

# Technical Advisory Council Meeting

June 18, 2020

 THE **LINUX** FOUNDATION

 LF AI

# Antitrust Policy Notice

- › Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.
- › Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrave of the firm of Gesmer Undergone LLP, which provides legal counsel to the Linux Foundation.

# Recording of Calls

## Reminder:

TAC calls are recorded and available for viewing on the [TAC Wiki](#)

# Reminder: LF AI Useful Links

Web site: [lfai.foundation](https://lfai.foundation)  
Wiki: [wiki.lfai.foundation](https://wiki.lfai.foundation)  
GitHub: [github.com/lfai](https://github.com/lfai)  
Landscape: [landscape.lfai.foundation](https://landscape.lfai.foundation) or [l.lfai.foundation](https://l.lfai.foundation)  
Mail Lists: <https://lists.lfai.foundation>

LF AI Logos: <https://github.com/lfai/artwork/tree/master/lfai>

LF AI Presentation Template:

[https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk\\_-czASlz2GTBRZk2/view?usp=sharing](https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing)

Events Page on LF AI Website: <https://lfai.foundation/events/>

Events Calendar on LF AI Wiki (subscribe available):

<https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091544>

Event Wiki Pages: <https://wiki.lfai.foundation/display/DL/LF+AI+Foundation+Events>

# Agenda

- › Roll Call
- › Approval of Minutes
- › IBM Trusted AI: Project Incubation Proposal + TAC Vote
- › Montreal AI Ethics Institute: Invited Presentation
- › LF AI General Updates
- › Upcoming TAC Meetings
- › Open Discussion

# TAC Voting Members

<b>Member</b>	<b>Contact</b>	<b>Email</b>
AT&T	Reuben Klein	<a href="mailto:rk1518@att.com">rk1518@att.com</a>
Baidu	Daxiang Dong	<a href="mailto:dongdaxiang@baidu.com">dongdaxiang@baidu.com</a>
Ericsson	Rani Yadav-Ranjan	<a href="mailto:rani.yadav-ranjan@ericsson.com">rani.yadav-ranjan@ericsson.com</a>
Huawei	Huang Zhipeng	<a href="mailto:huangzhipeng@huawei.com">huangzhipeng@huawei.com</a>
Nokia	Pantelis Monogioudis	<a href="mailto:pantelis.monogioudis@nokia.com">pantelis.monogioudis@nokia.com</a>
Tech Mahindra	Nikunj Nirmal	<a href="mailto:nn006444@techmahindra.com">nn006444@techmahindra.com</a>
Tencent	Bruce Tao	<a href="mailto:brucetao@tencent.com">brucetao@tencent.com</a>
Zilliz	Jun Gu	<a href="mailto:jun.gu@zilliz.com">jun.gu@zilliz.com</a>
ZTE	Wei Meng	<a href="mailto:meng.wei2@zte.com.cn">meng.wei2@zte.com.cn</a>
Acumos AI Project	Nat Subramanian	<a href="mailto:natarajan.subramanian@techmahindra.com">natarajan.subramanian@techmahindra.com</a>
Angel Project	Bruce Tao	<a href="mailto:brucetao@tencent.com">brucetao@tencent.com</a>
ONNX Project	Jim Spohrer*	<a href="mailto:spohrer@us.ibm.com">spohrer@us.ibm.com</a>

\* TAC Chairperson

# Approval of Minutes

Draft minutes from the June 4th meeting of the TAC were previously distributed to the TAC members

## **Proposed Resolution:**

- › That the minutes of the June 4th meeting of the Technical Advisory Council of the LF AI Foundation are hereby approved

# Project Contribution Proposal: IBM Trusted AI Projects

AI Fairness 360

Adversarial Robustness 360

AI Explainability 360



# Project Contribution Proposal: Review & Discussion

## IBM Trusted AI Projects

**IBM Trusted AI Projects are inclusive of three open sourced state of the art, trusted AI toolkits:**

- › AI Fairness 360 <https://github.com/IBM/AIF360>
- › Adversarial Robustness 360 <https://github.com/IBM/adversarial-robustness-toolbox>
- › AI Explainability 360 <https://github.com/IBM/AIX360>

**We are proposing to host all 3 projects as incubation projects under LF AI.**

- › **Projects Level:** Incubation
- › **Presenters:** Animesh Singh, Mathieu Sinn, and Mike Hind
- › **Proposal details:**  
<https://github.com/animeshsingh/proposing-projects/blob/trusted-ai/proposals/trusted-ai.adoc>

# Agenda

- › [Proposal](#) (15 mins)
- › Q&A (10 mins)
- › Vote (2 mins)



[Animesh Singh](#)

IBM Open Technologies  
Co-Chair of the LF AI  
[Trusted AI Committee](#)



[Mathieu Sinn](#)

IBM Research – Dublin  
[ART: Adversarial Robustness  
Toolbox](#)



[Mike Hind](#)

