

AI Advances, Responsibilities, and Governance

Eric Horvitz
Chief Scientific Officer

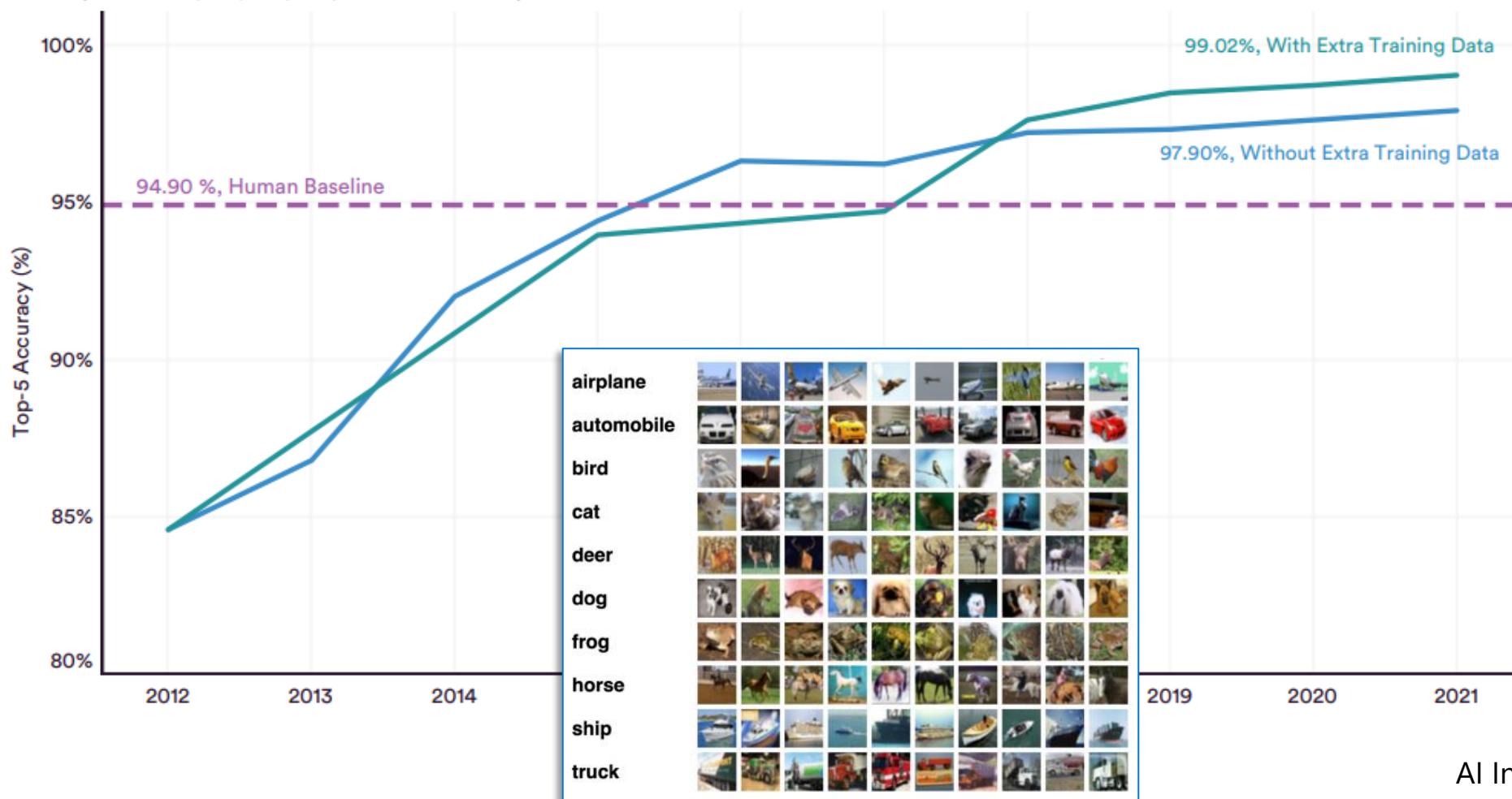
Microsoft

Digital Humanism Series
June 28, 2022

Fast-paced advances

IMAGENET CHALLENGE: TOP-5 ACCURACY

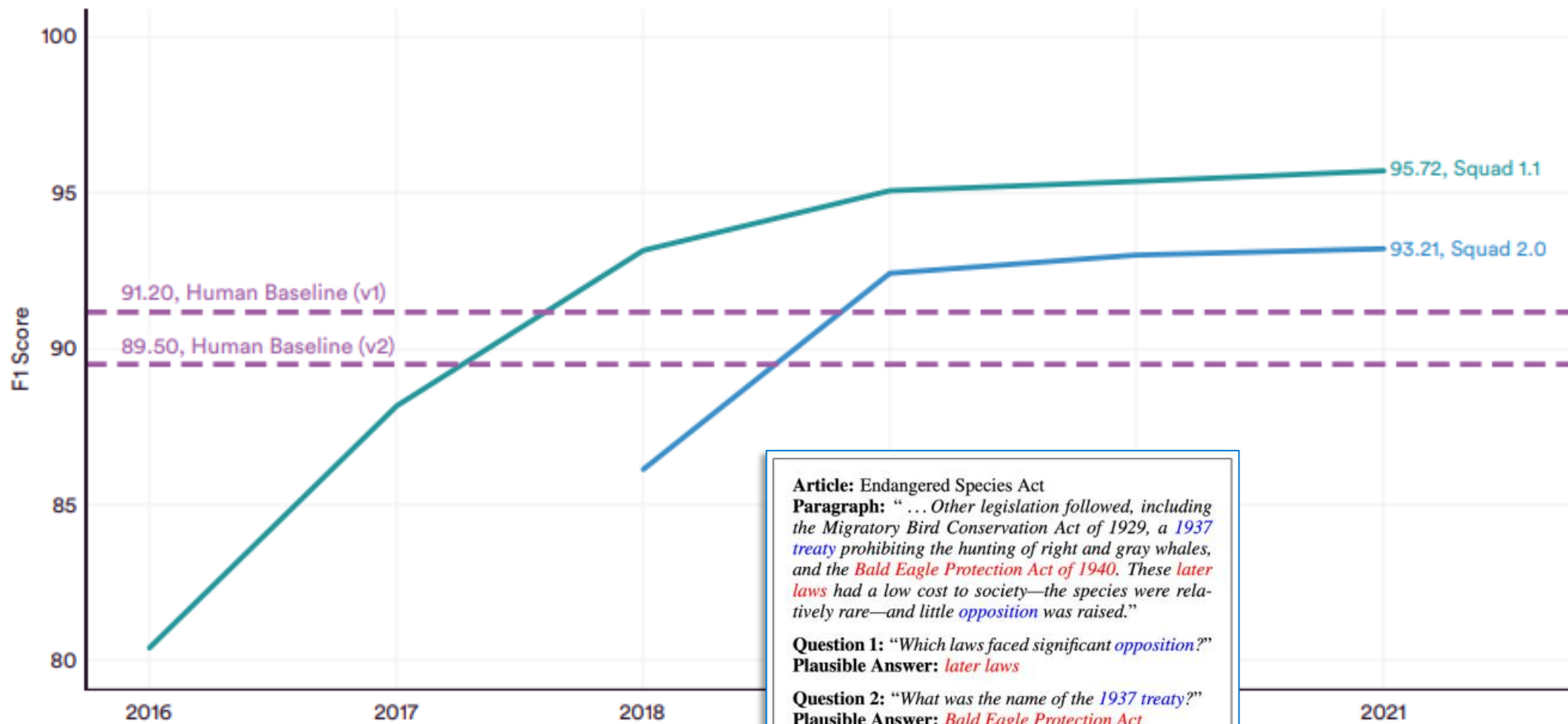
Source: Papers with Code, 2021; arXiv, 2021 | Chart: 2022 AI Index Report



Fast-paced advances

SQUAD 1.1 and SQUAD 2.0: F1 SCORE

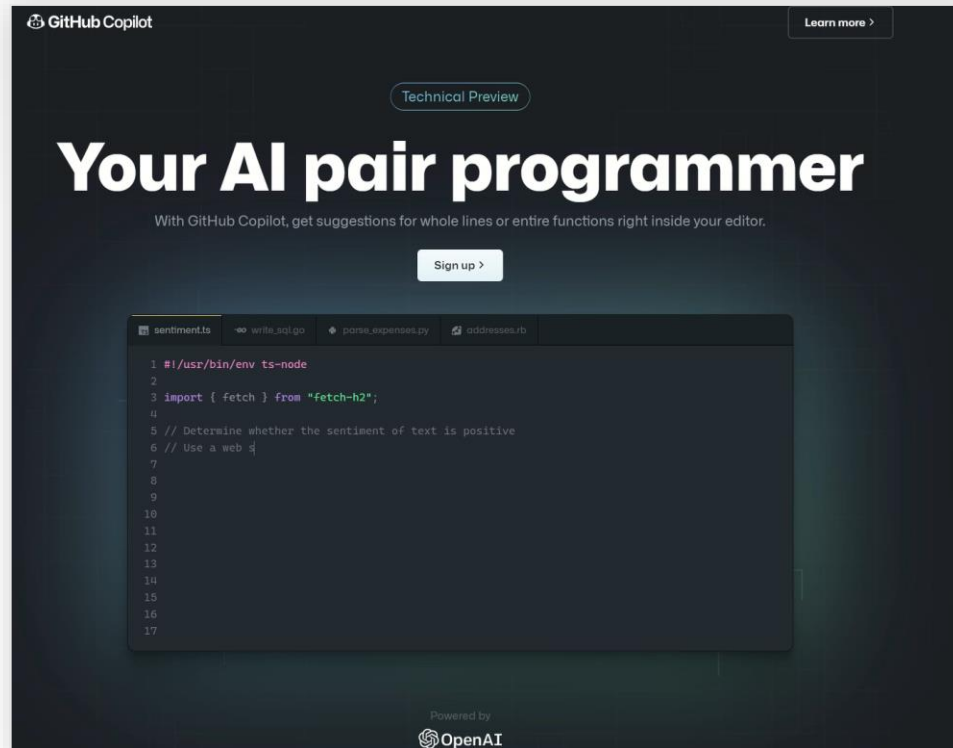
Source: SQuAD 1.1 and SQuAD 2.0, 2021 | Chart: 2022 AI Index Report



Disruptive Breakthroughs

Industry, academia,
government labs
pushing at frontiers

- Computer software generation
- Protein structure inference & protein engineering
- Fabrication of realistic content



The image shows the GitHub Copilot website. At the top left is the GitHub Copilot logo, and at the top right is a "Learn more >" link. Below the logo is a "Technical Preview" badge. The main heading is "Your AI pair programmer" in large white text. Underneath is the subtext: "With GitHub Copilot, get suggestions for whole lines or entire functions right inside your editor." Below this is a "Sign up >" button. The central part of the page features a dark-themed code editor window with several tabs: "sentiment.ts", "write_sqli.go", "parse_expenses.py", and "addresses.rb". The "sentiment.ts" tab is active and shows the following code:

```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web d
7
8
9
10
11
12
13
14
15
16
17
```

At the bottom of the page, it says "Powered by OpenAI" with the OpenAI logo.



Scientists are using software to design new biomolecules that treat cancer and block viral infection. (Photo by Ian Haydon, UW Medicine Institute for Protein Design.)

Governance

Pace of AI advancement → Governance scope & terrain



Corporate self-regulation with sharing of best practices
(Companies, Partnership on AI, etc.)



Professional societies, standards bodies, and safety organizations
(ISO, IEEE, etc.)

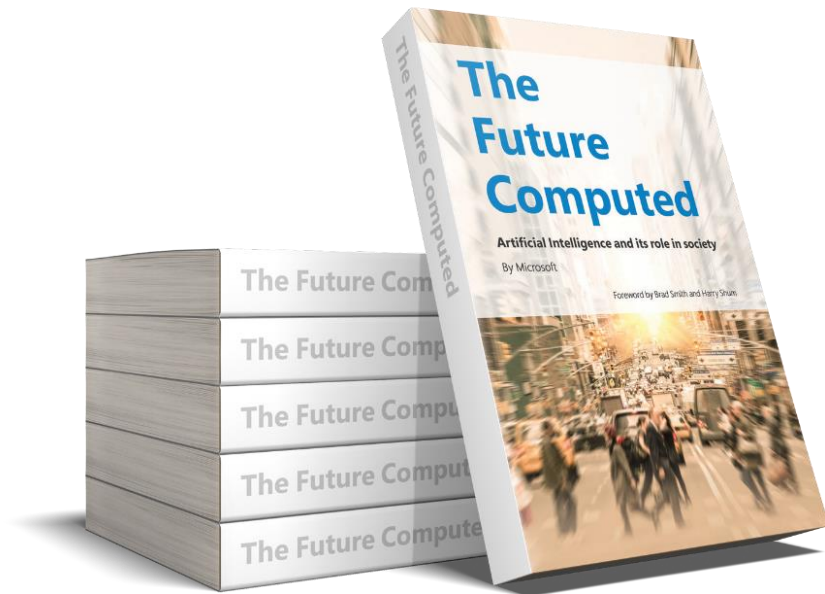


Federal and state government legislation and regulation
(FDA, FTC, CPSC, NHTSA, Uniform Law Commission, etc.)



Multinational understandings, coordination, and treaties
(OECD, UN, US Exec, State, Defense, NATO, US-China, UN, etc.)

