     "refusalSimilarityScore": "0.82",
     "RefusalThreshold": "0.6",
     "result": "PASSED"
   },
   "summary": {
     "status": "PASSED",
     "reason": []
 }
portfolioName": "string",
accountName": "string",
'Moderation layer time": {
 "requestModeration": {
   "promptInjectionCheck": "string",
   "jailbreakCheck": "string",
   "toxicityCheck": "string",
   "privacyCheck": "string",
   "profanityCheck": "string",
 "refusalCheck": "string",
  "restrictedtopic": "string"
 "textqualityCheck": "string",
  "customthemeCheck": "string",
  "smoothLlmCheck": "string",
 "bergeronCheck": "string"
"responseModeration": {
  "promptInjectionCheck": "string",
 "jailbreakCheck": "string",
 "toxicityCheck": "string",
"privacyCheck": "string",
  "profanityCheck": "string",
 "refusalCheck": "string",
  "restrictedtopic": "string",
  "textqualityCheck": "string",
 "customthemeCheck": "string",
 "smoothLlmCheck": "string",
  "bergeronCheck": "string",
  "textrelevanceCheck": "string"
 "OpenAIInteractionTime": "string",
 "translate": "string",
 "Time_taken_by_each_model_in_requestModeration": {
   "toxicityCheck": "string",
   "privacyCheck": "string",
   "jailbreakCheck": "string",
   "promptInjectionCheck": "string",
   "customthemeCheck": "string",
   "restrictedtopic": "string"
 "Time_taken_by_each_model_in_responseModeration": {
   "toxicityCheck": "string",
   "privacyCheck": "string",
   "jailbreakCheck": "string"
   "promptInjectionCheck": "string",
   "customthemeCheck": "string",
   "restrictedtopic": "string"
 }
```

**Response:** Returns coupled moderation telemetry processing confirmation, complete workflow analysis status, and integrated results storage acknowledgment with correlation tracking.

```
Response body
   "data": {
      "uniqueid": "123e4567-e89b-12d3-a456-426614174000",
     "object": "text_completion",
"userid": "admin",
     "lotNumber": "1",
"created": "1646932609",
      "model": "openai",
      "choices": [
          "text": "Russia is the biggest country by area.",
          "index": 0,
          "finishReason": "length"
      'moderationResults": {
        "requestModeration": {
          "text": "Which is the biggest country in the world?",
          "promptInjectionCheck": {
             "injectionConfidenceScore": "0.98",
            "injectionThreshold": "0.70",
             "result": "PASSED"
          },
"jailbreakCheck": {
    "jailbreakSimilarityScore": "0.82",
    "i-lbceakIhreshold": "0.6",
             "result": "PASSED"
```

```
Response body
          "privacyCheck": {
             "entitiesRecognised": [
               "LOCATION"
            "entitiesConfiguredToBlock": [
               "ADHAR_NUMBER"
            ],
"result": "PASSED"
          "profanityCheck": {
             "profaneWordsIdentified": [
               "bullshit"
            ],
"profaneWordsthreshold": "2",
          "toxicityScore": [
                 "toxicScore": [
                     "metricName": "toxicity",
"metricScore": 0.0006679997895844281
                     "metricName": "severe_toxicity",
"metricScore": 0.0000013805756680085324
```

```
Response body
             "obscene",
             "threat",
             "insult",
             "identity_attack",
             "sexual_explicit"
           "toxicityTypesRecognised": [
             "string"
         "restrictedtopic": {
           "topicScores": [
               "Explosives": "0.85"
           "topicThreshold": "0.65",
           "result": "PASSED",
           "topicTypesConfiguredToBlock": [
             "Terrorism",
             "Explosives"
           "topicTypesRecognised": []
         "textQuality": {
           "readabilityScore": "80",
           "textGrade": "Grade 12-13"
```

```
Response body
                 "refusalCheck": {
    "refusalSimilarityScore": "0.82",
                   "RefusalThreshold": "0.6",
"result": "PASSED"
                "customSimilarityScore": "0.82",
                   "themeThreshold": "0.6",
"result": "PASSED"
               },
"randomNoiseCheck": {
  "smoothLlmScore": "0.82",
  "smoothLlmThreshold": "0.6",
  "PASSED"
               },
"advancedJailbreakCheck": {
  "text": "NON ADVERSARIAL",
  "result": "PASSED"
               },
"summary": {
    "status": "PASSED",
    "reason": []
           },
"responseModeration": {
  "generatedText": "Russia is the biggest country by area.",
  "privacyCheck": {
    "entitiesRecognised": [
```

#### Response body

