Open Ethics Transparency Protocol

Exchanging nutrition labels and nutrition tables. How self-disclosure can help to bring safer digital diets?

Nikita Lukianets
Founder @OpenEthicsAI, PocketConfidant
Fellow, the Alliance of Democracies Foundation

Open Ethics

Abstract

The Open Ethics Transparency Protocol (OETP) is an application-level protocol for publishing and accessing ethical Disclosures of IT Products and their Components. The Protocol is describes exchange of information about the ethical "postures", provided in an open and standardized format. The scope of the Protocol covers Disclosures for systems such as Software as a Service (SaaS) Applications, Software Applications, Software Components, Application Programming Interfaces (API), Automated Decision-Making (ADM) systems, and systems using Artificial Intelligence (AI). OETP aims to bring more transparent, predictable, and safe environments for the end-users. The OETP Disclosure Format is an extensible JSON-based format.

Disclosure help us to make informed choices

Ingredients

Allergen labeling

Barcode

Sell-by date information

Caloric and nutritional value information and GDA

Manufacturer information

Identification mark

Mobile tagging

Links

Description	URL
IETF Draft	https://datatracker.ietf.org/doc/draft-lukianets-open-ethics-transparency-protocol/
Project Webpage	https://openethics.ai/oetp/
GitHub Page	https://github.com/OpenEthicsAI/OETP/
The labeling implementation to perform the initial Disclosure	https://openethics.ai/label/
The Open Ethics Manifesto	https://openethics.ai/manifesto/

Allowing informed choice

What if we bring labels to digital products?

Your choices on this panel will update the other panels on this page.

Allow adaptations of your work to be shared?

(?)

- ○ Yes
- ◉ No
- ○ Yes, as long as others share alike

Allow commercial uses of your work?

(?)

- ◉ Yes
- ○ No

Selected License

Attribution-NoDerivatives 4.0 International

This is not a Free Culture License.

Help others attribute you!

This part is optional, but filling it out will add machine-readable metadata to the suggested HTML!

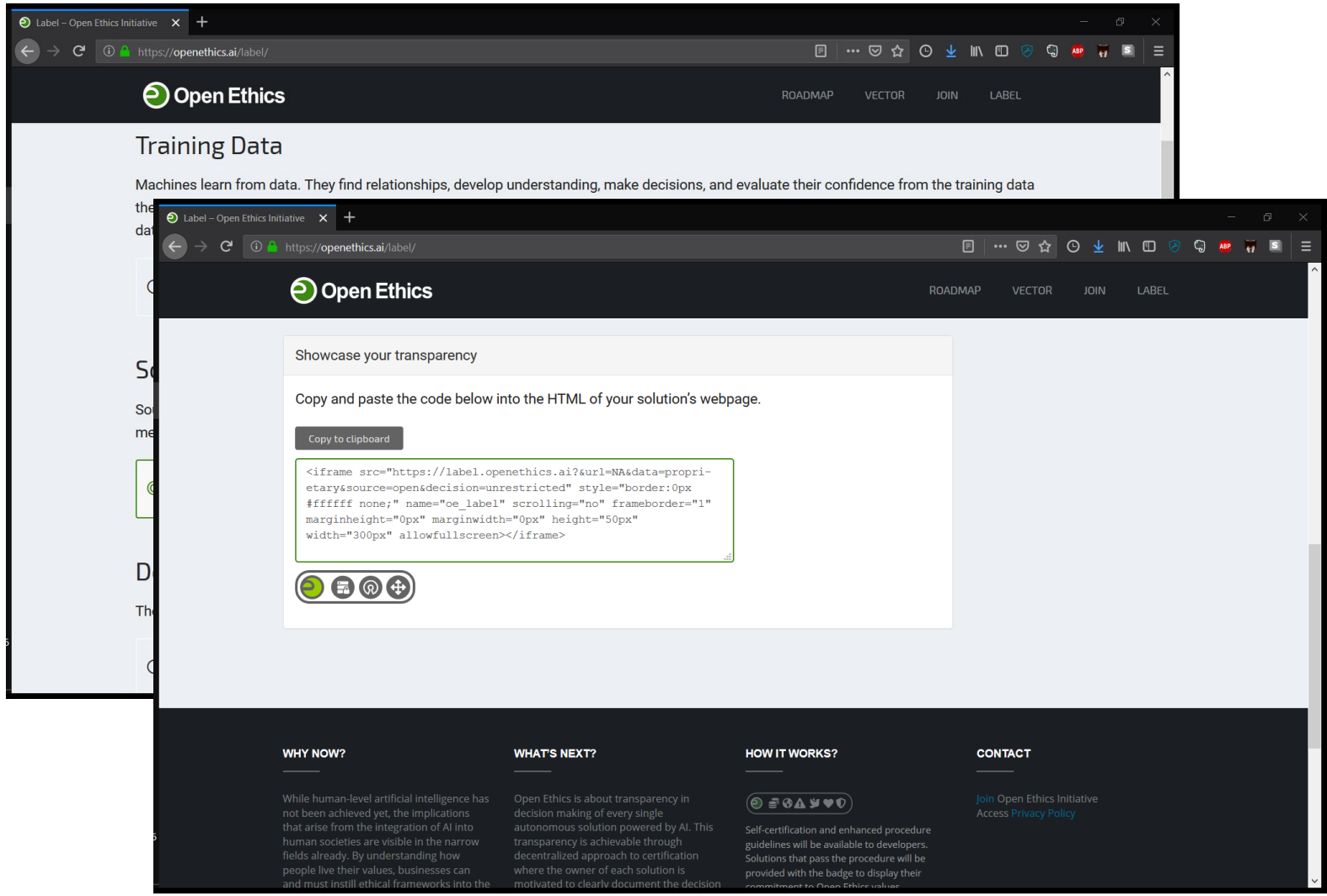
(?)

Have a web page?