Values



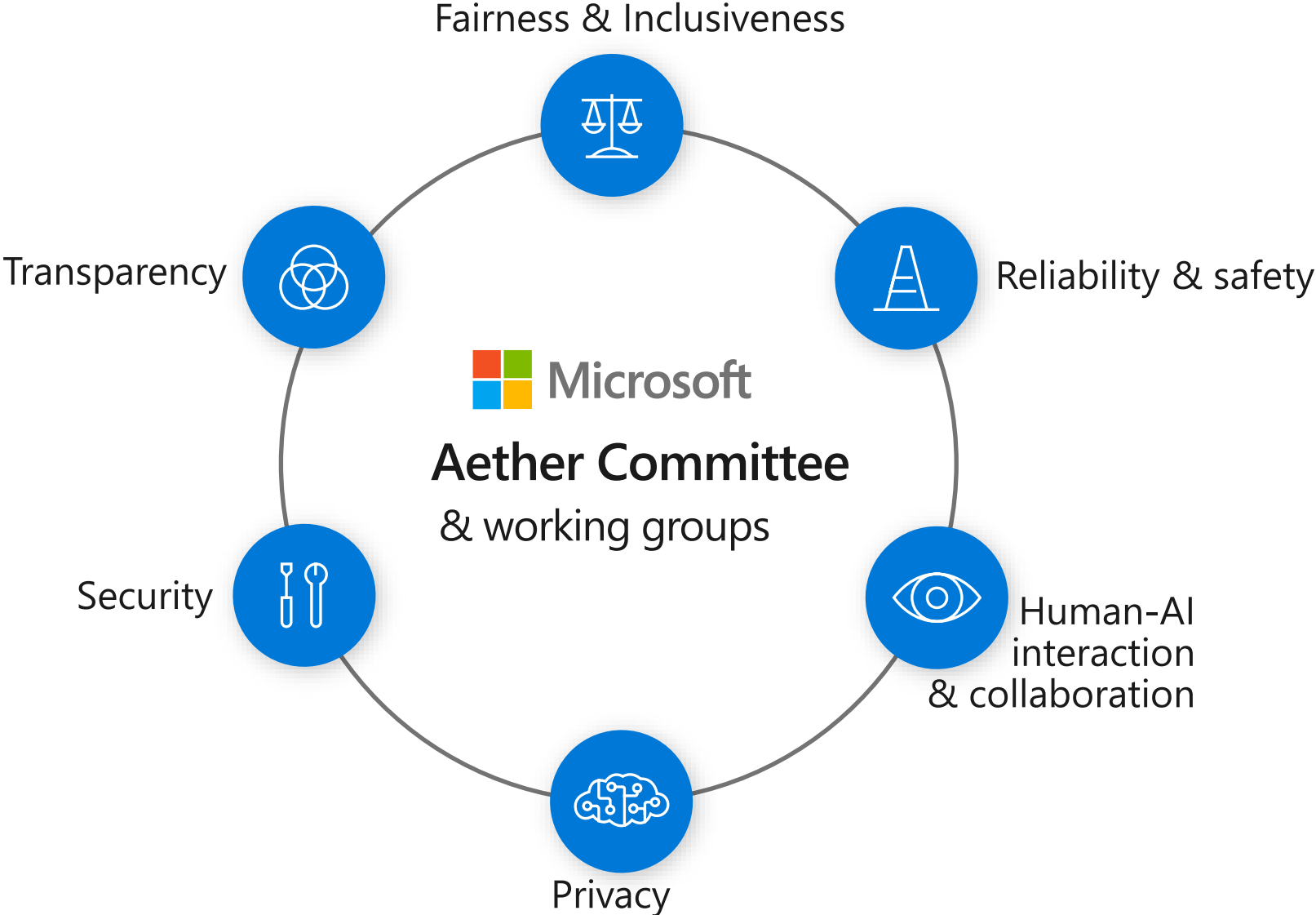
- Fairness
- Reliability
- Privacy & security
- Inclusiveness
- Transparency
- Accountability

Governance



Aether Committee
& working groups

Governance

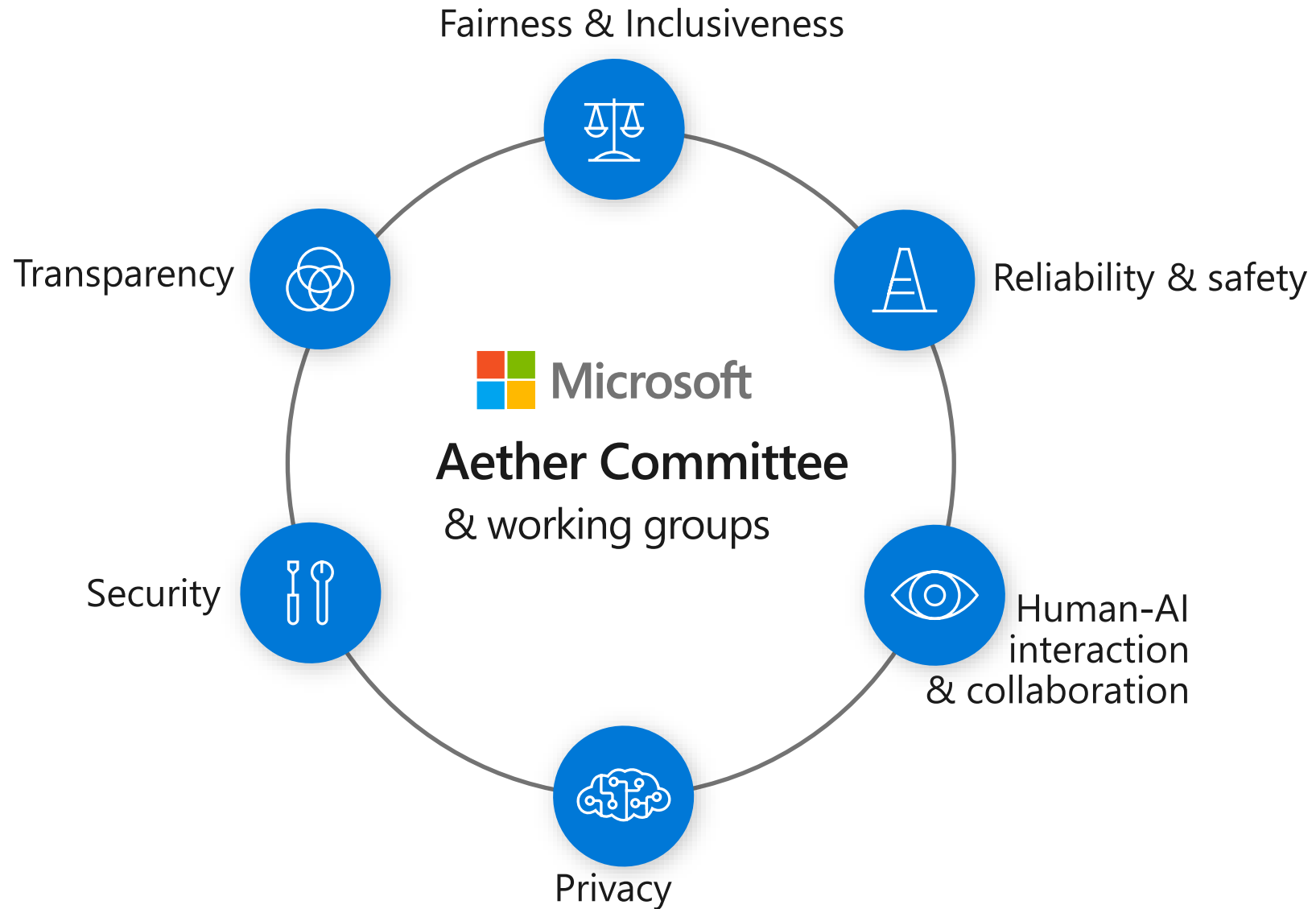


Aether Committee

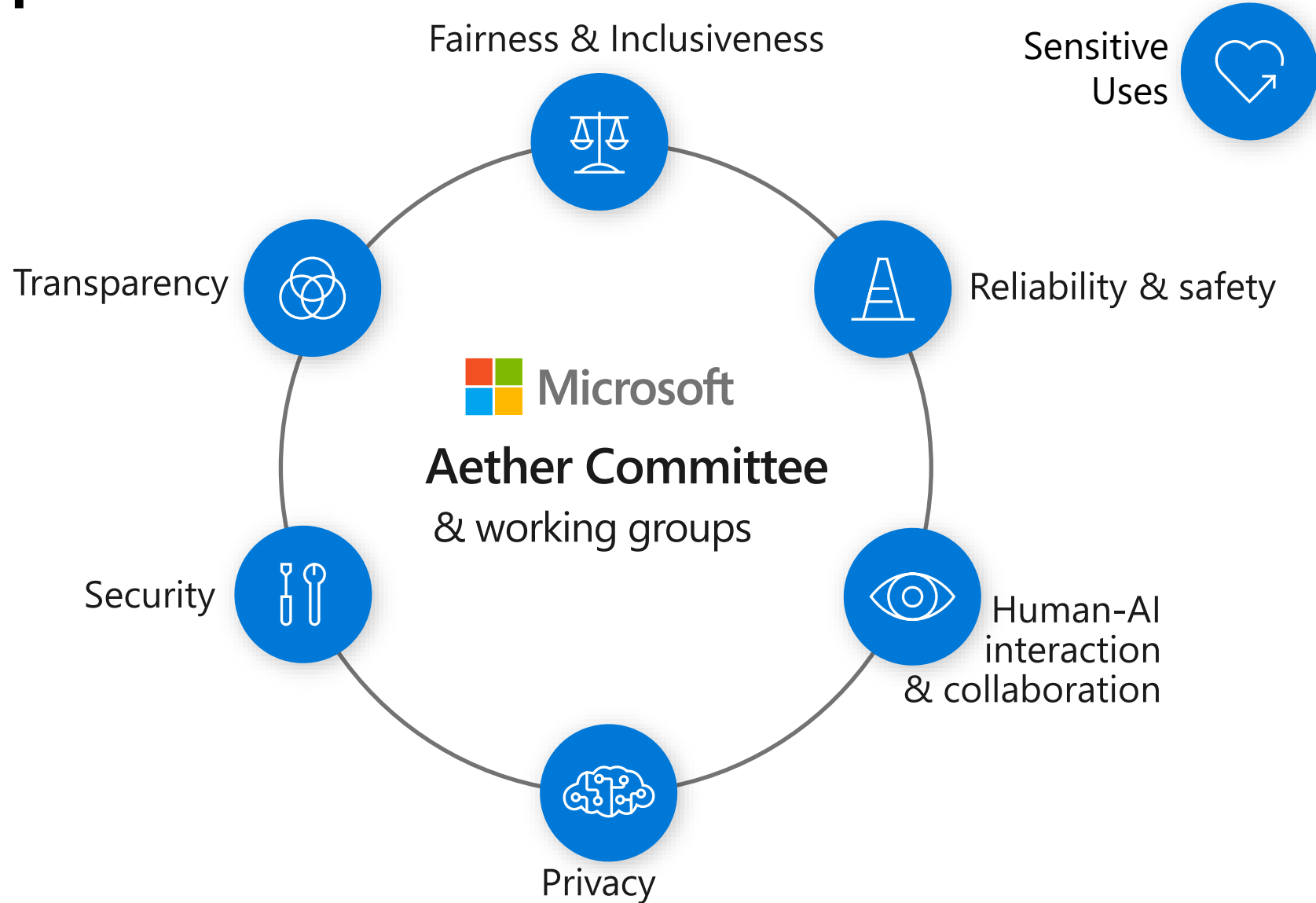
“As we make technological progress, we need to ensure that we are doing so responsibly. To this end, [we] have established Microsoft’s AI an Ethics in Engineering and Research (AETHER) Committee , bringing together senior leaders from across the company to focus on proactive formulation of internal policies and how to respond to specific issues in a responsible way.”

-Satya Nadella

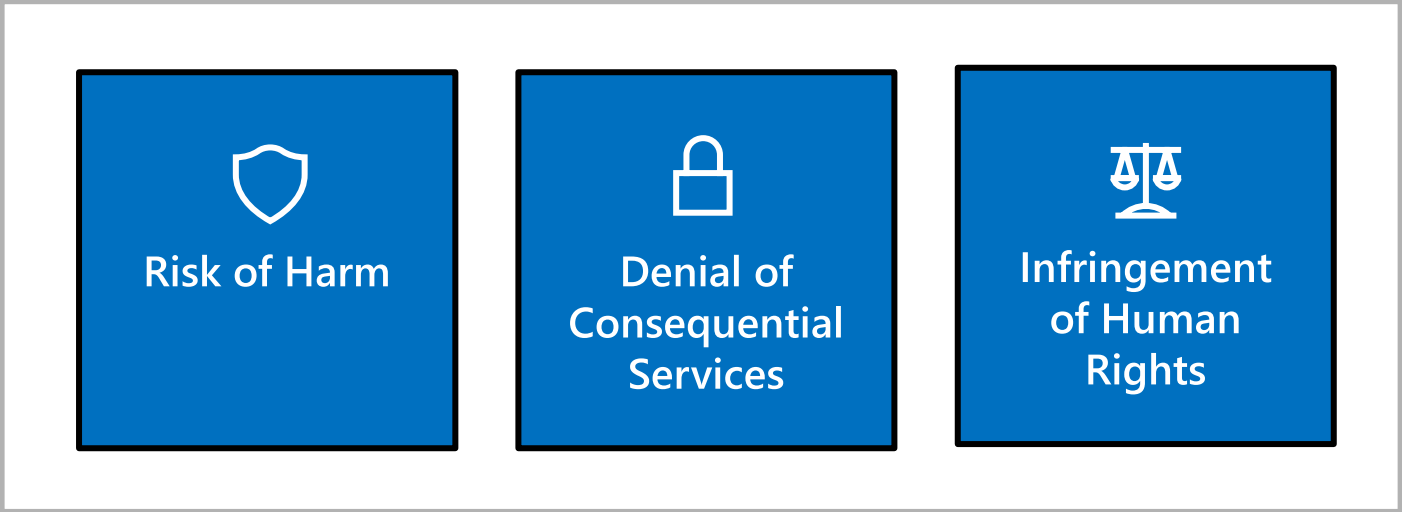
Approach



Approach



Sensitive Uses of AI



I. Risk of physical or psychological injury

The use or misuse of the AI system could result in significant physical or psychological injury to an individual.

II. Consequential impact on legal position or life opportunities

The use or misuse of the AI system could affect an individual's:

- Legal status, such as whether an individual is recognized as a minor, adult, parent, guardian, or person with a disability, as well as their marital, immigration, and citizenship status.
- Legal rights, particularly in the context of the criminal justice system.
- Access to credit, education, employment, healthcare, housing, insurance, and social welfare benefits, services, or opportunities, or the terms on which they are provided.

III. Threat to human rights

The use or misuse of the AI system could restrict, infringe upon, or undermine the ability to realize an individual's human rights.

- Human dignity and equality in enjoyment of rights.
- Freedom from discrimination.
- Life, liberty, and security of a person.
- Equal protection of the law and criminal justice systems.
- Protection against arbitrary interference with privacy.
- Freedom of movement.
- Freedom of thought, conscience, and religion.
- Freedom of opinion and expression.
- Peaceful assembly and association.

Sensitive Uses of AI



^ Sensitive Uses Overview and Categories

All employees are empowered to report sensitive uses of AI and seek guidance.