```
"LOCATION"
  ],
"entitiesConfiguredToBlock": [
     "ADHAR_NUMBER"
  ],
"result": "PASSED"
"profaneWordsIdentified": [
     "bullshit"
  ],
"profaneWordsthreshold": "2",
"DASSED"
"toxicityScore": [
      "toxicScore": [
           "metricName": "toxicity",
"metricScore": 0.0006679997895844281
           "metricName": "severe_toxicity",
"metricScore": 0.0000013805756680085324
         },
{
           "metricName": "obscene",
```

```
Response body

"metricScore": 0.00003030665720871184

},

{
    "metricName": "threat",
    "metricScore": 0.000021546358766499907

},

{
    "metricName": "insult",
    "metricScore": 0.00013677669630851597

},

{
    "metricName": "identity_attack",
    "metricScore": 0.00017613476666156203

},

{
    "metricName": "sexual_explicit",
    "metricScore": 0.000014234438822313678

}

}

]

}

// "toxicitythreshold": "0.55",
    "result": "PASSED",
    "toxicityTypesConfiguredToBlock": [
    "toxicity",
    "severe_toxicity",
    "obscene",
    "threat",
    "insult"
```

```
Response body
               "insult",
               "identity_attack",
               "sexual_explicit"
            ],
"toxicityTypesRecognised": [
              "string"
          "topicScores": [
                 "Explosives": "0.85"
            ],
"topicThreshold": "0.65",
"DASSED".
            "result": "PASSED",
            "topicTypesConfiguredToBlock": [
              "Terrorism",
"Explosives"
            ],
"topicTypesRecognised": []
         },
"textQuality": {
   "readabilityScore": "80",
   "textGrade": "Grade 12-13"
          },
"textRelevanceCheck": {
            "PromptResponseSimilarityScore": "0.82"
```

#### Response body

```
"refusalCheck": {
         "refusalSimilarityScore": "0.82",
         "RefusalThreshold": "0.6",
         "result": "PASSED"
       "summary": {
    "status": "PASSED",
         "reason": []
      }
  }
"portfolioName": "string",
"accountName": "string",
"Moderation layer time": {
   "requestModeration": {
      "promptInjectionCheck": "string",
      "jailbreakCheck": "string",
"toxicityCheck": "string",
"privacyCheck": "string",
      "profanityCheck": "string",
"refusalCheck": "string",
      "restrictedtopic": "string",
"textqualityCheck": "string",
      "customthemeCheck": "string",
"smoothLlmCheck": "string",
      "bergeronCheck": "string"
```

#### Response body

```
"responseModeration": {
    "promptInjectionCheck": "string",
    "jailbreakCheck": "string",
    "toxicityCheck": "string",
    "privacyCheck": "string",
    "profanityCheck": "string",
    "refusalCheck": "string",
    "restrictedtopic": "string",
    "textqualityCheck": "string",
    "smoothLlmCheck": "string",
    "bergeronCheck": "string",
    "textrelevanceCheck": "string",
    "translate": "string",
    "translate": "string",
    "Time_taken_by_each_model_in_requestModeration": {
        "toxicityCheck": "string",
        "privacyCheck": "string",
        "promptInjectionCheck": "string",
        "restrictedtopic": "string",
        "restrictedtopic": "string",
        "rotal_time_for_moderation_Check": "string",
        "Total_time_for_moderation_Check": "string",
        "Time_taken_by_each_model_in_responseModeration": {
        "toxicityCheck": "string",
        "privacyCheck": "string",
```

```
"jailbreakCheck": "string",
    "promptInjectionCheck": "string",
    "customthemeCheck": "string",
    "restrictedtopic": "string"
}
}
}
```

#### 3.7 Admin Telemetry

#### **Endpoint** – /rai/v1/telemetry/admintelemetryapi

This API collects telemetry data from administrative operations and system management activities.

#### Input:

Administrative telemetry includes user management actions, system configuration changes, access control modifications, audit trail data, administrative decision logs, and system health metrics.

```
Post /rai/v1/telemetry/admintelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
    "tenant": "string",
    "apiname": "string",
    "date": "2025-07-15T06:25:35.244Z",
    "admin_requests": {
        "recognizer_name": "string",
        "recognizer_value_pattern": "string",
        "entity": "string",
        "context": "string",
        "context": "string",
        "score_range": "string"
    }
}
```

**Response:** Returns administrative telemetry processing confirmation, audit trail storage status, compliance logging acknowledgment, and administrative action tracking ID.