This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0

Extending The Open Ethics Label

A trust-building platform.
Self-disclosure
for AI systems.



https://openethics.ai/label

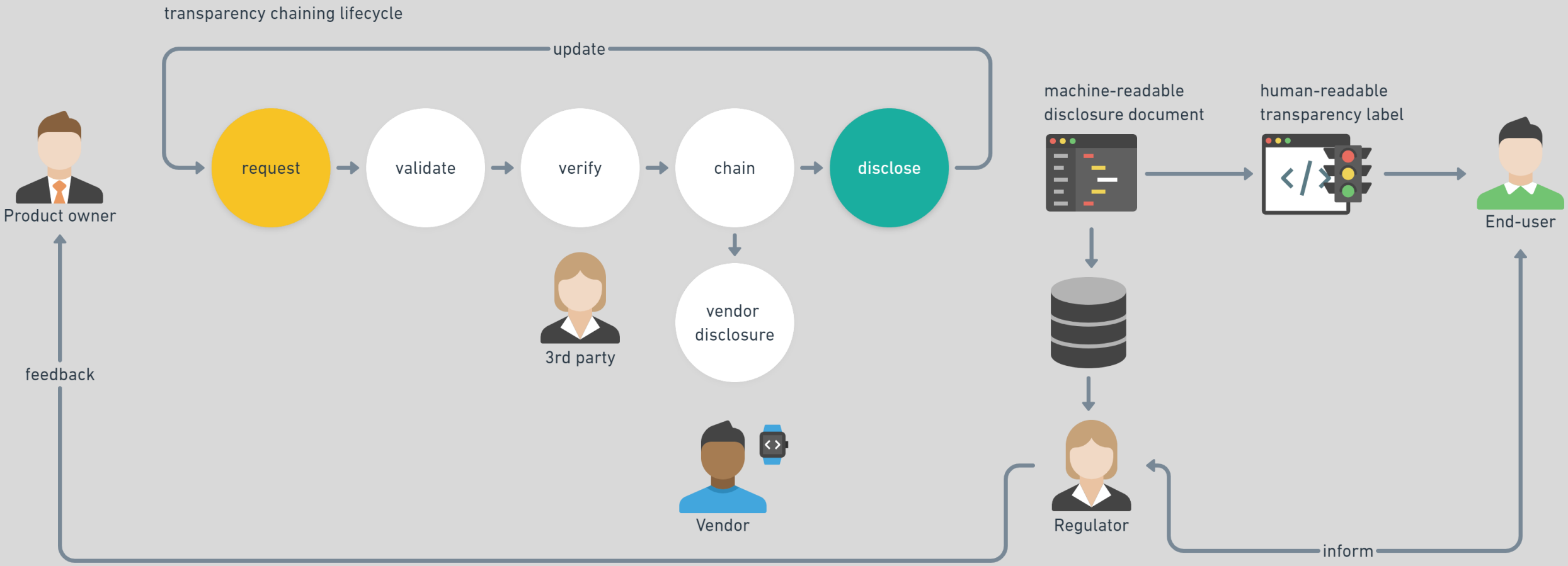
Open Ethics Transparency Protocol

```
{
    "schema": {
        "name": "Open Ethics Transparency Protocol",
        "version": "0.9.2 RFC"
    },
    "data": {
        "availability": "public",
        "batch": 14032507,
        "sources": [ ···
        ]

    },
    "algorithms": [ ···
    ],
    "decision-space": { ···
    },
    "validation": {
        "critical": true,
        "methods": [
            {
                "name": "HITL",
                "objective": "To confirm the diagnosis by
                human doctors before presenting to the
                patient",
                "type": "council"
            }
```

Display “ethical” posture in both human and machine-readable ways.

Open Ethics Transparency Protocol



https://openethics.ai/oetp/

Where help is needed?

1. How to better manage the identity of the components?
2. Identity of the validation providers?
3. How to extend the disclosure format?
4. What is the best way to generate composite disclosure?
5. Help to create use-cases
6. Help to standardize Privacy Disclosure and PII data-collection practices
7. Enhancing Label accessibility (ex. with ARIA W3C Recommendation)

How could you help?

- Where? Probably via the new WG
- You can contribute by submitting issues and by making pull requests to https://github.com/OpenEthicsAI/OETP/
- Via Discord? https://openethics.ai/discord/

