ProcessUnity

# Technical Overview: ProcessUnity Evidence Evaluator

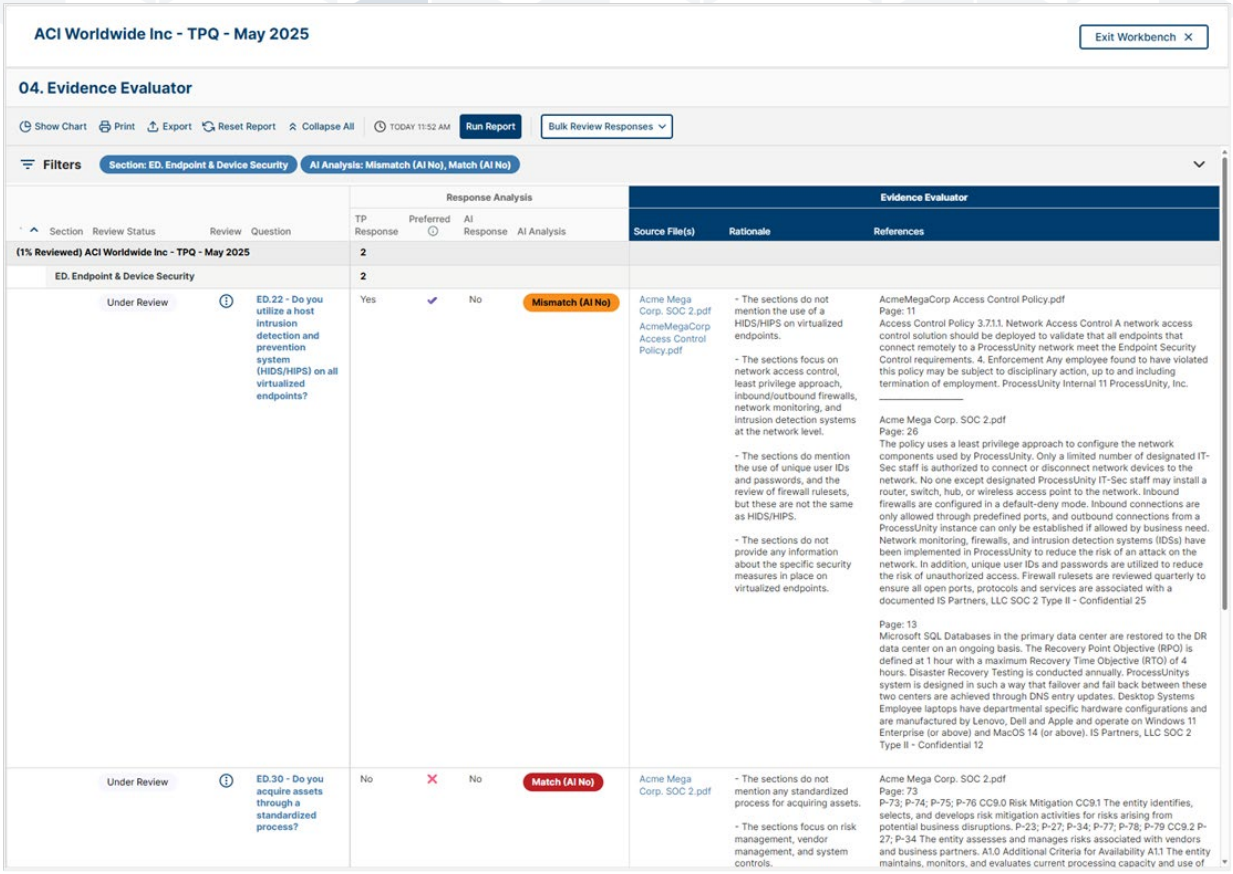
This document is a comprehensive overview of the training methodology for Evidence Evaluator and inference characteristics. The aim of this document is to detail the technology differentiator inherent to ProcessUnity's training data and explain how Evidence Evaluator produces results at scale.

## What is Evidence Evaluator?

Evidence Evaluator is an AI-enabled feature that accelerates assessment cycle times by reducing the time it takes a third-party analyst to review a third party's submitted evidence for control validation. This feature uses Natural Language Processing (NLP) and small language models to extract information from submitted evidence documents (security policies, completed questionnaires, etc.), then uses Generative AI to produce an AI response and final summarization of the results

## Evidence Evaluator Overview

Once a third party submits their assessment with the attached evidence documents, Evidence Evaluator will scan the attachments for relevance to the assessment questions. Then, Evidence Evaluator will choose the most relevant pieces of information to generate an answer to the assessment question and provide the reasoning and references supporting that answer. Data is encrypted both in-transit and at rest to protect sensitive data in the evidence documents. Furthermore, after each use of Evidence Evaluator, all uploaded data is removed leaving no trace of sensitive information.




***In the Assessment Report, Evidence Evaluator displays the assessment question, alongside the third-party response, a yes/no preferred response, the AI response, and the AI analysis. Evidence Evaluator provides the identified relevant documents for the question, along with extracted rationale and references inside the source documents.***

## Training Dataset and Methodology Overview

### Framework Mapper Dataset

The starting dataset contains over a quarter million in-house cybersecurity text pairs labelled by our security analysts, who have decades of experience in the field. This dataset is curated to represent question-to-passage pairs where the passage is the information relevant to the question. Each passage pair is weighted by relevancy as there can be many individual passages relevant to the same question. As of early 2025, *there are no standardized human labelled publicly available datasets that have cybersecurity text similarity pairs at this volume.*

| control1 | control2 | weight |
|---|---|---|
| Do you establish an authentication standard? | Cloud Storage Buckets Controls. | 0.5 |
| Do you establish an authentication standard? | Define and implement mechanisms for authentication, authorisation, identity lifecycle management, Single Sign On, Federation, Adaptive Authentication using Toll's endorsed Identity and Access Management Solution. This includes compliance to password management policies. | 1 |

The.framework.mapper.dataset.is.a.mapping.between.the.Cyber.Controls.Questionnaire.questions.and.controls.from.various.frameworks¡.Each.mapping.is.given.a.weight?assigned.by.a.human.expert?for.relevancy¡

## Extended Framework Pairs

The Framework Mapper dataset is further expanded using our Cyber Controls Questionnaire (CCQ) as steppingstones from one framework control to another.