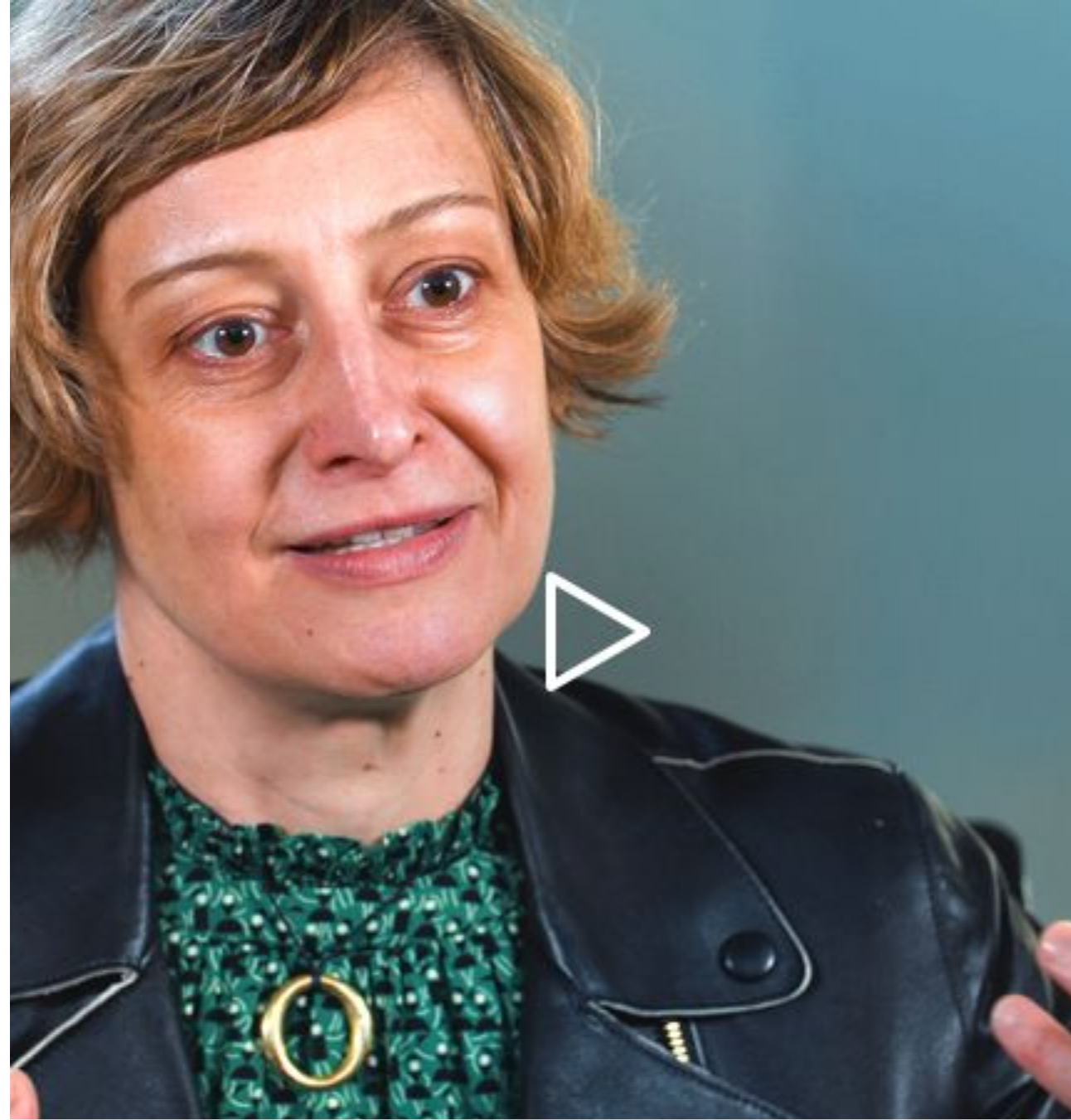
IBM Research – Yorktown  
[AIF360/AIX360](#): AI  
Fairness/Explainability

“Instrumenting trust into data sets and machine learning models will **accelerate the adoption of AI** and engender increased confidence in these general-purpose technologies.”

Aleksandra Mojsilovic

IBM Fellow

Head of Foundations of Trusted AI



“If we fail to make **ethical** and **inclusive** artificial intelligence we risk losing gains made in civil rights and gender equity under the guise of machine neutrality.”