Nikita Lukianets

n.lukianets@openethics.ai

Appendix

Implementation

Generate Open Ethics Label – O

https://openethics.ai/label/generate/

Open Ethics

MANIFESTO CANVAS EVENTS PROJECTS ABOUT JOIN

Human-in-the-Loop component

AI swarm + Human

Humans confirm

HITL design objective

Personal Data Record

Account ID

Source of the data record

permanei

Purpose of the data record processing

Consent obtained directly

Identifying record

Source

Privacy Policy

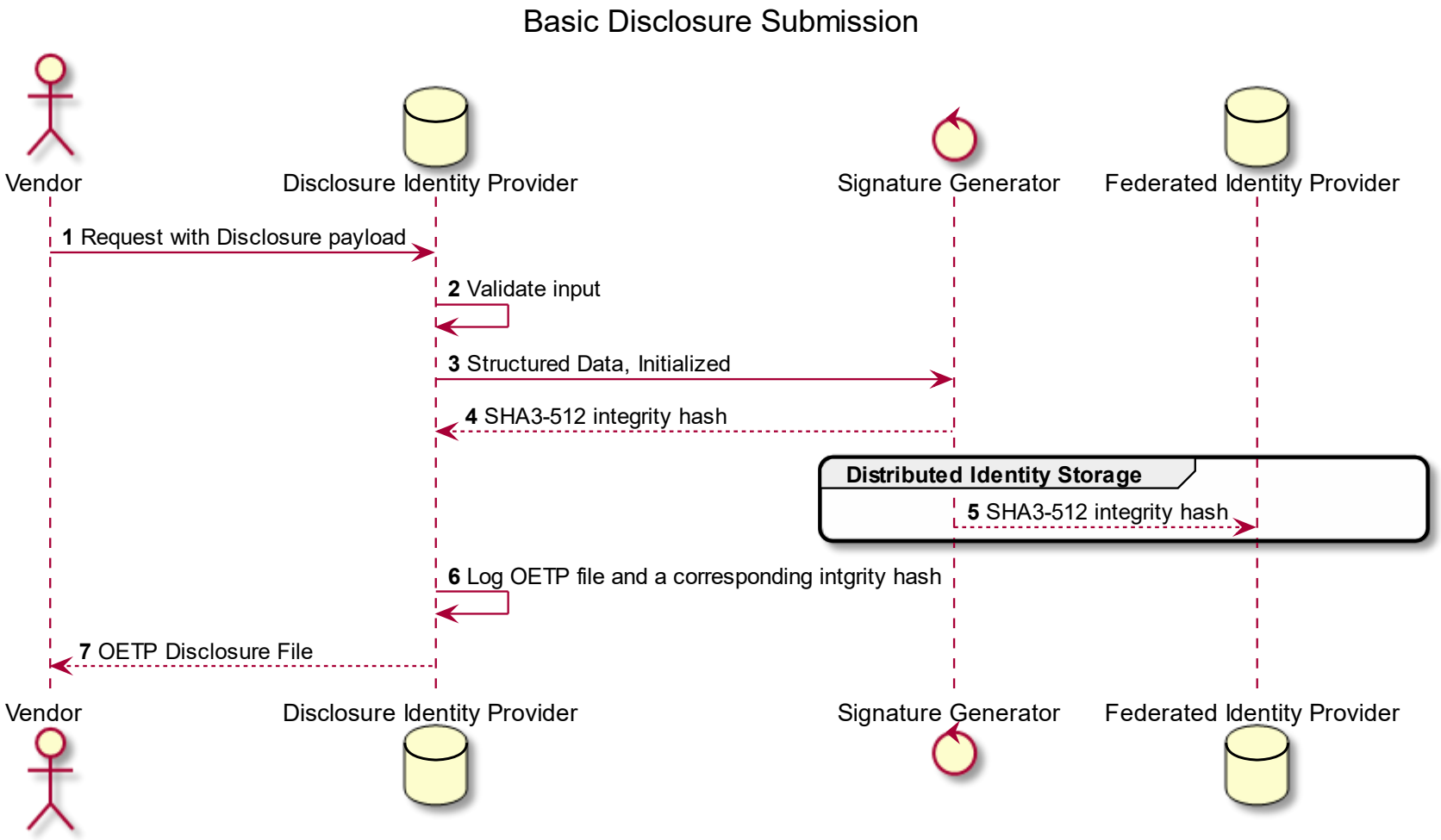
https://example.com/privacy

Add comments

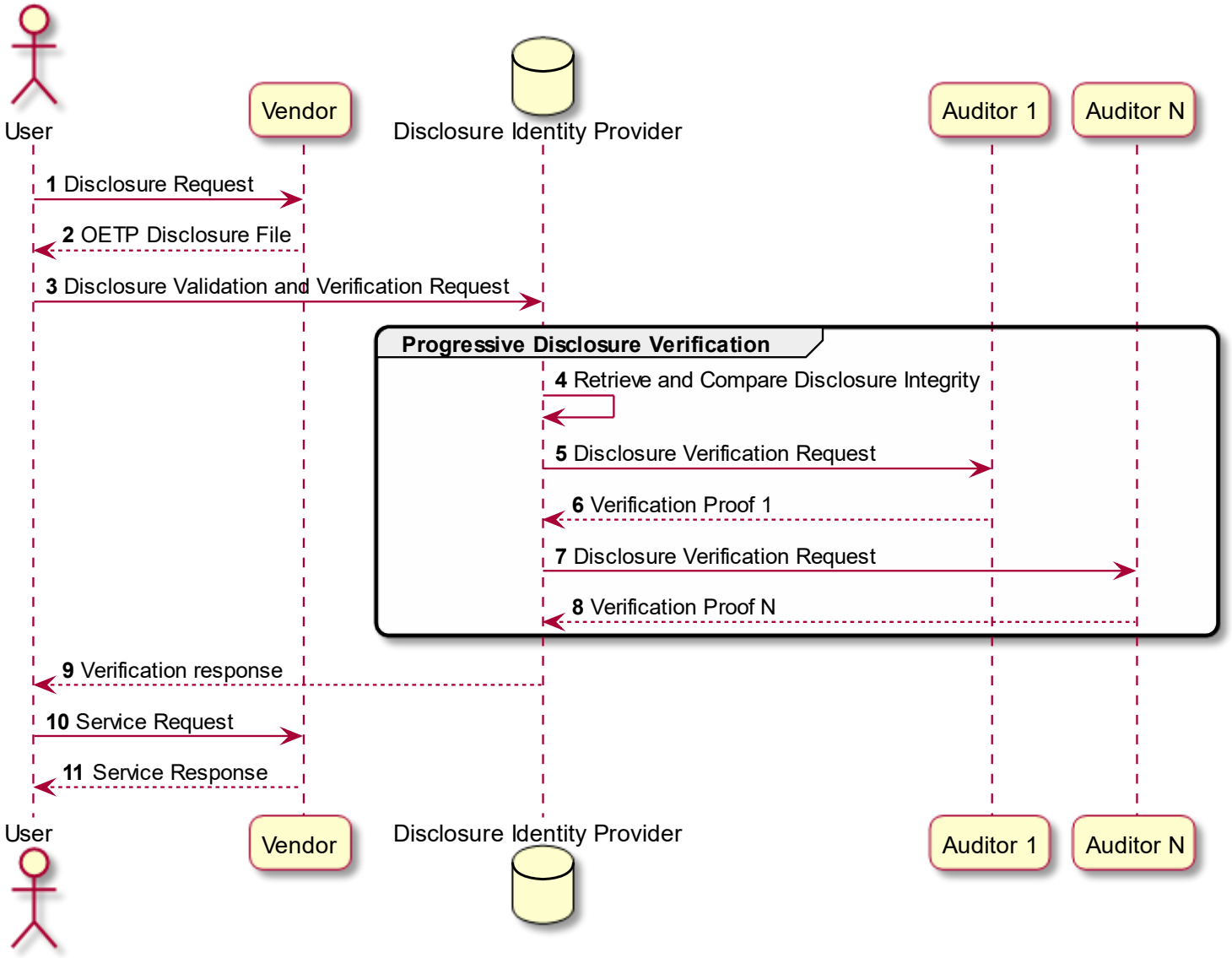
+ Personal Data Record

Display transparency

This is the preview. After submission of the form, you will receive the code to paste Open Ethics Label on your website.