Some potential uses of AI systems are particularly sensitive and impactful on individuals and

[Report Sensitive Uses](#)

[Frequently Asked Questions](#)

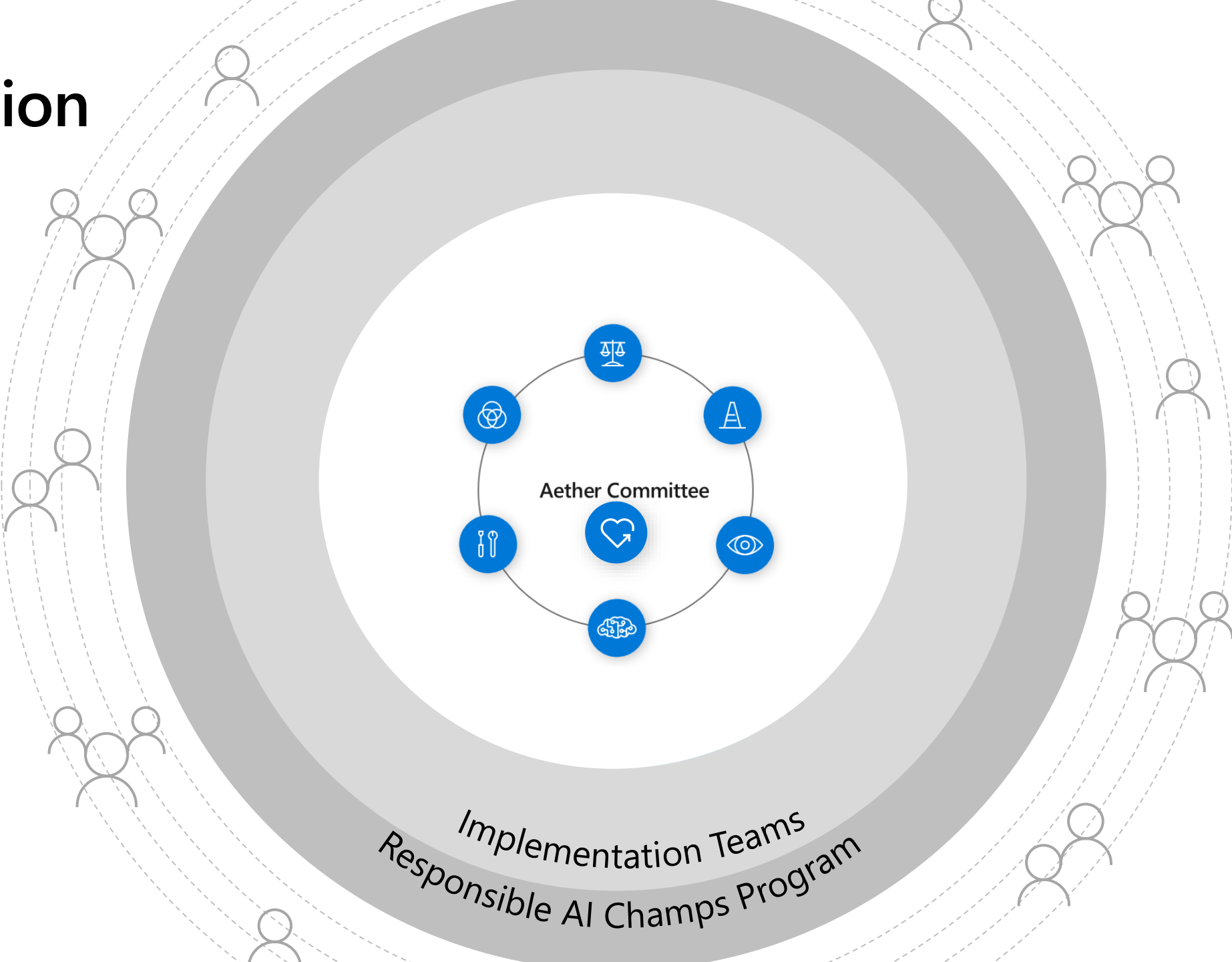
Early Case: Proceed with Constraints

"Microsoft will not design, develop or deliver advanced analytics capabilities, including without limitation, aggregation and/or correlation of data from social media feeds; audio or video detection of age, gender or ethnicity; facial recognition techniques; the use of machine learning or other predictive analytics to predict future events or to classify people or situations."

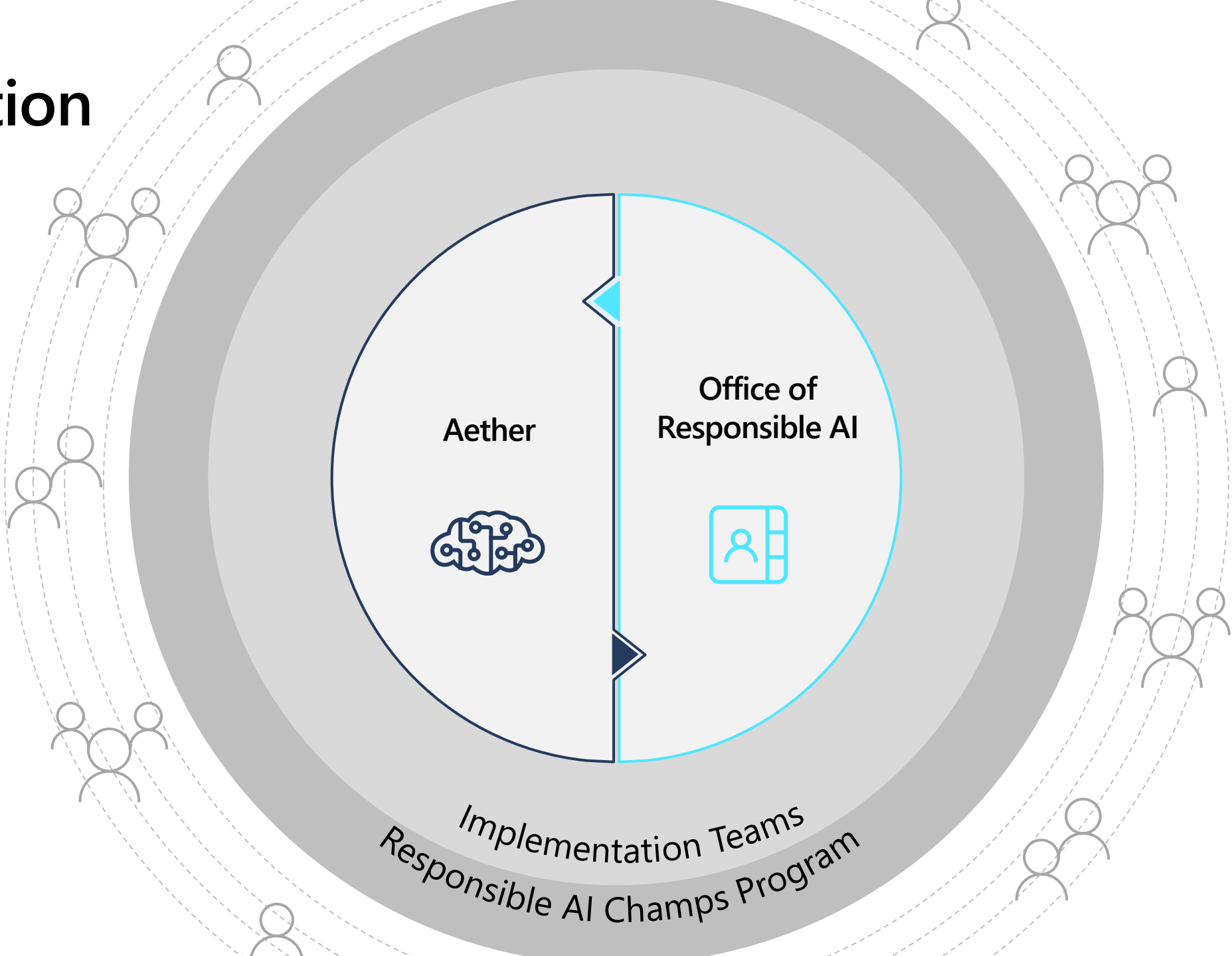
Evolution

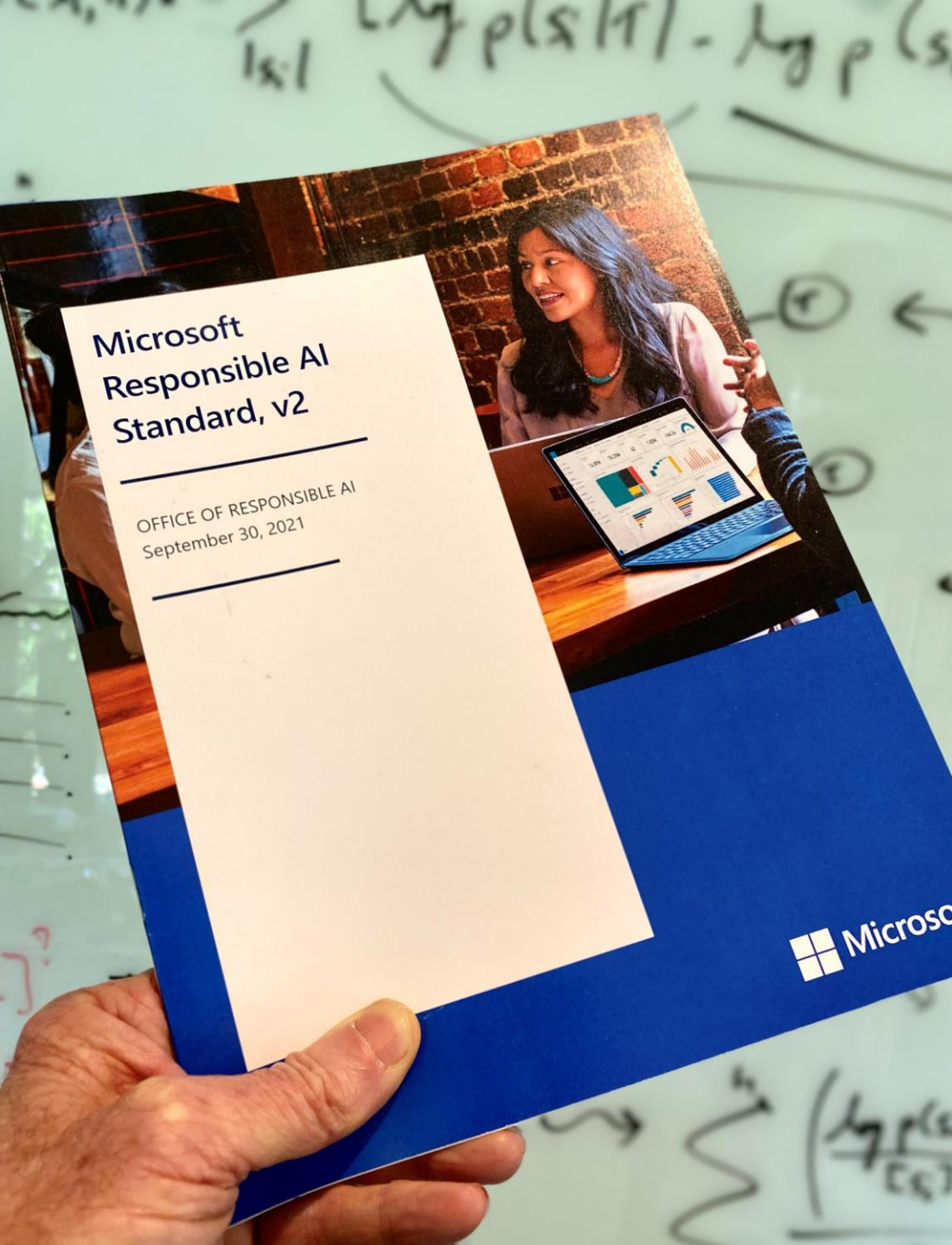


Evolution



Evolution





- **Accountability**

- Impact assessment
- Oversight of significant adverse influences
- Fit for purpose
- Data governance & management
- Human oversight & control

- **Transparency**

- System intelligibility
- Communication to stakeholders
- Disclosure of AI interaction

- **Fairness**

- Quality of service
- Allocation of resources & opportunities
- Minimize stereotyping, demeaning, erasure

- **Reliability & Safety**

- Reliability & safety guidance
- Failures & remediations
- Ongoing monitoring, feedback, evaluation


- **Privacy & Security**

- Secure per MS security policy

- **Inclusiveness**

- Inclusive design MS accessibility

Requirements,
tools, practices




**Microsoft
Responsible AI
Standard, v2**

GENERAL REQUIREMENTS

FOR EXTERNAL RELEASE

June 2022



**Microsoft
Responsible AI
Impact Assessment
Template**

FOR EXTERNAL RELEASE

June 2022

The Responsible AI Impact Assessment Template is the product of a multi-year effort at Microsoft to define a process for assessing the impact an AI system may have on people, organizations, and society. We are releasing our Impact Assessment Template externally to share what we have learned, invite feedback from others, and contribute to the discussion about building better norms and practices around AI.

We invite your feedback on our approach:
<https://aka.ms/ResponsibleAIQuestions>



Transparency Goals

Goal T1: System intelligibility for decision making

Microsoft AI systems that inform decision making by or about people are designed to support stakeholder needs for intelligibility of system behavior.

Applies to: All AI systems when the intended use of the generated outputs is to inform decision making by or about people.

Requirements

T1.1 Identify:

- 1) stakeholders who will use the outputs of the system to make decisions, and
- 2) stakeholders who are subject to decisions informed by the system.

Document these stakeholders using the Impact Assessment template.

Tags: Impact Assessment.

T1.2 Design the system, including, when possible, the system UX, features, reporting functions, and educational materials, so that stakeholders identified in requirement T1.1 can:

- 1) understand the system's intended uses,
- 2) interpret relevant system behavior effectively (i.e., in a way that supports informed decision making), and
- 3) remain aware of the possible tendency of over-relying on outputs produced by the system ("automation bias").

For the two categories of stakeholders identified in requirement T1.1, document:

- 1) how the system design will support their understanding of the system's intended uses, and
- 2) how the system aids their ability to interpret relevant system responses, and
- 3) how the system design discourages automation bias.

T1.3 Define and document the method to be used to evaluate whether each stakeholder who will make decisions or be subject to decisions based on the behavior of the system can interpret the relevant system responses reasonably well. Include the metrics or rubrics that will be used in the evaluations.

Tags: Ongoing Evaluation Checkpoint.

T1.4 Define and document a Responsible Release Plan, to include Responsible Release Criteria to achieve this Goal.

Tags: Ongoing Evaluation Checkpoint.