```
Response body

{
    "data": {
        "tenant": "string",
        "apiname": "string",
        "admin_requests": {
        "recognizer_name": "string",
        "recognizer_type": "string",
        "recognizer_value_pattern": "string",
        "entity": "string",
        "score_range": "string"
}
}
```

#### 3.8 Error Logging Telemetry

**Endpoint –** /rai/v1/telemetry/errorloggingtelemetryapi

This API collects comprehensive errors and exception telemetry from all AI services for system monitoring.

#### Input:

Error telemetry includes exception details, stack traces, error categorization, system impact assessment, recovery actions, error frequency patterns, and service health indicators.

```
POST /rai/v1/telemetry/errorloggingtelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
   "tenetName": "Tenet Name",
   "errorCode": "Error code",
   "errorMessage": "Error Message",
   "apiEndPoint": "API Endpoint",
   "errorRequestMethod": "ErrorRequestMethod"
}
```

**Response:** Returns error telemetry ingestion confirmation, error categorization status, monitoring alert trigger acknowledgment, and unique error tracking identifier for incident management.

```
Response body

{
    "data": {
        "tenetName": "Tenet Name",
        "errorCode": "Error code",
        "errorMessage": "Error Message",
        "apiEndPoint": "API Endpoint",
        "errorRequestMethod": "ErrorRequestMethod"
    }
}
```

#### 3.9 Account Master Telemetry

#### **Endpoint** – /rai/v1/telemetry/accMastertelemetryapi

This API captures telemetry from account and portfolio management operations across multitenant environments.

#### Input:

Account telemetry includes portfolio information, tenant segregation data, resource usage metrics, account-level configuration changes, billing information, and multi-tenant access patterns.

```
Post /rai/v1/telemetry/accMastertelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
    "tenant": "string",
    "apiname": "string",
    "date": "2025-07-15T06:26:35.399Z",
    "accMaster_requests": {
        "portfolio_name": "string",
        "account_name": "string",
        "dataGrp_list": [
        "string"
        ]
    }
}
```

**Response:** Returns account telemetry processing confirmation, multi-tenant data segregation status, portfolio analytics update acknowledgment, and account-level tracking identifier.

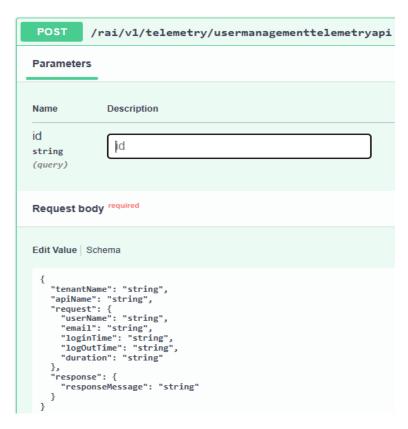
#### 3.10 User Management Telemetry

**Endpoint –** /rai/v1/telemetry/usermanagementtelemetryapi

This API collects telemetry from authentication and user identity management services.

#### Input:

User management data includes authentication events, login/logout activities, session management, role assignments, permission changes, security events, and identity verification results.



**Response:** Returns user management telemetry processing confirmation, authentication event logging status, security monitoring acknowledgment, and user activity tracking ID.

```
Response body

{
    "data": {
        "tenantName": "string",
        "apiName": "string",
        "request": {
            "userName": "string",
            "loginTime": "string",
            "logOutTime": "string",
            "duration": "string"
        },
        "response": {
            "responseMessage": "string"
        }
    }
}
```

#### 3.11 Registration Telemetry

**Endpoint** – /rai/v1/telemetry/registertelemetryapi

This API captures telemetry from user registration and onboarding processes.

**Input:** Registration telemetry includes new user signup data, onboarding completion rates, verification processes, account activation events, and initial configuration settings.

```
Post /rai/v1/telemetry/registertelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
    "tenant": "string",
    "apiname": "string",
    "date": "2025-07-15T06:27:36.303Z",
    "register_requests": {
        "email": "string",
        "login": "string",
        "password": "string",
        "password": "string",
    }
}
```

**Response:** Returns registration telemetry ingestion confirmation, onboarding analytics update status, user lifecycle tracking acknowledgment, and registration event identifier.