Basic Disclosure Submission
Vendor
Disclosure Identity Provider
Signature Generator
Federated Identity Provider
1 Request with Disclosure payload
2 Validate input
3 Structured Data, Initialized
4 SHA3-512 integrity hash
Distributed Identity Storage
5 SHA3-512 integrity hash
6 Log OETP file and a corresponding intgrity hash
7 OETP Disclosure File
Vendor
Disclosure Identity Provider
Signature Generator
Federated Identity Provider

Progressive Verification with multiple Auditors



Integrity

Open Ethics has generated the following SHA3-512 hash for the snapshot of submitted data.

```
093b41c91413496d71c839105f0c2c882d3d34ded7456103cdd84e0a4c4aa6ac6ca35141865cd00c99bb2adaa1240ff306325dd80555e2fe9c7f9d7f0b2752cc
```

Open Ethics Transparency Protocol

Save the code below as a text file named *oetp.json* in the root of your product's website. It will allow integrity validation, as well as machine readable scenarios for AI transparency management.

Download oetp.json

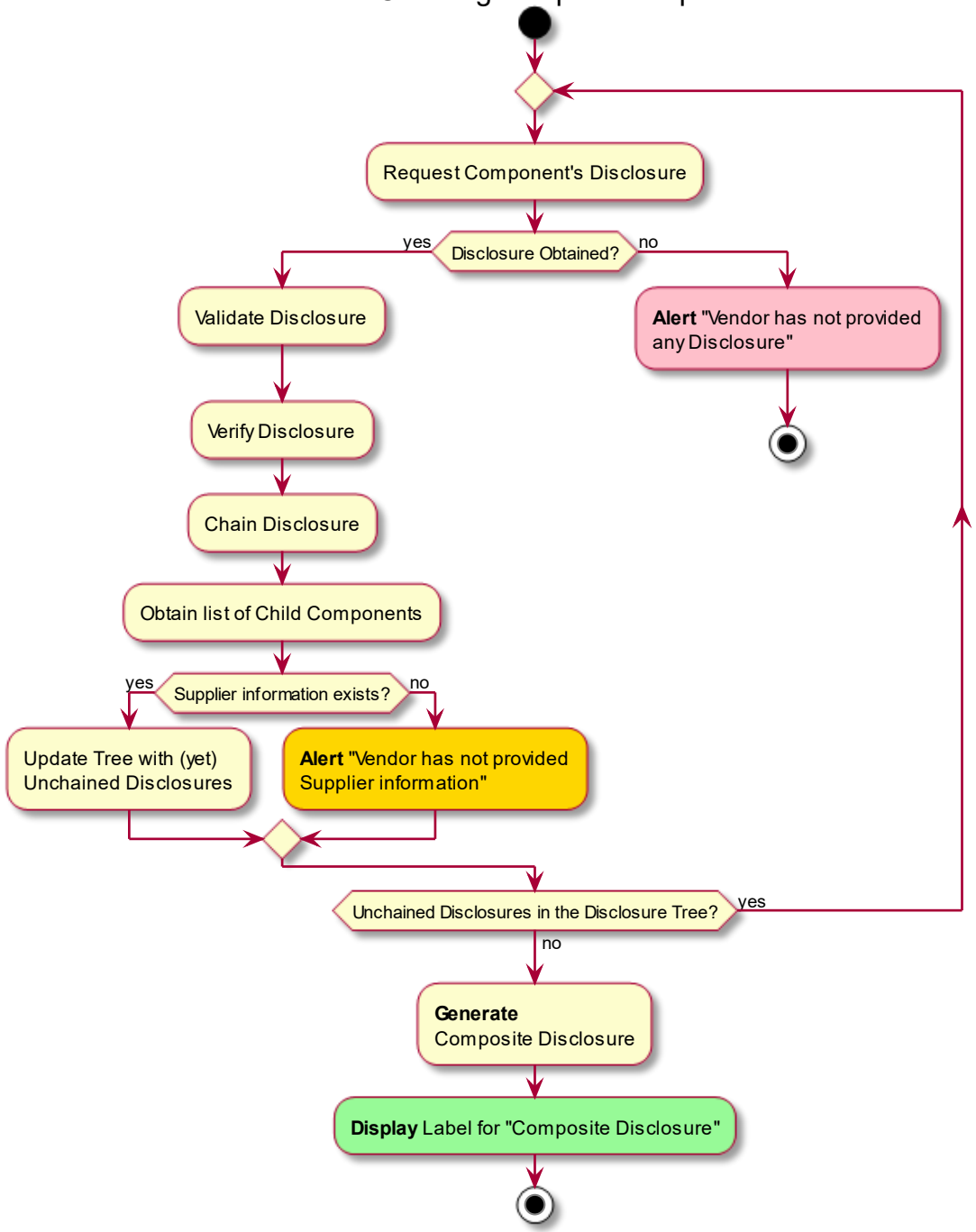

```
{
    "schema": {
        "name": "Open Ethics Transparency Protocol",
        "version": "0.9.3 RFC",
        "integrity": "093b41c91413496d71c839105f0c2c882d3d34ded7456103cdd84e0a4c4aa6ac6ca35141865cd00c99bb2adaa1240ff306325dd80555e2fe9c7f9d7f0b2752cc"
    },
    "snapshot": {
        "product": {
            "url": "ietftest.com",
            "description": ""
```