T1.5 Conduct evaluations defined by requirement T1.3. Document the pre-release results of the evaluations.

Determine and document how often ongoing evaluation should be conducted to continue supporting this Goal.

Tags: Ongoing Evaluation Checkpoint.

T1.6 If there are Responsible Release Criteria for metrics or rubrics that have not been met, consult with the reviewers named in the Impact Assessment, and in the case of Sensitive Uses, with the Office of Responsible AI, to develop a plan detailing how the gap will be managed until it can be closed. Document that plan.

Goal T2: Communication to stakeholders

Microsoft provides information about the capabilities and limitations of our AI systems to support stakeholders in making informed choices about those systems.

Applies to: All AI systems.

Requirements

T2.1 Identify:

- 1) stakeholders who make decisions about whether to employ a system for particular tasks, and
- 2) stakeholders who develop or deploy systems that integrate with this system.

Document these stakeholders in the Impact Assessment template.

Tags: Impact Assessment.

T2.2 Publish documentation for the system so that stakeholders defined in T2.1 can understand the system.

Include:

- 1) capabilities,
- 2) intended uses,
- 3) uses that require extra care or guidance,
- 4) operational factors and settings that allow for effective and responsible system use,
- 5) limitations, including uses for which the system was not designed or evaluated, and
- 6) evidence of system accuracy and performance as well as a description of the extent to which these results are generalizable across use cases that were not part of the evaluation.

When the system is a platform service made available to external customers or partners, a Transparency Note is required.

Tags: Transparency Note.

T2.3 Review and update documentation annually or when any of the following events occur:

- 1) new uses are added,
- 2) functionality changes,
- 3) the product moves to a new release stage,
- 4) new information about reliable and safe performance becomes known as defined by requirement RS3.3, or
- 5) new information about system accuracy and performance becomes available.

When the system is a platform service made available to external customers or partners, include this information in the required Transparency Note.

Tags: Transparency Note.

Goal T3: Disclosure of AI interaction

Microsoft AI systems are designed to inform people that they are interacting with an AI system or are using a system that generates or manipulates image, audio, or video content that could falsely appear to be authentic.

Applies to: AI systems that impersonate interactions with humans, unless it is obvious from the circumstances or context of use that an AI system is in use. AI systems that generate or manipulate image, audio, or video content that could falsely appear to be authentic.

Requirements

T3.1 Identify stakeholders who will use or be exposed to the system, in accordance with the Impact Assessment requirements. Document these stakeholders using the Impact Assessment template.

Tags: Impact Assessment.

T3.2 Design the system, including system UX, features, reporting functions, educational materials, and outputs so that stakeholders identified in T3.1 will be informed of the type of AI system they are interacting with or exposed to. Ensure that any image, audio, or video outputs that are intended to be used outside the system are labelled as being produced by AI.

T3.3 Define and document the method to be used to evaluate whether each stakeholder identified in T3.1 is informed of the type of AI system they are interacting with or exposed to.

Tags: Ongoing Evaluation Checkpoint.

T3.4 Define and document Responsible Release Criteria to achieve this Goal.

Tags: Ongoing Evaluation Checkpoint.

T3.5 Conduct evaluations defined by requirement T3.3. Document the pre-release results of the evaluations. Determine and document how often ongoing evaluation should be conducted to continue supporting this goal.

Tags: Ongoing Evaluation Checkpoint.

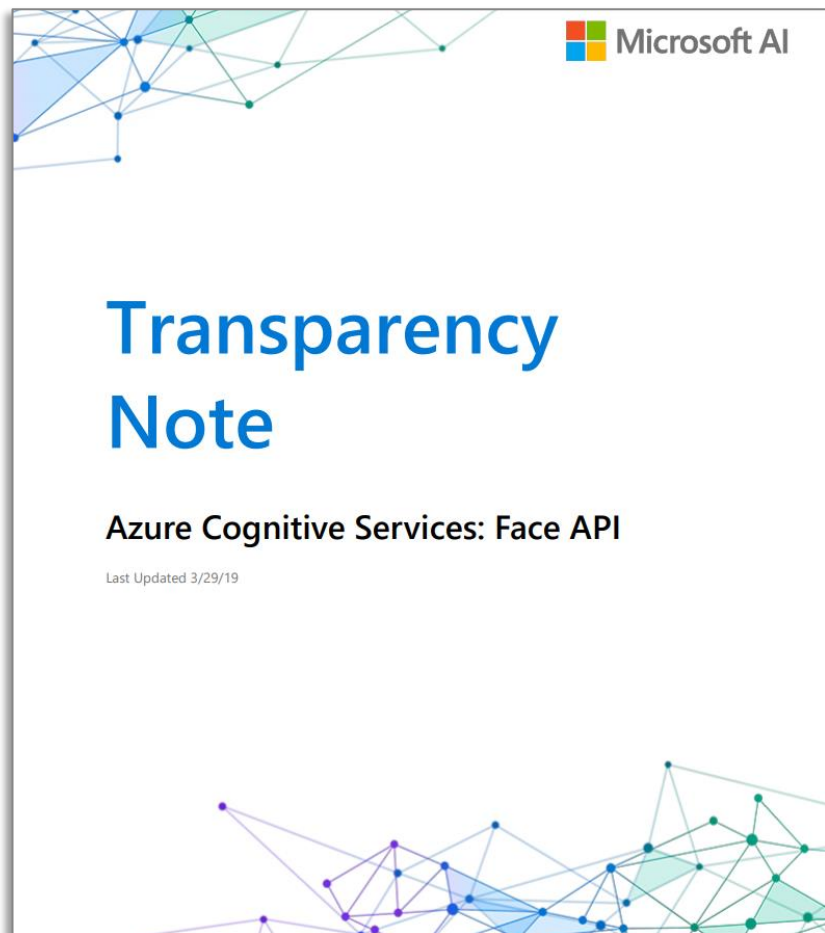
Tools and practices

Recommendation T1.2.1 Follow the Guidelines for Human-AI Interaction when designing the system.

Recommendation T1.2.2 Use one or more techniques available as part of the Interpret ML toolkit to understand the impact of features on system behavior. This may help stakeholders who need to understand model predictions.

Recommendation T1.3.1 Assign user researchers to define, design, and prioritize evaluations in appropriately realistic contexts of use.

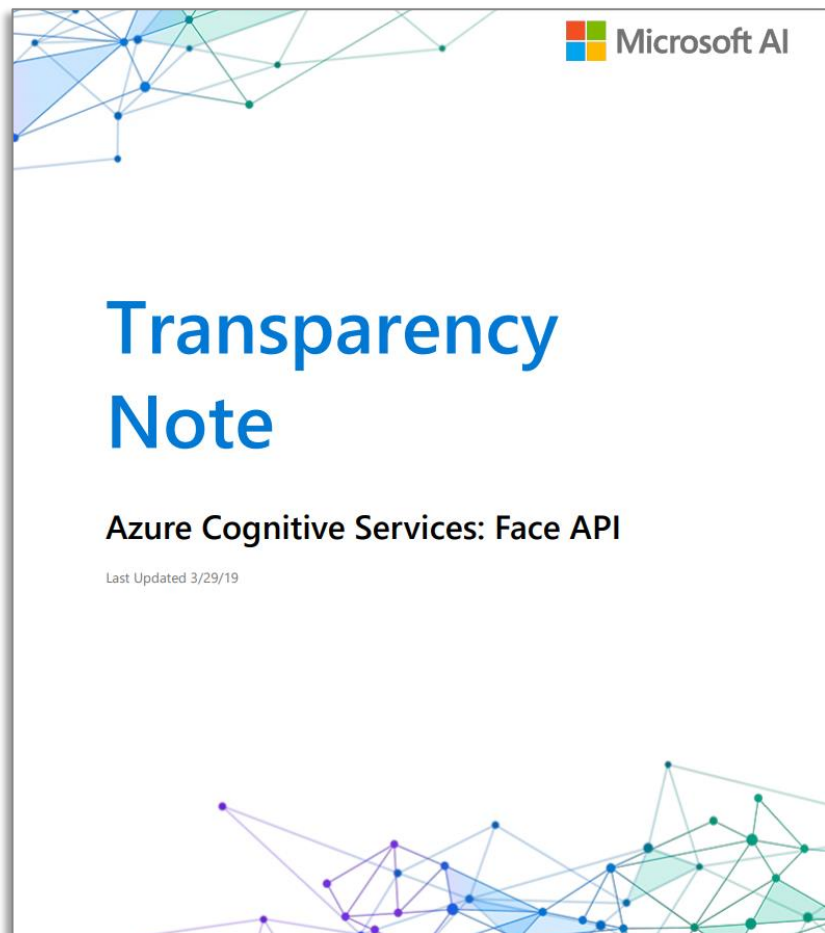
Transparency Note



Contents

About this Transparency Note.....	2
The basics of Face API.....	2
Key facial recognition terms.....	2
Face API functions.....	3
Understanding accuracy and errors.....	3
How accurate is Face API?.....	3
The language of accuracy.....	4
Match scores, match thresholds, and candidate lists.....	5
Best practices for improving accuracy.....	6
Plan for an evaluation phase.....	6
Meet image quality specifications.....	6
Control image capture environment.....	7
Plan for variations in subject appearance and behavior.....	7
Design the system to support human judgment.....	8
Use multiple factors for authentication.....	8
Deploying responsible facial recognition systems.....	9
Evaluate stakeholder concerns and design the experience to address them.....	9
Develop transparent communication and escalation processes for stakeholder concerns.....	9
Provide training and evaluate the effectiveness of people who make final judgments based on facial recognition.....	9
Update privacy policies and implement necessary changes.....	9
Learn more about Face API.....	10
Contact us.....	10
About this document.....	10

Transparency Note



Probe image

A probe image is an image submitted to a facial recognition system to be compared to enrolled individuals. Probe images are also converted to probe templates. As with enrollment templates, high-quality images result in high-quality templates.

Face API functions

Face API Detection ("Detection") answers the question, *"Are there one or more human faces in this image?"* Detection finds human faces in an image and returns bounding boxes indicating their locations. All other functions are dependent on Detection: before Face API can identify or verify a person (see below), it must know the locations of the faces to be recognized.

The Detection function can also be used to predict attributes about each face, including age and gender. These attribute prediction functions are completely separate from the verification and identification functions of Face API. Face API does not predict an individual's age or gender as a precursor to verifying or identifying them.

Face API Verification ("Verification") builds on Detection and addresses the question, *"Are these two images the same person?"*. In security or access scenarios, Verification relies on the existence of a primary identifier (such as a customer ID) and facial recognition is used as a second factor to *verify* the person's identity. Verification is also called "one-to-one" matching because the probe template (one person) is only compared to the template stored for the (one) person associated with the identification presented.

Face API Identification ("Identification") also starts with Detection and answers the question, *"Can this unknown person be matched to an enrolled template?"* Identification compares a probe template to all enrollment templates stored in your private repository, so it is also called "one-to-many" matching. Candidate matches are returned based on how closely the probe template matches each of the enrolled templates.

Face API can answer the questions:

1. Are there one or more human faces in this image?
2. Are these two images the same person?
3. Can this unknown person be matched to an enrolled template?

Face API documentation


For more information on all of the functions of Face API, see the [Face API documentation](#)

Understanding accuracy and errors

How accurate is Face API?

Because Face API is a building block for creating a facial recognition system to which other building blocks must be added, it is not possible to provide a universally applicable estimate of accuracy for the actual system you are planning to deploy. Companies may share accuracy as measured by public benchmark competitions, but these accuracies depend on details of the benchmark and therefore won't be the same as the accuracy of a deployed system. Ultimately, system accuracy depends on a number of factors, including the technology and how it is configured, environmental conditions, the use case for the system, how people to be recognized interact with the camera, and how people interpret the system's output. The following section is intended to help you understand key concepts that describe accuracy in the context of a facial recognition system. With that understanding, we then describe [system design choices](#) and how they influence accuracy.

Transparency Note







Transparency Note

Azure Cognitive Services: Face API

Last Updated 3/29/19

The language of accuracy

The **accuracy** of a facial recognition system is based on a combination of two things: how often the system correctly identifies a person who *is enrolled* in the system and how often the system correctly finds no match for a person who *is not enrolled*. These two conditions, which are referred to as the "true" conditions, combine with two "false" conditions to describe all possible outcomes of a facial recognition system:

True positive or true accept 	The person in the probe image is enrolled and they are correctly matched.
True negative or true reject 	The person in the probe image is not enrolled and they are not matched.
False positive or false accept 	Either the person in the probe image is <i>not enrolled</i> but is matched to an <i>enrolled</i> person OR the person in the probe image is enrolled but is matched with the wrong person.
False negative or false reject 	The person in the probe image is enrolled, but they are not matched.

The consequences of a false positive or a false negative vary depending on the purpose of the facial recognition system. The examples below illustrate this variation and how choices you make in designing the system affect the experience of those people who are subject to it.

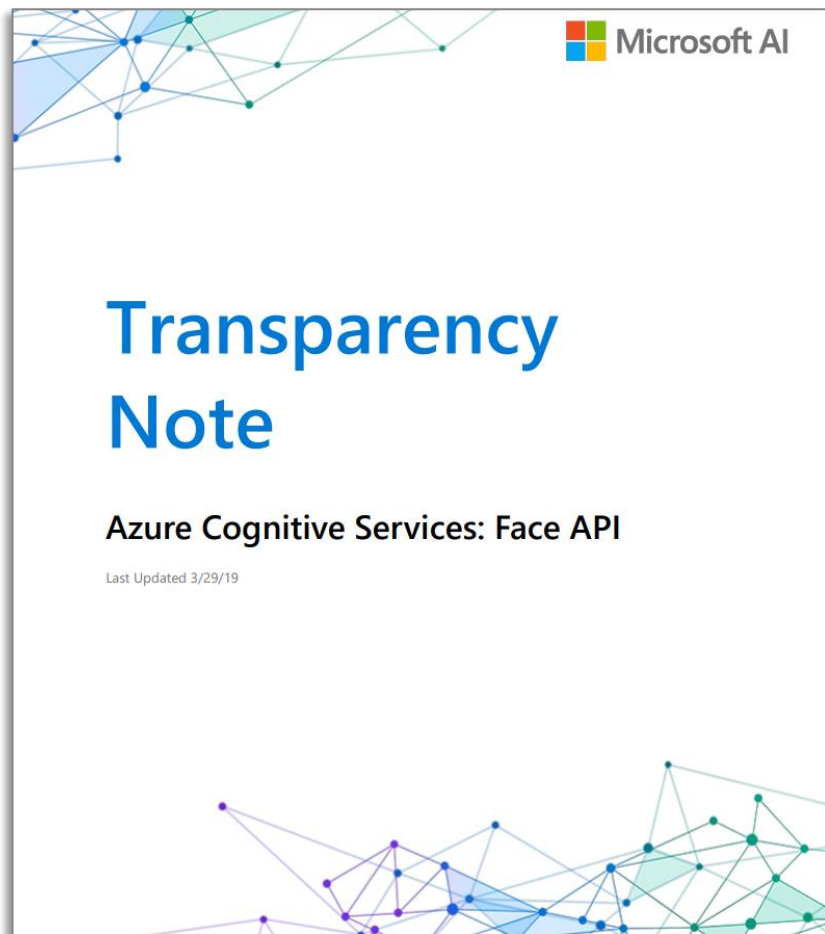
Logging into a banking app

Facial recognition can provide an added layer of security in addition to a PIN or other primary identification. A false positive for this application reduces customer security because it results in an incorrect match, while a false negative could prevent the customer from accessing their account. Because the purpose of the system is security, false positives must be minimized and as a result, most errors will be false negatives (account access fails). To address this limitation, system owners can provide a fallback mechanism, like pushing a notification to the customer's phone with an access code. The customer's experience may be less efficient in this case, but account access is not blocked, and security is prioritized.