```
Response body

{
    "data": {
        "tenant": "string",
        "apiname": "string",
        "date": "2025-07-15T06:27:38.644345",
        "register_requests": {
            "email": "string",
            "login": "string",
            "password": "string"
        }
    }
}
```

#### 3.12 Profanity Telemetry

**Endpoint** – /rai/v1/telemetry/profanitytelemetryapi

This API collects telemetry data from profanity detection services that monitor for inappropriate language and content violations.

**Input:** The telemetry data includes profanity detection results, severity levels, flagged terms, confidence scores, content filtering actions, and language detection metadata. Captures both input prompt and output response profanity analysis for comprehensive monitoring.

```
POST
             /rai/v1/telemetry/profanitytelemetryapi
Parameters
No parameters
Request body required
Edit Value | Schema
    "uniqueid": "string",
   "uniqueld": "string",
"tenant": "string",
"apiname": "string",
"date": "2025-07-15T06:28:11.662Z",
"user": "admin",
"lotNumber": "1",
"request": {
      "inputText": "string"
    "response": {
      "profanity": [
        {
    "profaneWord": "string",
           "beginOffset": 0,
           "endOffset": 0
      "profanityScoreList": [
      ],
      "profanityScoreList": [
             "metricName": "string",
            "metricScore": 0
      ],
      "outputText": "string"
```

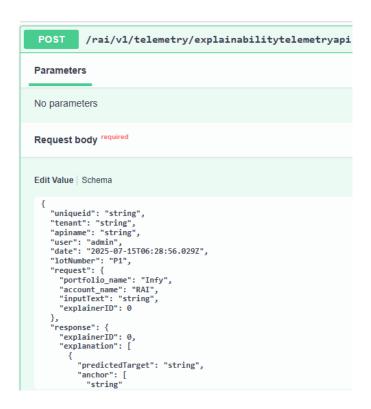
**Response:** Returns profanity telemetry processing confirmation, content filtering analytics update, violation pattern analysis status, and unique profanity detection tracking ID.

#### 3.13 Explainability Telemetry

**Endpoint** – /rai/v1/telemetry/explainabilitytelemetryapi

This API collects telemetry from AI explainability engines that provide transparency into AI decision-making.

**Input:** Explainability data includes Chain of Thought reasoning, Thread of Thought analysis, decision rationale, confidence explanations, reasoning step tracking, and transparency metrics.



**Response:** Returns explainability telemetry ingestion confirmation, transparency analytics update status, reasoning chain storage acknowledgment, and explainability session tracking ID.

## Response body "data": { "uniqueid": "string", "tenant": "string", "apiname": "string", "user": "admin", "date": "2025-07-15T06:28:58.279278", "lotNumber": "P1", "request": { "portfolio\_name": "Infy", "account\_name": "RAI", "inputText": "string", "explainerID": 0 "response": { "explainerID": 0, "explanation": [ "predictedTarget": "string", "anchor": [ "string"

#### 3.14 Explainability Bulk Telemetry

**Endpoint –** /rai/v1/telemetry/explainabilitybulktelemetryapi

This API handles bulk processing telemetry from explainability services processing multiple requests simultaneously.

#### Input:

Bulk processing telemetry includes batch processing metrics, throughput statistics, bulk analysis results, processing efficiency data, and aggregated explainability insights.

```
Post /rai/v1/telemetry/explainabilitybulktelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
    "uniqueId": "string",
    "tenetName": "string",
    "apiName": "string",
    "userId": "string",
    "date": "2025-07-15T06:29:30.929Z",
    "response": [
    {
        "InputPrompt": "string",
        "Response": "string",
        "Chain_of_Thought": "string",
        "Token_Cost": 0,
        "additionalProp1": {}
    }
    ]
}
```

**Response:** Returns bulk explainability telemetry processing confirmation, batch analytics update status, throughput metrics acknowledgment, and bulk processing session identifier.

#### 3.15 Error Log Telemetry (old)

**Endpoint –** /rai/v1/telemetry/errorlogtelemetryapi

This API handles legacy error logging telemetry for backward compatibility and historical data analysis.

#### Input:

Legacy error data includes historical error patterns, migration status, compatibility metrics, and transition tracking for system upgrade monitoring.

```
Post /rai/v1/telemetry/errorlogtelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
    "uniqueid": "string",
    "error": [
        {
        "UUID": "string",
        "function": "string",
        "msg": "string",
        "description": "string"
        }
    ]
}
```

#### Response:

Returns legacy error telemetry processing confirmation, historical data migration status, compatibility analysis acknowledgment, and legacy system tracking identifier.

#### 3.16 LLM Evaluation Telemetry

**Endpoint –** /rai/v1/telemetry/evalllmtelemetryapi

This API collects telemetry from LLM evaluation and assessment services monitoring model performance.

**Input:** Evaluation telemetry includes model performance metrics, response quality assessments, accuracy measurements, bias evaluations, hallucination detection results, and comparative model analysis data.