Display transparency

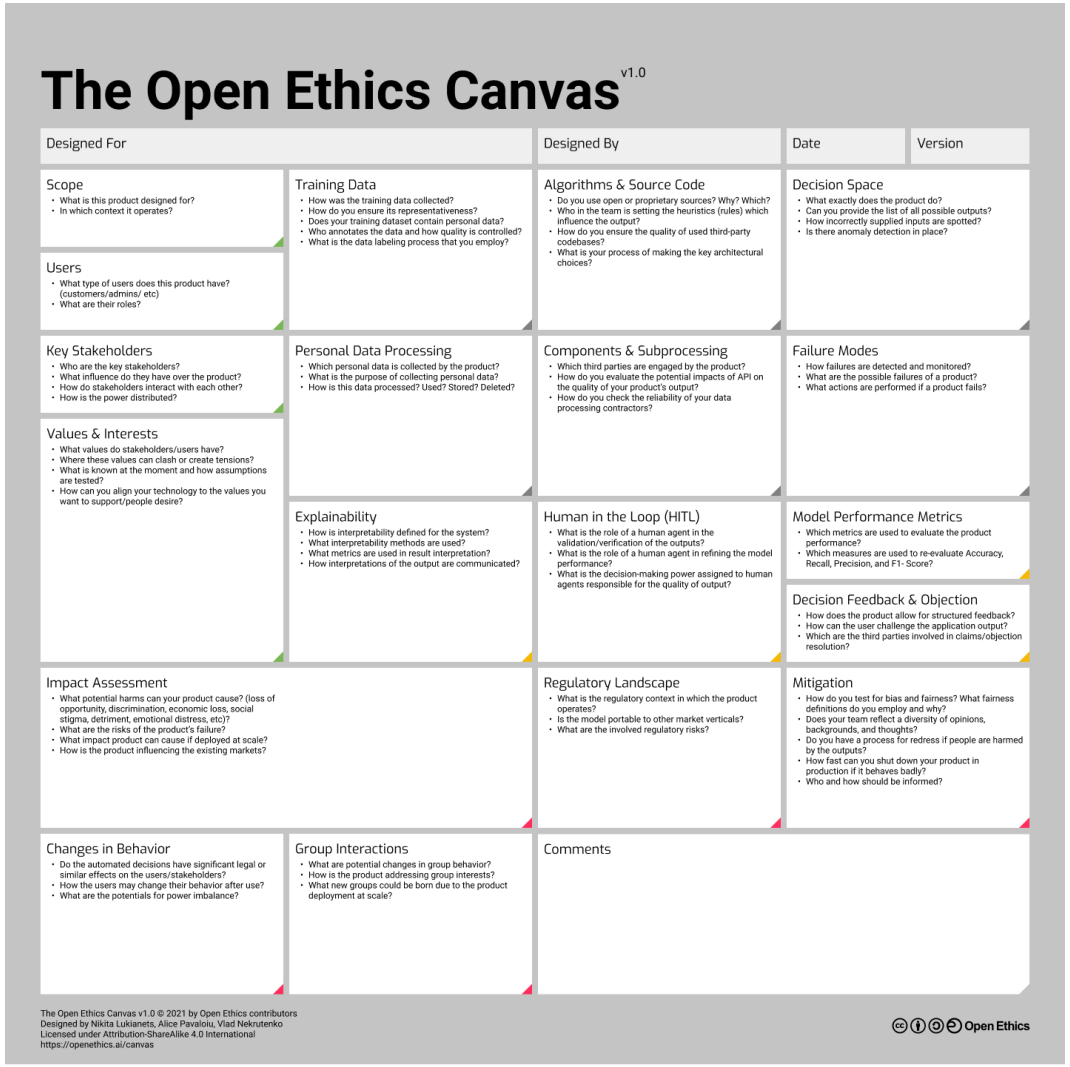
Save and paste the Open Ethics Label HTML code into the HTML of your product's webpage.

Download oel.html

Disclosure Chaining: Request-Response



Talk #ethics, include your #tech team



The Open Ethics Canvas v1.0

Designed For		Designed By	Date	Version

Scope
- What is this product designed for?
- In which context it operates?

Users
- What type of users does this product have? (customers/admins/ etc)
- What are their roles?

Training Data
- How was the training data collected?
- How do you ensure its representativeness?
- Does your training dataset contain personal data?
- Who annotates the data and how quality is controlled?
- What is the data labeling process that you employ?

Algorithms & Source Code
- Do you use open or proprietary sources? Why? Which?
- Who in the team is setting the heuristics (rules) which influence the output?
- How do you ensure the quality of used third-party codebases?
- What is your process of making the key architectural choices?

Decision Space
- What exactly does the product do?
- Can you provide the list of all possible outputs?
- How incorrectly supplied inputs are spotted?
- Is there anomaly detection in place?

Key Stakeholders
- Who are the key stakeholders?
- What influence do they have over the product?
- How do stakeholders interact with each other?
- How is the power distributed?

Personal Data Processing
- Which personal data is collected by the product?
- What is the purpose of collecting personal data?
- How is this data processed? Used? Stored? Deleted?

Components & Subprocessing
- Which third parties are engaged by the product?
- How do you evaluate the potential impacts of API on the quality of your product's output?
- How do you check the reliability of your data processing contractors?

Failure Modes
- How failures are detected and monitored?
- What are the possible failures of a product?
- What actions are performed if a product fails?

Values & Interests
- What values do stakeholders/users have?
- Where these values can clash or create tensions?
- What is known at the moment and how assumptions are tested?
- How can you align your technology to the values you want to support/people desire?