Organizing photographs

Many photo organizing apps help you find pictures of a specific person across your photo collection using facial recognition. In this instance, the customer is using the app to choose photos for a retirement party. Because the customer will be reviewing the photos and choosing photos they wish to use, false positives may not be particularly important: facial recognition is making the search task easier for the customer, and if they review a few more photos than necessary, they can still easily complete their task. On the other hand, if they have scanned old family photographs that are somewhat degraded, the app may not be able to find relatives in these photographs (false negatives) and the customer may be frustrated with the app.

Transparency Note



Deploying responsible facial recognition systems

In addition to addressing accuracy, here are some additional considerations for successful deployment.

Evaluate stakeholder concerns and design the experience to address them

Understand both the perceived value of the facial recognition system and the concerns that people may have about it. Engage your research team to help understand how your customers, employees, and other stakeholders can help you deploy a system that supports your critical needs and those of the people who will be involved.

Develop transparent communication and escalation processes for stakeholder concerns

People may still have questions and concerns. Part of any release plan should include both proactive and reactive communication, a documented escalation process, and clear explanations for how feedback will be addressed.

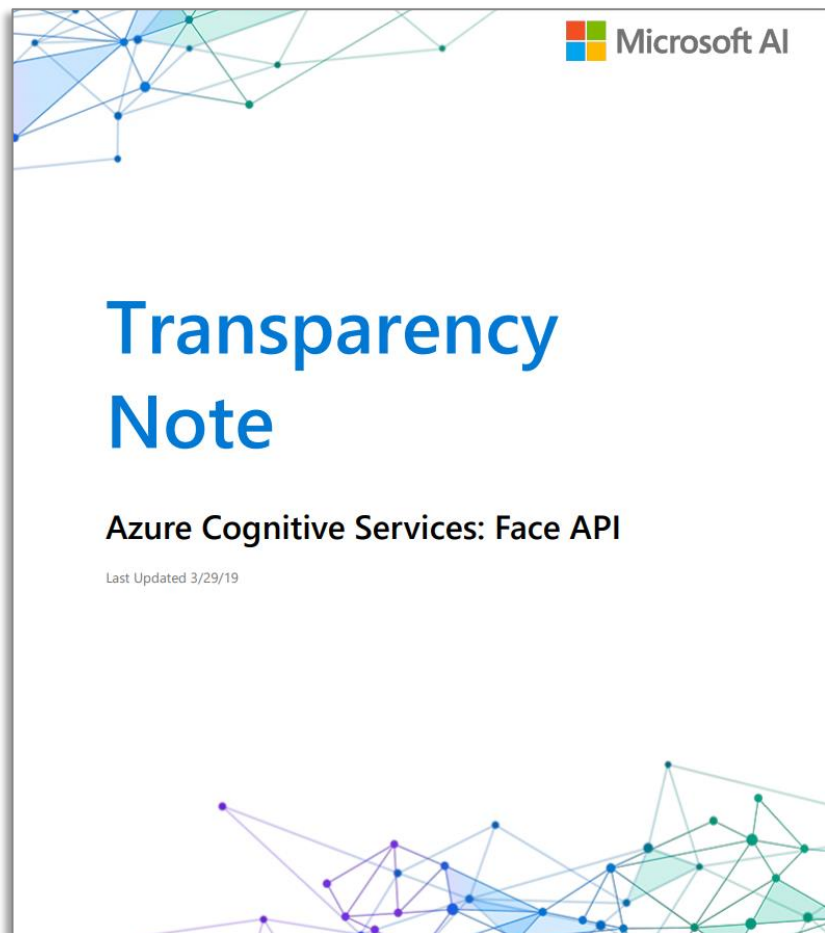
Provide training and evaluate the effectiveness of people who make final judgments based on facial recognition

Microsoft strongly recommends that customers develop training for people who will use the output of systems or who will decide whether the system output is correct. Customers should also evaluate whether these employees can make correct judgments based on the output of the system and determine whether any unfair biases are introduced.

Update privacy policies and implement necessary changes

Microsoft strongly recommends that private sector customers provide conspicuous notice to and secure consent from individuals before capturing their images for use with facial recognition technology. System owners should also establish responsible data handling practices (including limits on retention and reuse of images) and ensure that those practices are communicated clearly to individuals subject to the system. Remember to include considerations for children who may be subject to recognition. In some jurisdictions, there may be additional legal requirements, and customers are responsible for compliance with all applicable laws.

Transparency Note



Design the system to support human judgment

In most cases, Microsoft recommends using Face API's facial recognition capabilities to support people making more accurate and efficient judgements rather than fully automating a process. Meaningful human review is important to:

- Detect and resolve cases of misidentification or other failures.
- Provide support to people who believe their results were incorrect.
- Identify and resolve changes in accuracy due to changing conditions (like lighting or sensor cleanliness).

For example, when using Face API for building security, a trained security officer can help when the facial recognition fails to match someone who believes they are enrolled by deciding whether a person should be admitted to the building. In this case, Face API helps the security officer work more efficiently, requiring a judgment to admit someone only when the person is not recognized.

The user experience that you create to support the people who will use the system output should be designed and evaluated with those people to understand how well they can interpret the output, what additional information they might need, how they can get answers to their questions, and ultimately, how well the system supports their abilities to make more accurate judgments.

Face API supports facial recognition with still images: there are *no anti-spoofing countermeasures* built into Face API, such as depth or motion detection. In cases where facial recognition is supporting human judgment and improving efficiency, this is generally not a key limitation: humans can easily detect when a person is holding up a picture to a camera.

Use multiple factors for authentication

Use Face API along with one or more other factors when creating authentication systems, such as confirming passengers who are about to board a plane or confirming a banking transaction. As discussed above, Verification makes use of facial recognition as a second factor for identifying someone rather than a single or

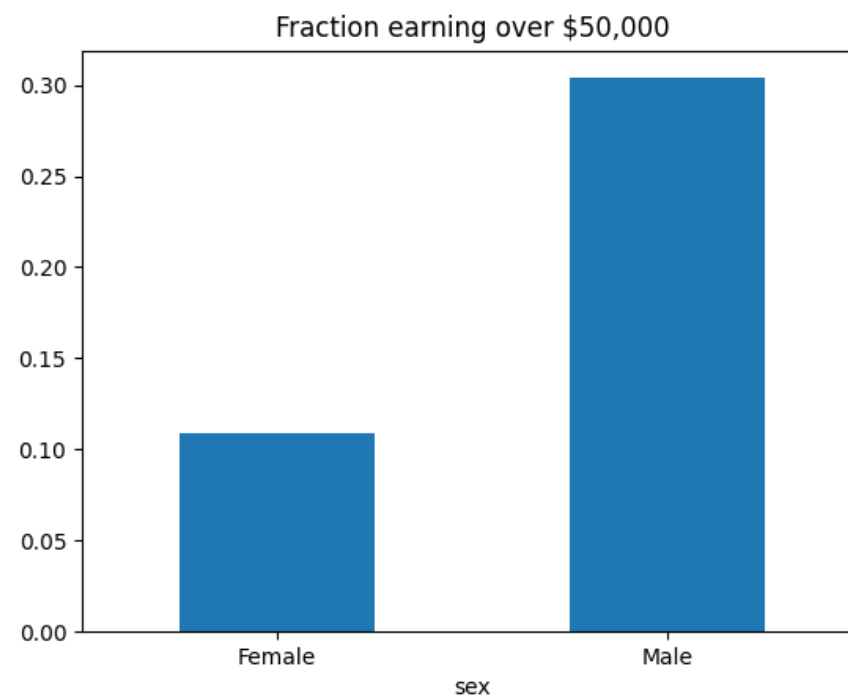
Tools: Fairlearn

Fairlearn

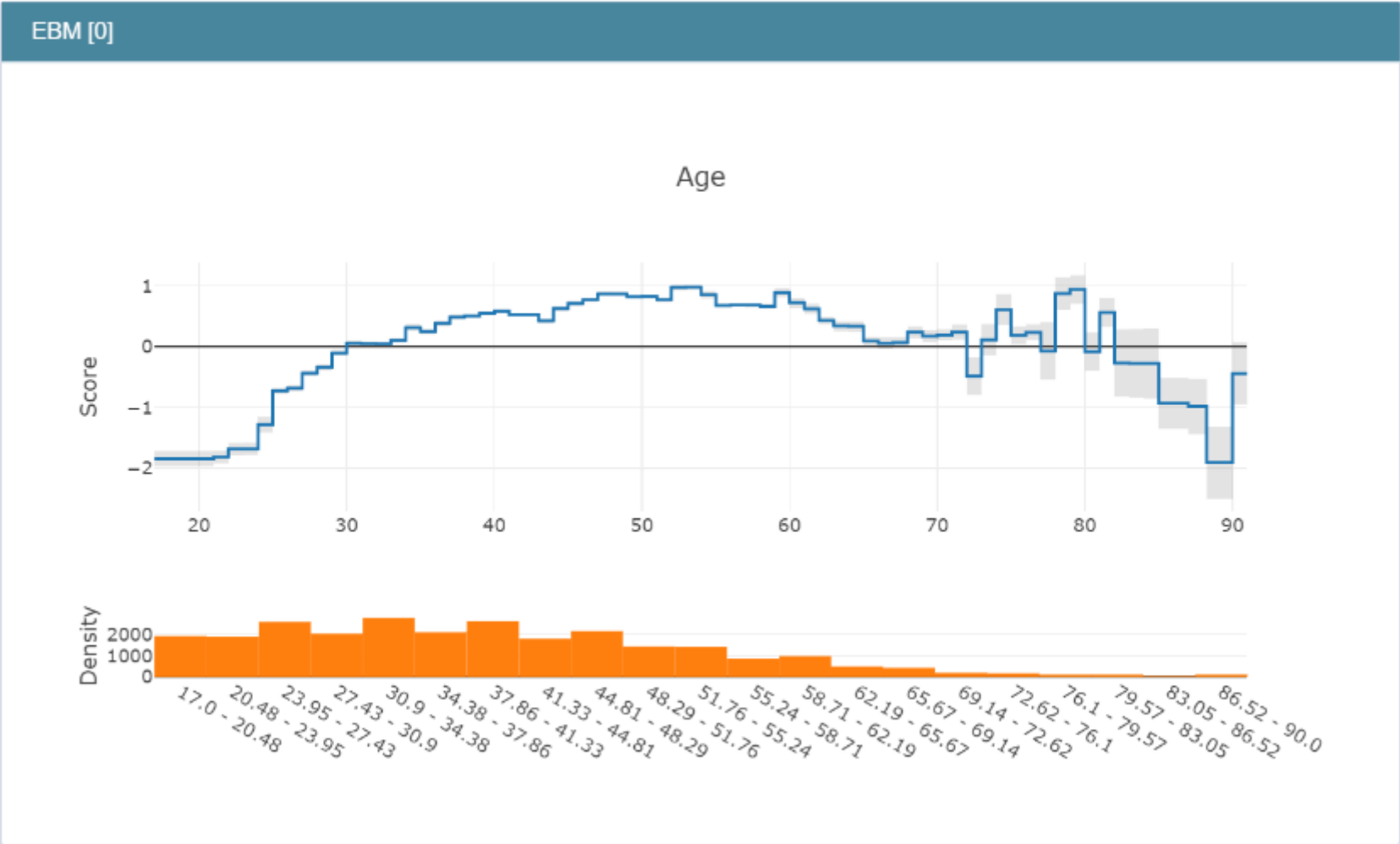
Evaluating fairness-related metrics

Firstly, Fairlearn provides fairness-related metrics that can be compared between groups and for the overall population. Using existing metric definitions from [scikit-learn](#) we can evaluate metrics for subgroups within the data as below:

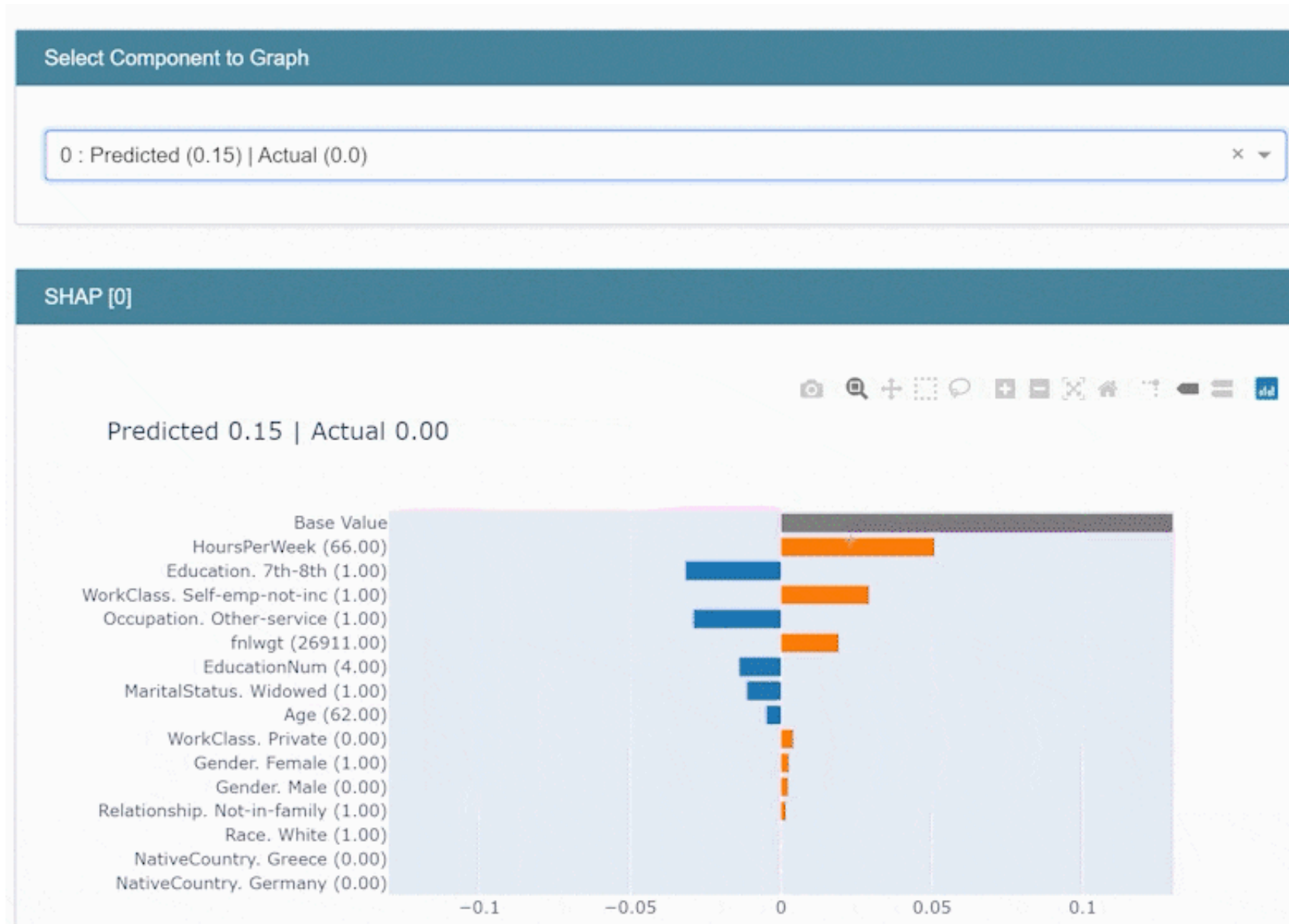
```
>>> from fairlearn.metrics import MetricFrame
>>> from sklearn.metrics import accuracy_score
>>> from sklearn.tree import DecisionTreeClassifier
>>>
>>> classifier = DecisionTreeClassifier(min_samples_leaf=10, max_depth=4)
>>> classifier.fit(X, y_true)
DecisionTreeClassifier(...)
>>> y_pred = classifier.predict(X)
>>> gm = MetricFrame(metrics=accuracy_score, y_true=y_true, y_pred=y_pred, sensitive_features=sex)
>>> print(gm.overall)
0.8443...
>>> print(gm.by_group)
sex
Female    0.9251...
Male     0.8042...
Name: accuracy_score, dtype: object
```



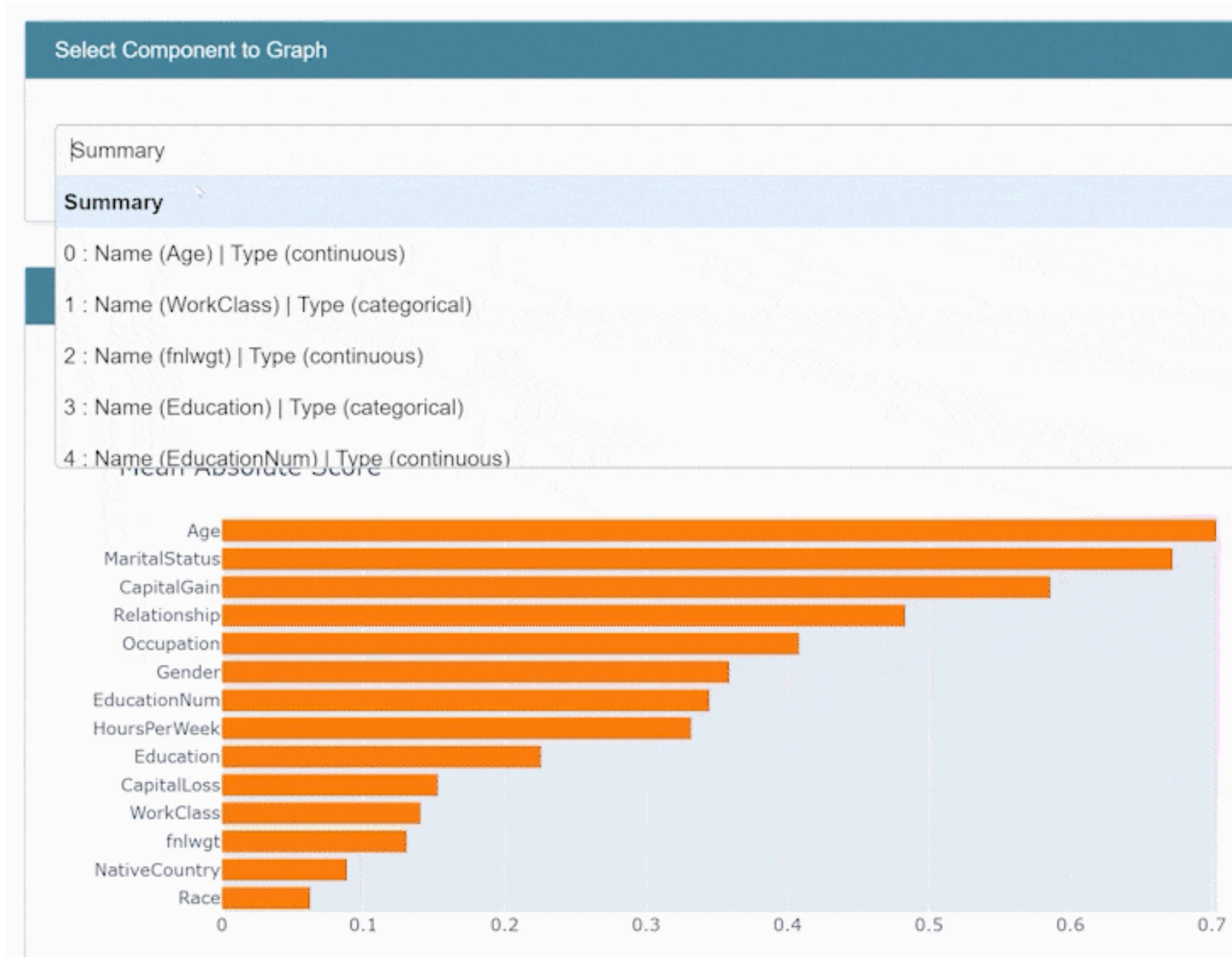
Tools: InterpretML



Tools: InterpretML



Tools: InterpretML



Tools: Error Analysis

Global cohort: All data (default) [Switch global cohort](#) [+ Create new cohort](#)

[Cohort settings](#) [Dashboard navigation](#)

Error analysis

[Tree map](#) [Heat map](#) [Feature list](#)

The tree visualization uses the mutual information between each feature and the error to best separate error instances from success instances hierarchically in the data. This simplifies the process of discovering and highlighting common failure patterns. To find important failure patterns, look for nodes with a stronger red color (i.e., high error rate) and a higher fill line (i.e., high error coverage). To edit the list of features being used in the tree, click on "Feature list."

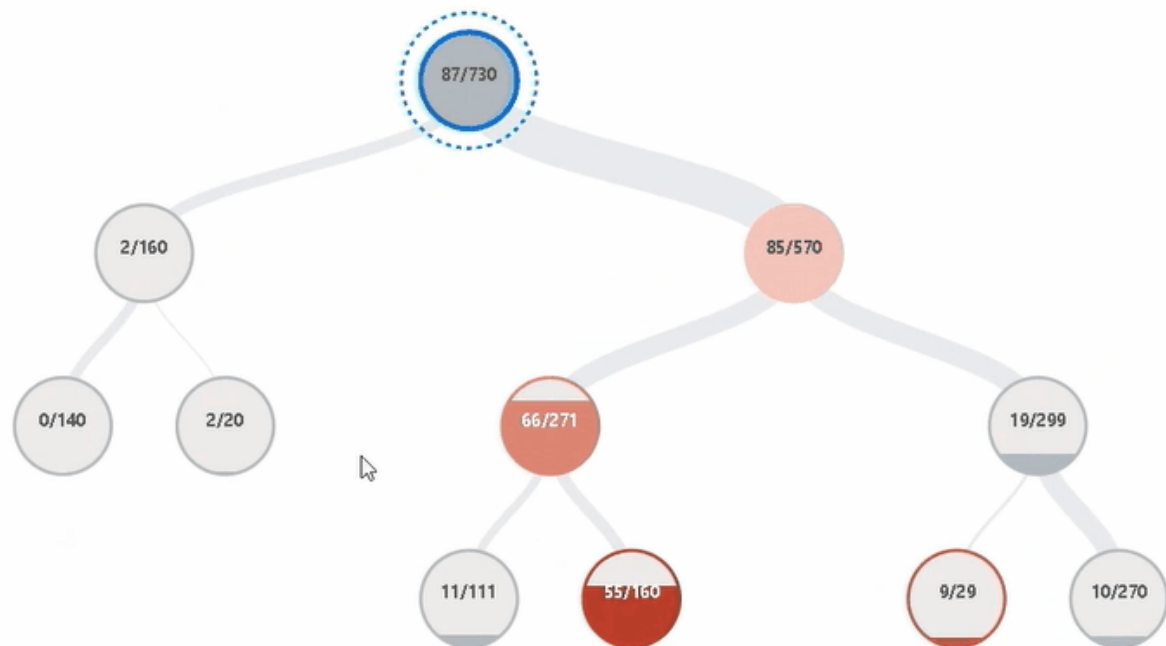
Select metric

Error rate

Error coverage [ⓘ]
100.00%



Error rate [ⓘ]
11.92%



Save as a new cohort

Basic Information

All data (0 filters)