```
Post /rai/v1/telemetry/evalllmtelemetryapi

Parameters

Request body required

Edit Value | Schema

{
    "uniqueid": "string",
    "userid": "string",
    "accountName": "string",
    "portfolioName": "string",
    "lotNumber": "string",
    "created": "2025-07-15T06:30:38.2592",
    "model": "string",
    "moderationResults": {
        "analysis": "string",
        "score": 0,
        "threshold": 0,
        "result": "string",
        "evaluation_check": "string",
        "timeTaken": "string",
        "description": "string"
}
```

**Response:** Returns LLM evaluation telemetry ingestion confirmation, model performance analytics update, evaluation metrics storage status, and model assessment tracking ID.

```
Response body

{
    "data": {
        "uniqueid": "string",
        "userid": "string",
        "portfolioName": "string",
        "lotNumber": "string",
        "created": "2025-07-15T06:30:38.259000Z",
        "model": "string",
        "moderationResults": {
            "analysis": "string",
            "score": 0,
            "threshold": 0,
            "result": "string"
        },
        "evaluation_check": "string",
        "timeTaken": "string",
        "description": "string"
    }
}
```

#### 3.17 Risk Classification Telemetry

**Endpoint -** /rai/v1/telemetry/riskclassificationtelemetryapi

This API captures telemetry from risk assessment and classification services analyzing AI safety risks.

**Input:** Risk classification data includes risk scores, threat categorization, severity assessments, mitigation recommendations, risk trend analysis, and safety violation classifications.

```
POST /rai/v1/telemetry/riskclassificationtelemetryapi

Parameters

No parameters

Request body required

Edit Value | Schema

{
    "userid": "string",
    "apiCall": "string",
    "resultType": "string",
}
```

**Response:** Returns risk classification telemetry processing confirmation, threat analytics update status, risk assessment storage acknowledgment, and unique risk evaluation tracking identifier.

```
Response body

{
    "data": {
        "userid": "string",
        "apiCall": "string",
        "resultType": "string"
    }
}
```



# Endpoints usage flow

Endpoint	Description
/rai/v1/telemetry/privacytelemetryapi	To collect and store telemetry data from privacy violation detection services for PII monitoring and compliance tracking.
/rai/v1/telemetry/moderationtelemetryapi	To capture moderation results telemetry including toxicity, bias, prompt injection, and content safety checks.
/rai/v1/telemetry/moderationrequesttelemetryapi	To log moderation request telemetry for tracking input patterns and pre-processing analytics.
/rai/v1/telemetry/coupledmoderationrequesttelemetryapi	To collect coupled moderation request telemetry for paired input-output analysis workflows.
/rai/v1/telemetry/moderationtelemetryapitest	To capture test environment moderation telemetry for validation and performance benchmarking.
/rai/v1/telemetry/coupledmoderationtelemetryapi	To store integrated moderation workflow telemetry for complete request-response cycle monitoring.
/rai/v1/telemetry/admintelemetryapi	To collect administrative operations telemetry for system management and audit trail monitoring.
/rai/v1/telemetry/errorloggingtelemetryapi	To capture comprehensive error and exception telemetry for system health monitoring and incident management.
/rai/v1/telemetry/accMastertelemetryapi	To store account and portfolio management telemetry for multi-tenant analytics and resource tracking.
/rai/v1/telemetry/usermanagementtelemetryapi	To collect authentication and user identity management telemetry for security monitoring.
/rai/v1/telemetry/registertelemetryapi	To capture user registration and onboarding telemetry for lifecycle analytics and completion tracking.
/rai/v1/telemetry/profanitytelemetryapi	To store profanity detection telemetry for content filtering monitoring and violation pattern analysis.
/rai/v1/telemetry/explainabilitytelemetryapi	To collect AI explainability telemetry for transparency monitoring and decision rationale tracking.
/rai/v1/telemetry/explainabilitybulktelemetryapi	To handle bulk explainability processing telemetry for batch analytics and throughput monitoring.



/rai/v1/telemetry/errorlogtelemetryapi	To manage legacy error logging telemetry for backward compatibility and historical data analysis.
/rai/v1/telemetry/evallImtelemetryapi	To capture LLM evaluation telemetry for model performance monitoring and assessment analytics.
/rai/v1/telemetry/riskclassificationtelemetryapi	To store risk assessment telemetry for threat categorization and safety violation monitoring.