Explainability
- How is interpretability defined for the system?
- What interpretability methods are used?
- What metrics are used in result interpretation?
- How interpretations of the output are communicated?

Human in the Loop (HITL)
- What is the role of a human agent in the validation/verification of the outputs?
- What is the role of a human agent in refining the model performance?
- What is the decision-making power assigned to human agents responsible for the quality of output?

Model Performance Metrics
- Which metrics are used to evaluate the product performance?
- Which measures are used to re-evaluate Accuracy, Recall, Precision, and F1- Score?

Decision Feedback & Objection
- How does the product allow for structured feedback?
- How can the user challenge the application output?
- Which are the third parties involved in claims/objection resolution?

Impact Assessment
- What potential harms can your product cause? (loss of opportunity, discrimination, economic loss, social stigma, detriment, emotional distress, etc)?
- What are the risks of the product's failure?
- What impact product can cause if deployed at scale?
- How is the product influencing the existing markets?

Regulatory Landscape
- What is the regulatory context in which the product operates?
- Is the model portable to other market verticals?
- What are the involved regulatory risks?

Mitigation
- How do you test for bias and fairness? What fairness definitions do you employ and why?
- Does your team reflect a diversity of opinions, backgrounds, and thoughts?
- Do you have a process for redress if people are harmed by the outputs?
- How fast can you shut down your product in production if it behaves badly?
- Who and how should be informed?

Changes in Behavior
- Do the automated decisions have significant legal or similar effects on the users/stakeholders?
- How the users may change their behavior after use?
- What are the potentials for power imbalance?

Group Interactions
- What are potential changes in group behavior?
- How is the product addressing group interests?
- What new groups could be born due to the product deployment at scale?

Comments

The Open Ethics Canvas v1.0 © 2021 by Open Ethics contributors
Designed by Nikita Lukianets, Alice Pavaloiu, Vlad Nekrutenko
Licensed under Attribution-ShareAlike 4.0 International
https://openethics.ai/canvas

Open Ethics

Download and use it for free

https://openethics.ai/canvas/

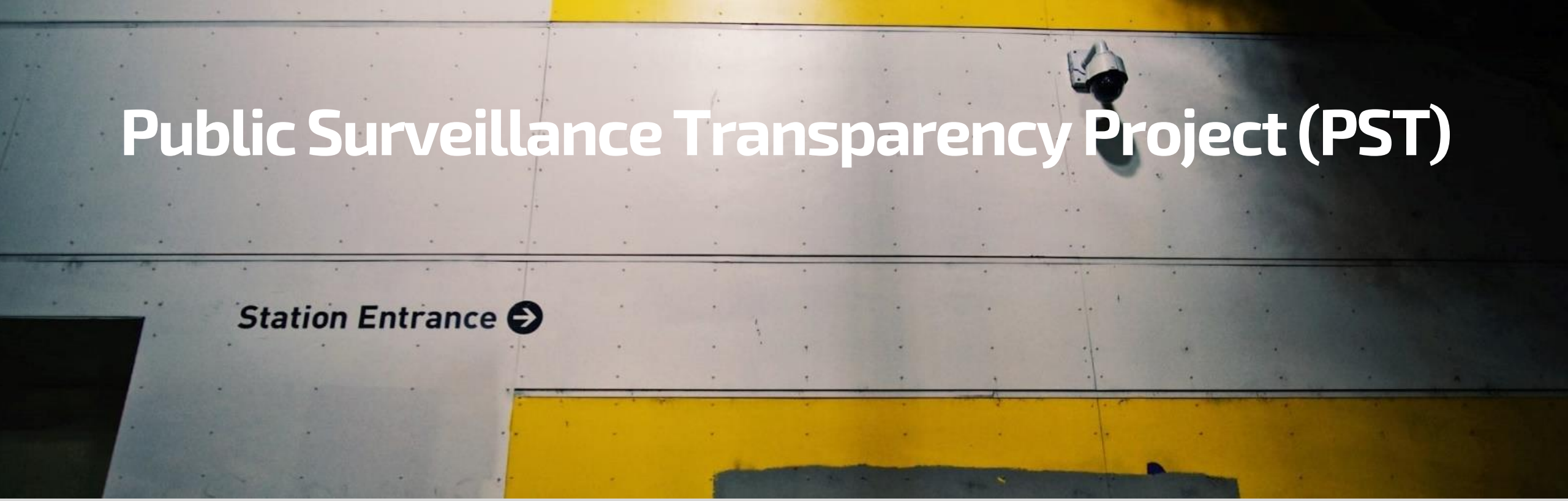
Motivation

- **Informed consumer choices** : End-users able to make informed choices based on their own ethical preferences and ADM disclosure.
- **Industrial scale monitoring** : Discovery of best and worst practices within market verticals, technology stacks, and product value offerings.
- **Legally-agnostic guidelines** : Suggestions for developers and product-owners, formulated in factual language, which are legally-agnostic and could be easily transformed into product requirements and safeguards.
- **Iterative improvement** : Digital products, specifically, the ones powered by artificial intelligence could receive a nearly real-time feedback on how their performance and ethical posture could be improved to cover security, privacy, diversity, fairness, power-balance, non-discrimination, and other requirements.
- **Labeling and certification** : Mapping to existing and future regulatory initiatives and standards.