Instances in base cohort

Total 730

Correct 643

Incorrect 87

Instances in the selected cohort

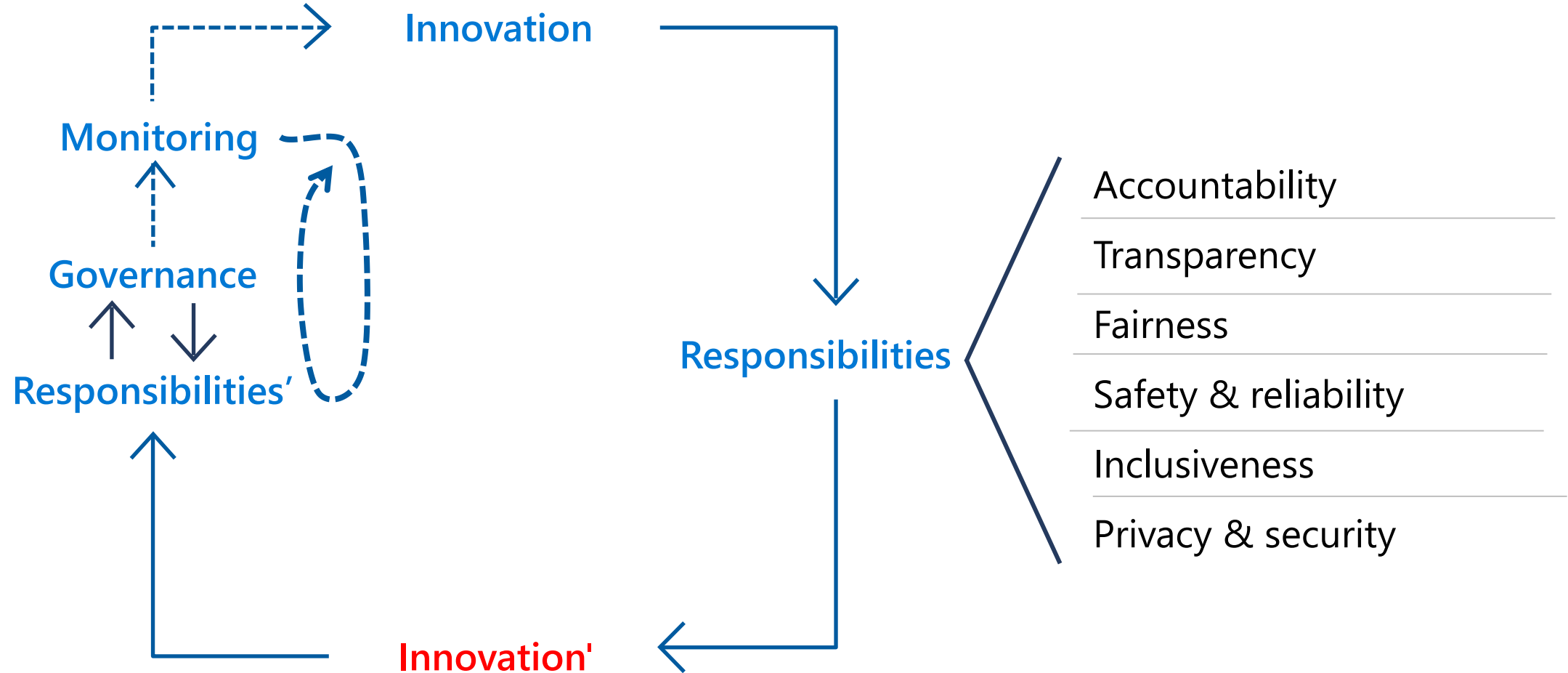
Total 730

Correct 643

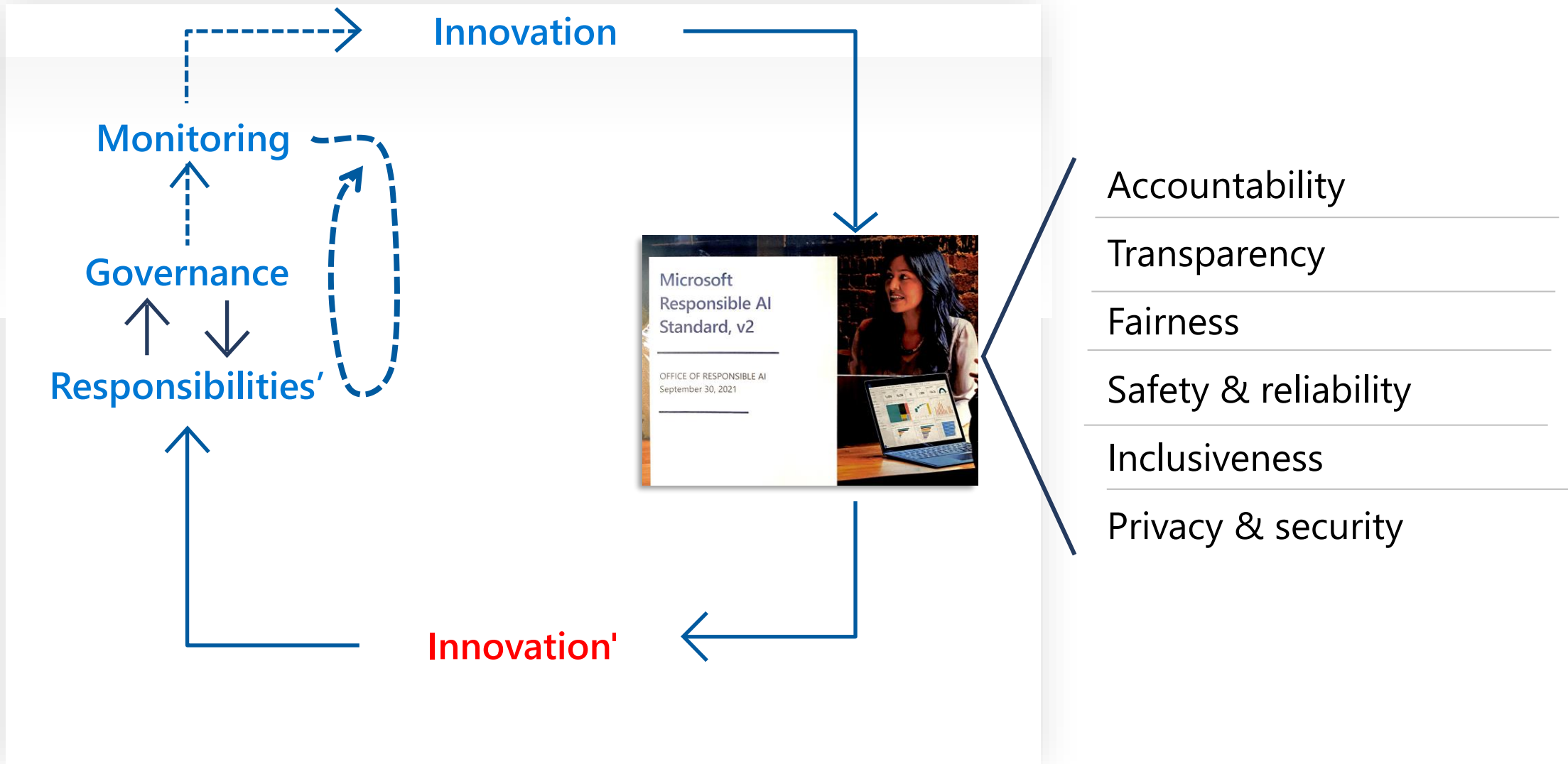
Incorrect 87

Prediction path (filters)

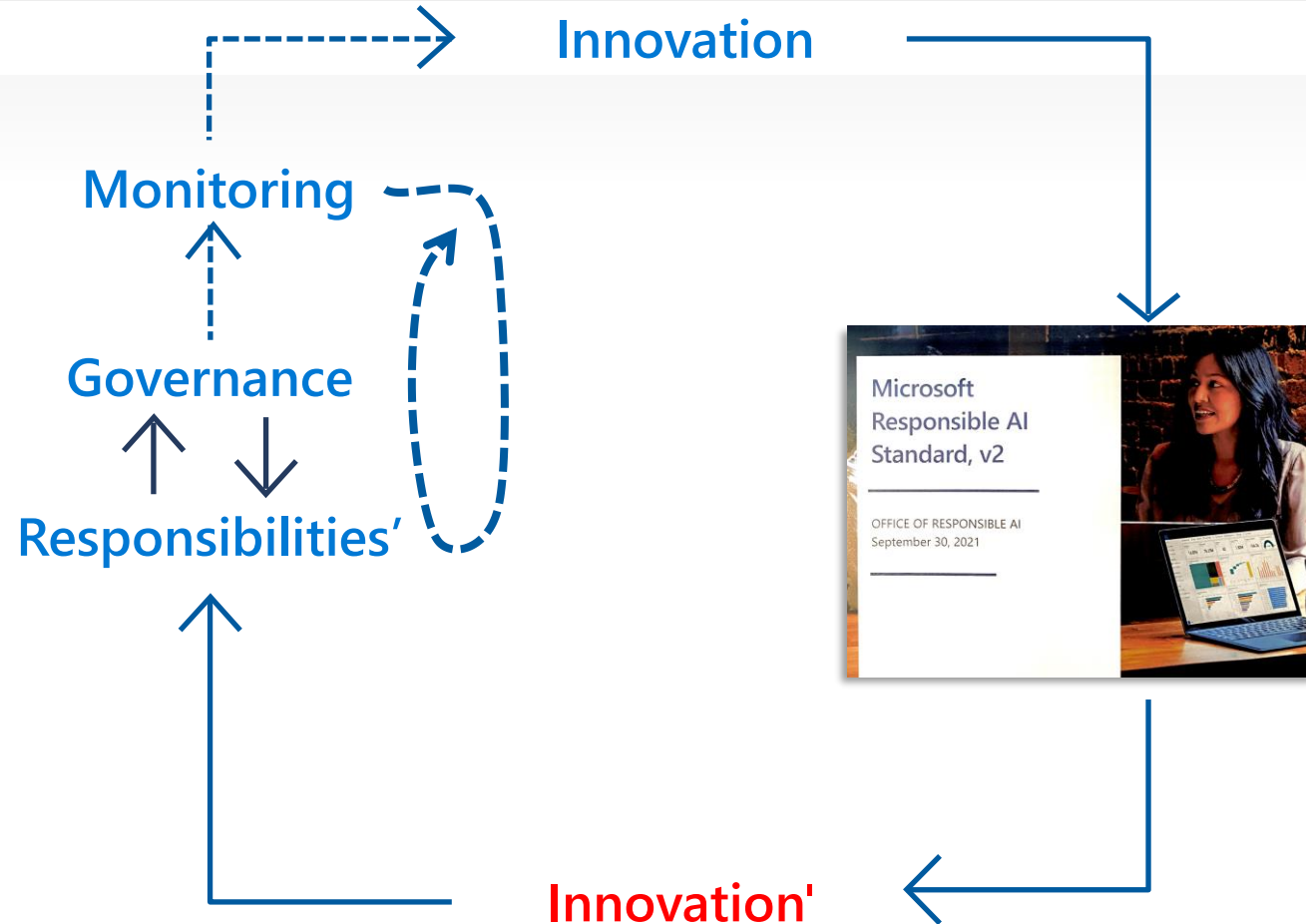
Cycle of Responsible Innovation



Cycle of Responsible Innovation

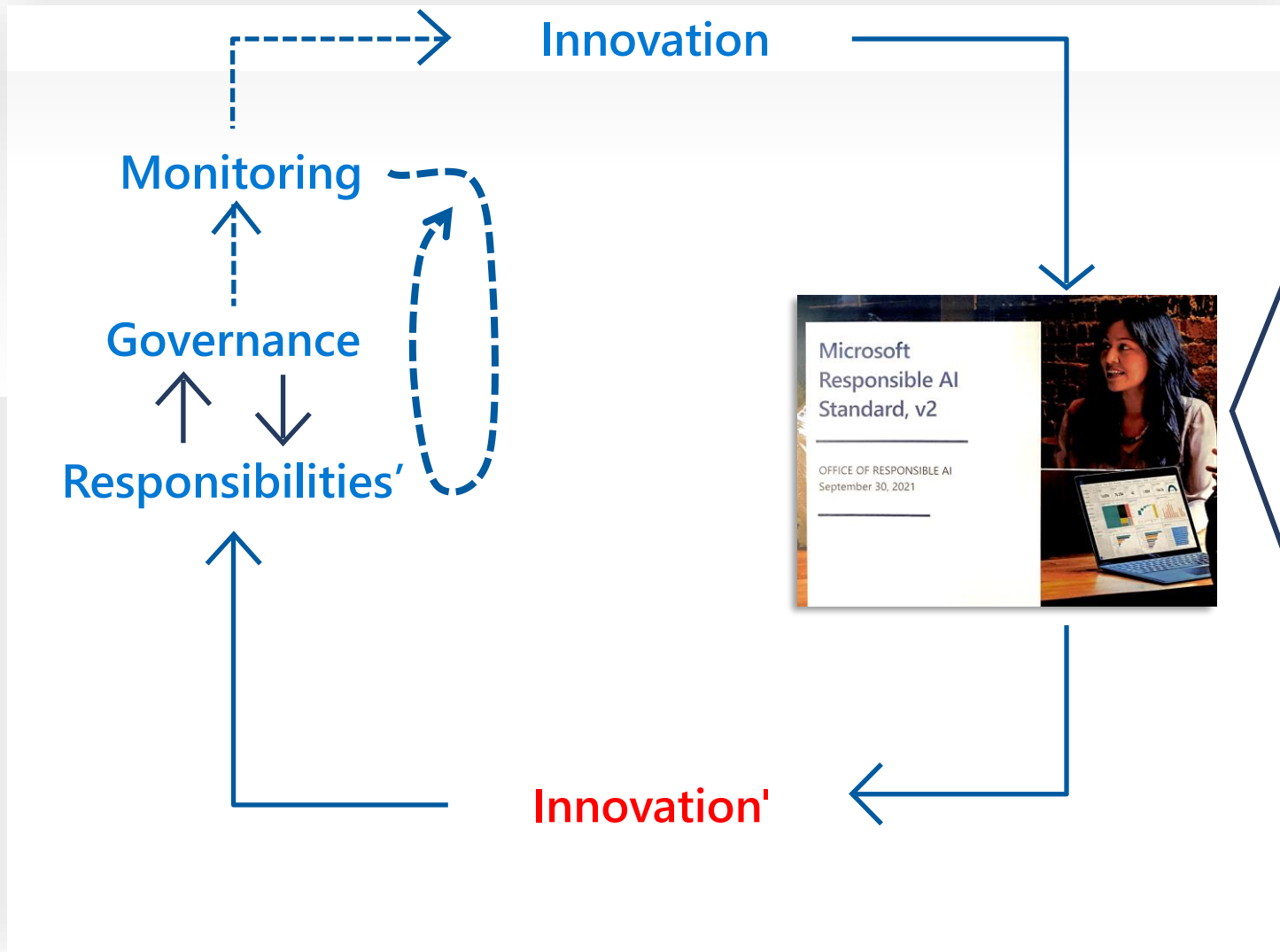


Cycle of Responsible Innovation



- **Accountability**
 - Impact assessment
 - Oversight of significant adverse influences
 - Fit for purpose
 - Data governance & management
 - Human oversight & control
- **Transparency**
 - System intelligibility
 - Communication to stakeholders
 - Disclosure of AI interaction
- **Fairness**
 - Quality of service
 - Allocation of resources & opportunities
 - Minimize stereotyping, demeaning, erasure
- **Reliability & Safety**
 - Reliability & safety guidance
 - Failures & remediations
 - Ongoing monitoring, feedback, evaluation
- **Privacy & Security**
 - Secure per MS security policy
- **Inclusiveness**
 - Inclusive design MS accessibility

Cycle of Responsible Innovation



Practices & tools

[Error Analysis](#): Analyzes model errors

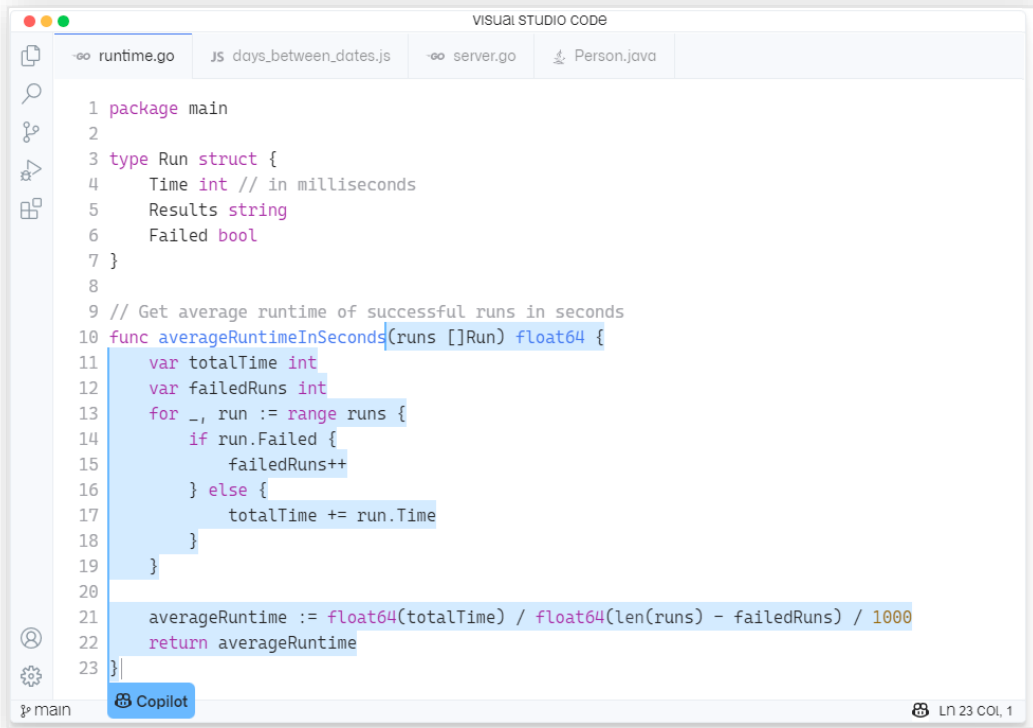
[Fairlearn](#): Assess & mitigate bias

[InterpretML](#): Debug data & inference

[HAX Toolkit](#): Human-AI collaboration

Disruptive capability: Copilot

e.g., Copilot: Assists programmers via code generation



```
1 package main
2
3 type Run struct {
4     Time int // in milliseconds
5     Results string
6     Failed bool
7 }
8
9 // Get average runtime of successful runs in seconds
10 func averageRuntimeInSeconds(runs []Run) float64 {
11     var totalTime int
12     var failedRuns int
13     for _, run := range runs {
14         if run.Failed {
15             failedRuns++
16         } else {
17             totalTime += run.Time
18         }
19     }
20
21     averageRuntime := float64(totalTime) / float64(len(runs) - failedRuns) / 1000
22     return averageRuntime
23 }
```

The screenshot shows the Visual Studio Code interface with a Go file open. The code is a function that calculates the average runtime of successful runs. The Copilot logo is visible in the bottom left corner of the editor.