We can extend these mappings to map frameworks to frameworks, not just frameworks to CCQ controls, by relating two controls from different frameworks through the CCQ.

**Original Mappings:**

| Framework 1 | CCQ |
|---|---|
| A1 | Q1 |
| B1 | Q2 |
| C1 | Q3 |

| Framework 2 | CCQ |
|---|---|
| A2 | Q1 |
| B2 | Q2 |
| C2 | Q3 |

**Relation:**

| Framework 1 | CCQ | Framework 2 |
|---|---|---|
| A1 | Q1 | A2 |
| B1 | Q2 | B2 |
| C1 | Q3 | C2 |

**Extension:**

| Framework 1 | Framework 2 |
|---|---|
| A1 | A2 |
| B1 | B2 |
| C1 | C2 |

The resulting unified dataset characterizes:

- Upwards of 40-million pairs
- All 40-million pairs are topic clustered through the CCQ
- All 40-million passages are weighted by using the minimum of the weights for each pair that was derived from being primary (1.0) to the CCQ, or supporting (0.5) to the CCQ

This amount of data enables us to efficiently aggregate, filter, fine-tune, and iterate on models capable of retrieving sections of text from documents.

An example of the resulting training dataset of pairs is shown below after filtering, aggregating framework controls, and averaging scores.

| question | positive passage | score | negative passage | score |
|---|---|---|---|---|
| Do you have a disaster recovery plan and emergency mode operation plan? | Data is appropriately backed up and archived in compliance to Toll's backup and retention policies. Requirements for continuity of the system in adverse situations, e.g. during a crisis or disaster are determined and implemented for all applications and services that process or hold sensitive information. Define and document business impact of any disruption that considers dependencies, processes, applications, business partners and third party providers. | 0.8333 | Data protection and associated media protection must be in accordance with the Cyber Security Standard – Data Security. Encryption must be applied in accordancewith the Cyber Security Standard – Data Security. Maintain a centralised inventory(CMDB) and record / register Information Services including ownership,data classification, business impact assessment, risk rating, and location.Update as underlying assets in the service change. Create and securely storebackups based on business system owner agreed backup cycles (daily, weekly, monthly).Encrypt backups where required. | 0 |

This.is.an.example.row.of.the.resulting.training.dataset.with.aggregated.pairs.to.create.positive.(relevant).passages.and.negative.(similar.language?not.relevant).passages¡

### Cybersecurity Relevant Document Test Dataset

Evidence Evaluator is tested on real-world cybersecurity data to verify its ability to identify the documents and text passages as we would expect to see in practice.

### Hard Negatives

One of the training techniques used is the method of pairing similar passages in the training set with different questions. These pairs are not relevant to each other, i.e. negative pairs, but exhibit similar language use. By pairing similar passages in both positive and negative examples, the model is forced to learn distinctions in context, language, and topic, rather than simply identifying related words.

## Inference at Scale

### No Performance Hardware Dependency

Our proprietary relevancy model is light weight at a few hundred megabytes.
The model is compiled with the inference code, allowing it to be as mobile as the code is when packaged. The model only requires CPU hardware as opposed to more expensive, higher carbon-emitting, difficult to provision, inflexible hardware while still performing at a high level. This allows

for Evidence Evaluator to achieve much higher and more consistent uptime than depending on specialized hardware.

### No Queueing Mechanism

With performance hardware, there needs to be a queueing mechanism as requests are limited by the amount of onboard hardware memory. We are not constrained by queueing requests that cause halts for customer results. Being independent of a queueing mechanism opens inference to be horizontally scalable while remaining cost-effective. In our tests using real world documents, we find that five-hundred documents can be processed in under two minutes.

### AWS Infrastructure

Evidence Evaluator is hosted in AWS cloud infrastructure for scalability and managed services, further reducing our development time and exposure to security vulnerabilities.

### Dynamically scalable

The inference pipeline is scalable with the ability to support 5,000 *simultaneous requests per minute of roughly 500 documents per request. Since the documents are processed in parallel, the average runtime of one minute is unchanged as the documents and requests increase*.

### Multi-Region Deployment

The necessary components are also light as to be implemented easily in multiple regions through AWS. We maintain stateless components that only require CPUs and storage that can be initialized across many AWS regions to support our customers' needs.

## Data Security and Privacy

### Data Security

The data uploaded from each request to Evidence Evaluator is encrypted in transit and at rest. Any documents and questions delivered to Evidence Evaluator are removed from the processing pipeline and all databases within the pipeline after the inference run completes. Any tables, sections, or numerical representations produced of the content in the data are also removed.

### AWS Services

Any content in the documents that are used is scrubbed for sensitive information, sectioned to a handful of short paragraphs, and only those relevant to the given question are retained before being processed in any AWS service. Additionally, the data used within an AWS service is not retained for training or improving the AWS product.

ProcessUnity

## Additional Support

Have additional questions regarding Evidence Evaluator? Please reach out to your account representative to get connected with our product team for additional support.