Use-case

example

Public Surveillance Transparency Project (PST)



Citizen in the public space → Scan QR code on the sticker → Read the disclosure → Provide feedback

1. Visual transparency labels for citizens
2. Machine-readable disclosure
3. Disclosure as Open Data [governance/market]

Market creation: Location-based services

Open Data

McKinsey reference: **5.2 bn EUR**

ODI estimates: **1.2 bn EUR**

LBS

in Europe alone estimates (2025): **10 bn EUR**

https://theodi.org/article/the-value-of-open-data-for-the-private-sector/

https://www.cio.com/article/3626390/6-ways-to-harness-the-value-of-open-data.html

Implications

AI alignment

Safety

Failure Transparency

Judicial Transparency

Responsibility

Value Alignment

Personal Privacy

Liberty and Privacy

Shared Benefit

Human Control

Non-subversion

Compassion

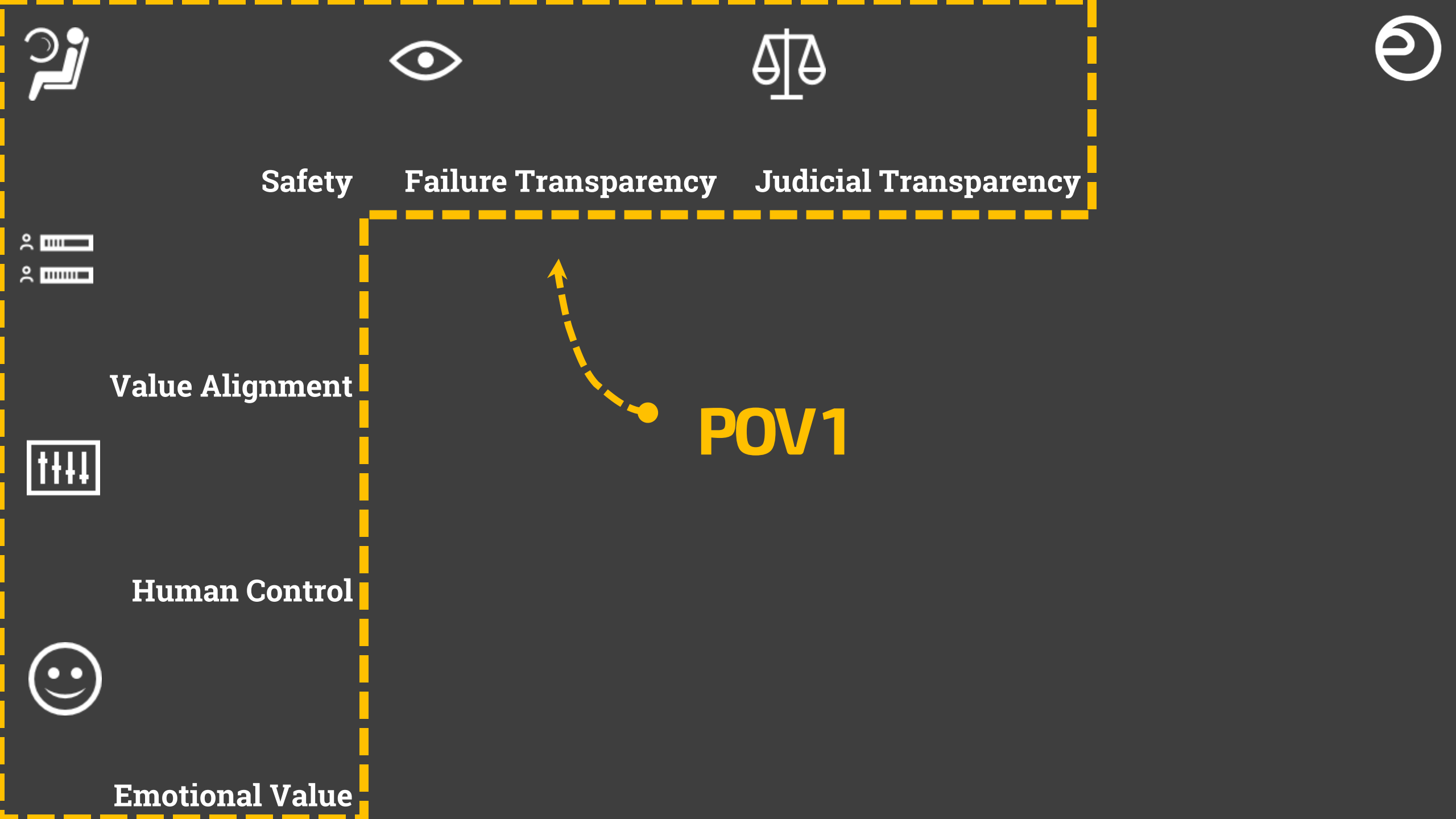
Shared Prosperity

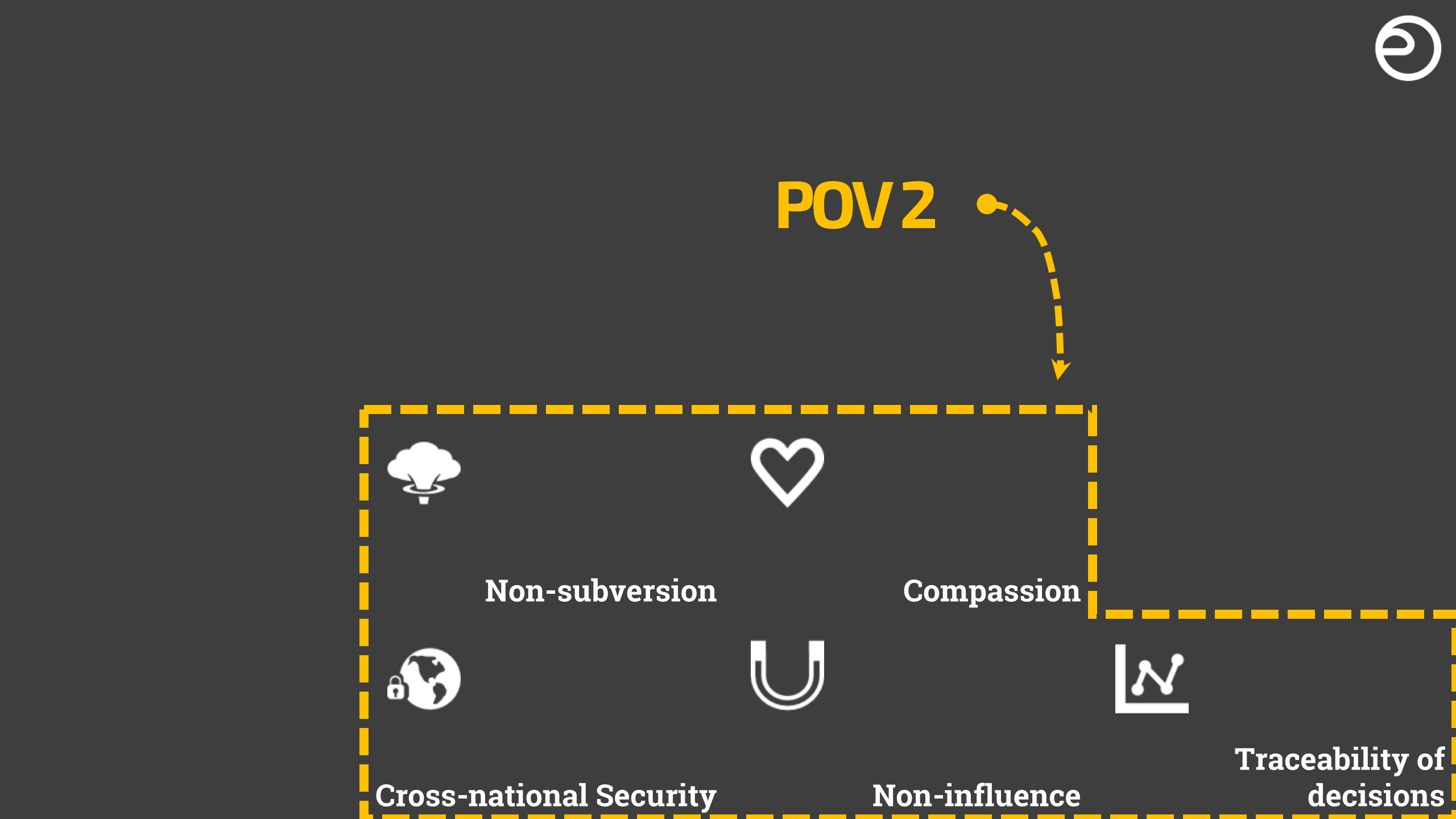
Emotional Value

Cross-national Security

Non-influence

Traceability of decisions


Safety
Failure Transparency
Judicial Transparency
Value Alignment
POV1
Human Control
Emotional Value


POV 2
Non-subversion
Compassion