Special Aether study

Identified:

- Security vulnerabilities
 - Injection of exploits
 - Malware at scale
- Programmer overreliance
- Leaks of information
- Offensive content

Outcome:

30+ requests

Cross-org mobilization

- ✓ New safety features
- ✓ Intensive monitoring & analysis for emerging issues.



Disruptive Capability: Synthetic Voice

Context:

AI can generate realistic voices based on small amounts of training data.



Text input



Text analyzer



Neural acoustic model



Neural vocoder



Audio input

Findings:

New kinds of fraud, impersonation, deception

Need to develop policies that restrict uses but enable acceptable uses

Outcome/Progress:

Sensitive uses review of **Custom Neural Voice**

Established principles

Policies and controls

Note

As part of Microsoft's commitment to designing responsible AI, we have limited the use of Custom Neural Voice. You may gain access to the technology only after your applications are reviewed and you have committed to using it in alignment with our responsible AI principles. Learn more about our policy on the limit access and apply [here](#).



Disruptive Capability: AI-Generated Synthetic Media

Context:

Generative AI use in synthetic & manipulated media poses risk to trusted journalism.

Threat to democracy: Loss of trust, acceleration of disinformation

Findings:

Assessment: **AI detection methods will fail**

Innovation: New media provenance technologies

Certify origin and history of changes to digital content

Outcome/Progress:

Challenge: Can we build tech for “glass-to-glass” authentication of media provenance?

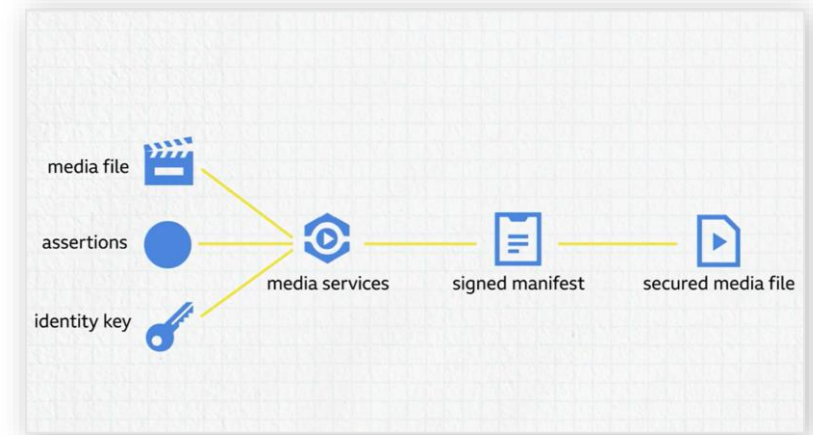
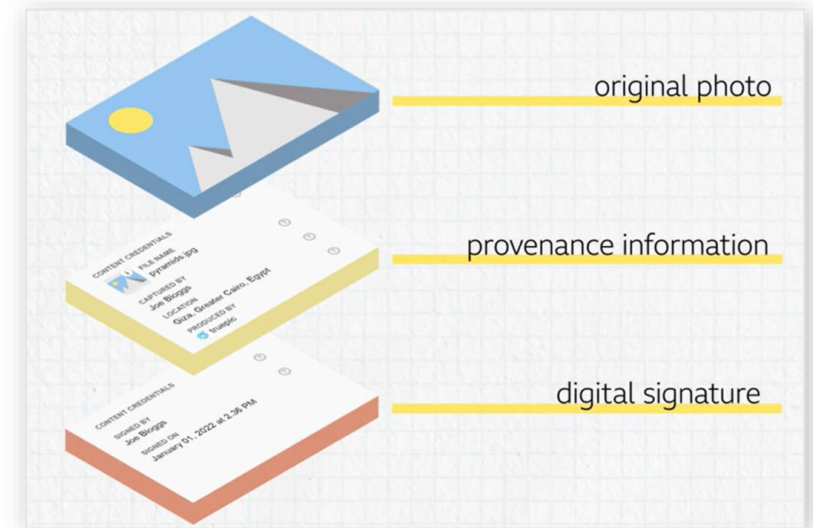
Intensive outreach & teaming:

- Project Origin with MSFT, BBC, NYTimes
- Coalition for Content Provenance and Authenticity (C2PA):

MSFT, Adobe, Arm, BBC, Intel, and Truepic

[C2PA open standard released Jan 26, 2022](#)

Bill in Congress: Portman & Peters’ Deepfake Task Force Act



Disruptive Capability: AI-Generated Synthetic Media

Azure Content Trust Service (ACTS) video service



Video verification

Adobe photo app

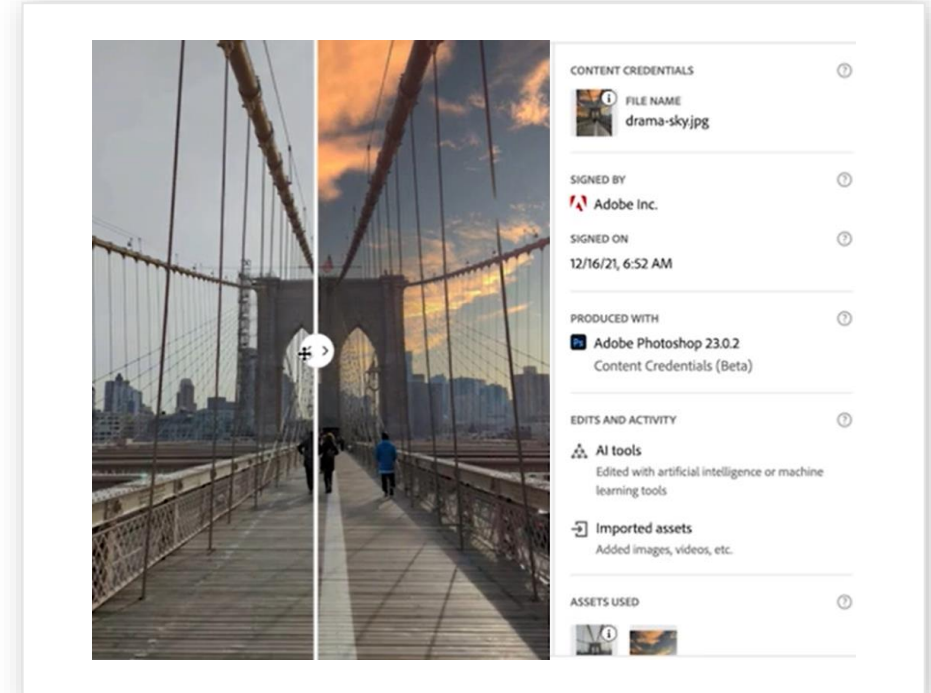


Photo verification

Disruptive Capability: Multimodal Modals

DALL·E creates images from text captions for a wide range of concepts expressible in natural language.

TEXT PROMPT an armchair in the shape of an avocado. . . .

AI-GENERATED
IMAGES



[Edit prompt or view more images](#) ↓

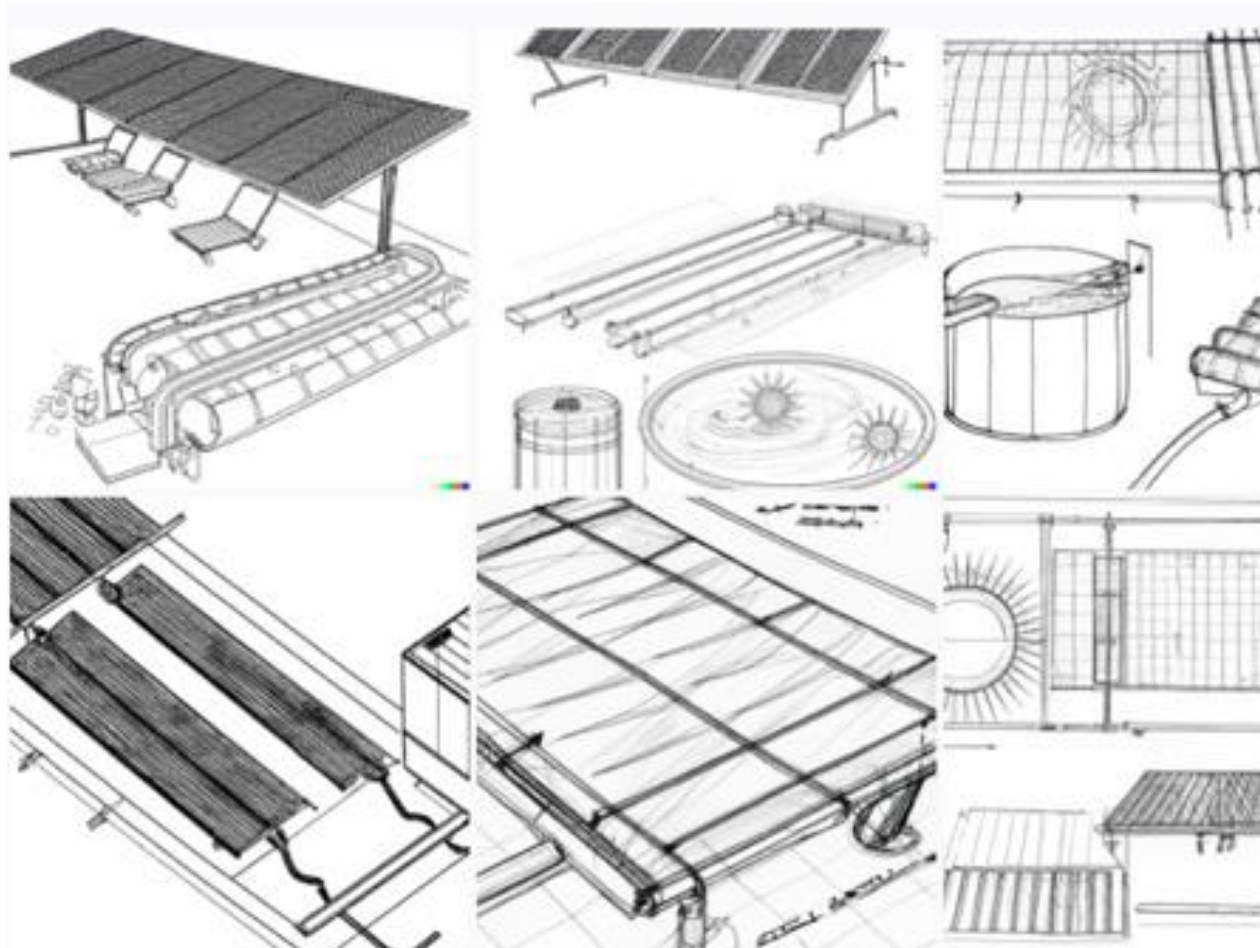
Disruptive Capability: Multimodal Modals

DALL·E creates images from text captions for a wide range of concepts expressible in natural language.



Disruptive Capability: Multimodal Modals

DALL·E creates images from text captions for a wide range of concepts expressible in natural language.



Moving Forward

Expect disruptive innovations & capabilities

AI principles and applications evolving quickly

Invest in understanding & addressing failures, costs, surprises

Pursue mitigations, best practices, and regulations for sociotechnical, geopolitical, civil liberties, ethical challenges.

Tight interleaving

Interleave core innovations with intensive efforts on responsibilities & governance.

Monitor advances, applications, influences.



Learnings, Insights → Governance

Rapid pace of AI advancements & applications
→ multiple forms of governance.



Corporate self-regulation with sharing of best practices
(Companies, Partnership on AI, etc.)



Professional societies, standards bodies, and safety organizations
(ISO, IEEE, etc.)



Federal and state government legislation and regulation
(FDA, FTC, CPSC, NHTSA, Uniform Law Commission, etc.)



Multinational understandings, coordination, and treaties
(Exec, State, Defense, e.g., NATO, US-China, UN, etc.)

Music selection

回忆 Memories

把你写过的日记
埋藏在我心底
写下我所有的记忆
把回忆留给自己
把你写在我心里
写满了岁月的痕迹

不是因为我知道
让我想念你的微笑
让我听见你的心跳
自从遇见你那一秒



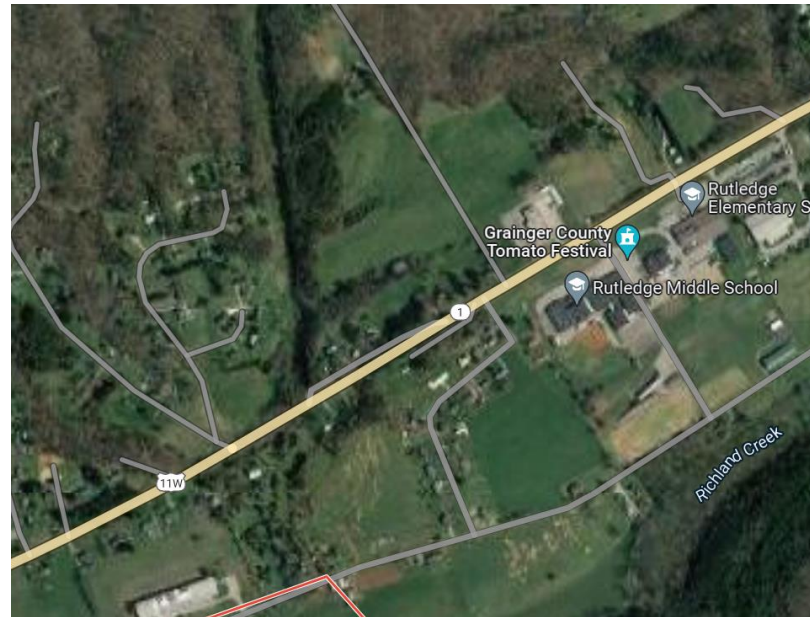


Dixieland Delight

Ronnie Rogers, 1982

Alabama, 1983

“Ronnie Rogers was driving down Highway 11W in Rutledge, Tennessee when these experiences streamed into his mind.”





Dixieland Delight

Ronnie Rogers, 1982

Alabama, 1983

“Ronnie Rogers was driving down Highway 11W in Rutledge, Tennessee when these experiences streamed into his mind.”