Cross-national Security
Non-influence
Traceability of decisions

Alternative 1

Safety

Failure Transparency

Judicial Transparency

Responsibility

Value Alignment

Personal Privacy

Liberty and Privacy

Shared Benefit

Human Control

Non-subversion

Compassion

Shared Prosperity

Emotional Value

Cross-national Security

Non-influence

Traceability of decisions

Alternative 2

Safety

Failure Transparency

Judicial Transparency

Responsibility

Value Alignment

Personal Privacy

Liberty and Privacy

Shared Benefit

Human Control

Non-subversion

Compassion

Shared Prosperity

Emotional Value

Cross-national Security

Non-influence

Traceability of decisions

Alternative 3

Safety

Failure Transparency

Judicial Transparency

Responsibility

Value Alignment

Personal Privacy

Liberty and Privacy

Shared Benefit

Human Control

Non-subversion

Compassion

Shared Prosperity

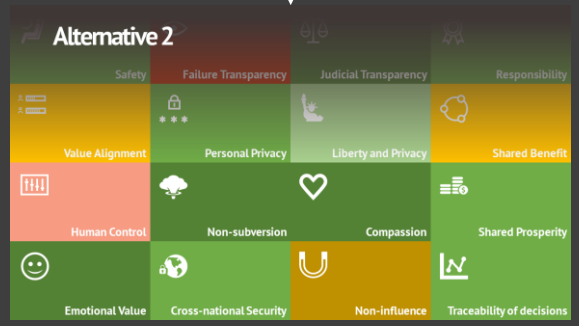
Emotional Value

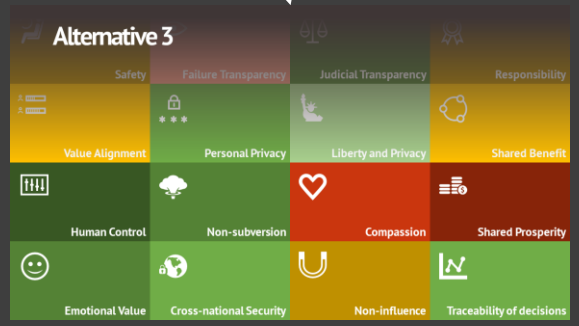
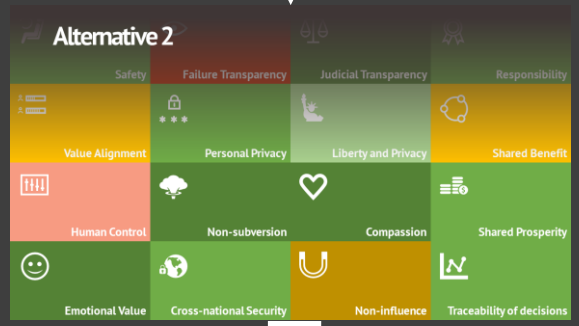
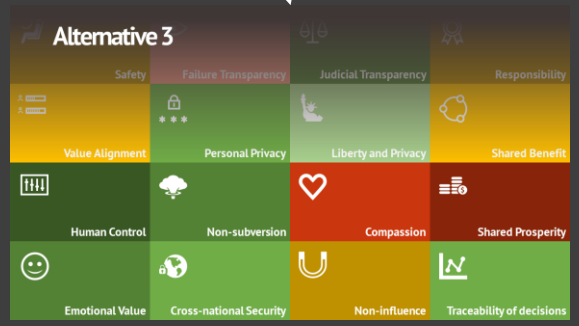
Cross-national Security

Non-influence

Traceability of decisions

Alternatives





Aligned choice