Joy Buolamwini  
Gender Shades  
MIT Media Lab



# LFAI Trusted AI Committee

<https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee>

Bring Trust, Transparency and Responsibility into AI

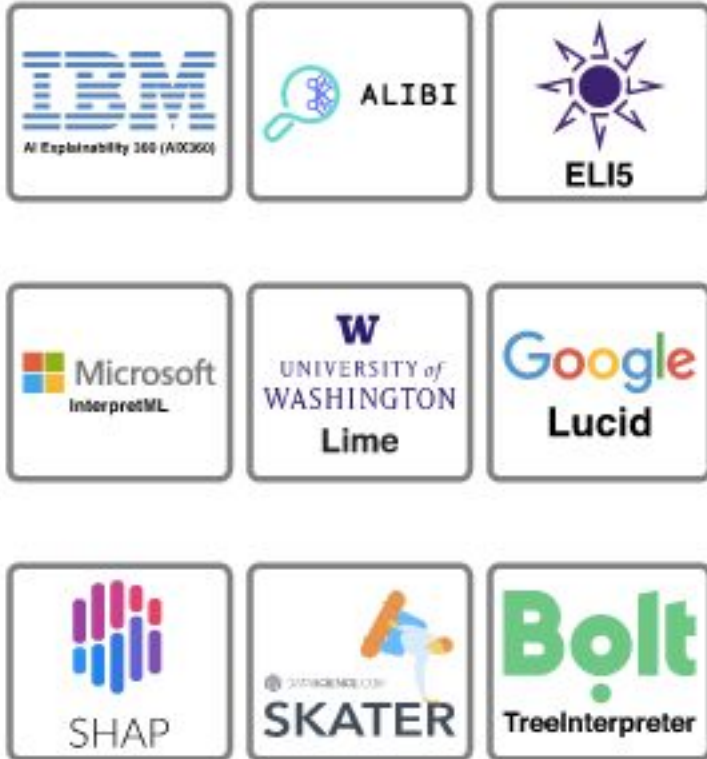
- ✓ Principles Working Group
- ✓ Technical Working Group

Chairs	Region	Company
Animesh Singh	North America	IBM
Souad Ouali	Europe	Orange
Jeff Cao	Asia	Tencent



Trusted & Responsible AI

Explainability



Adversarial



Bias & Fairness



# So what does it take to trust a decision made by a machine?

(Other than that it is 99% accurate)?



**Did anyone tamper with it?**



**Is it fair?**



**Is it easy to understand?**



**Is it accountable?**

# Our vision for Trusted AI

Pillars of trust, woven into the lifecycle of an AI application



**ROBUSTNESS**

**Did anyone  
tamper with it?**



**FAIRNESS**

**Is it fair?**



**EXPLAINABILITY**

**Is it easy to  
understand?**



**LINEAGE**

**Is it accountable?**



# Adversarial Robustness

## 360 ↳ (ART)

<https://github.com/IBM/adversarial-robustness-toolbox>

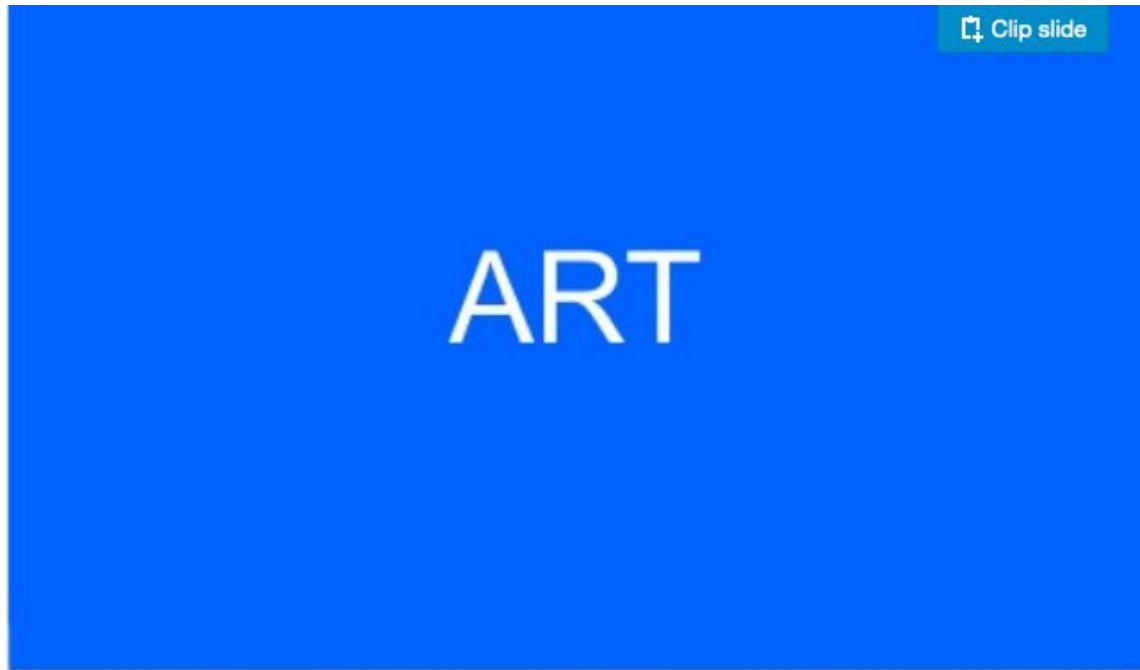
ART is a library dedicated to adversarial machine learning. Its purpose is to allow rapid crafting and analysis of **attack, defense and detection methods** for machine learning models. Applicable domains include finance, self driving vehicles etc.

The Adversarial Robustness Toolbox provides an implementation for many state-of-the-art methods for attacking and defending classifiers.

### Toolbox: Attacks, defenses, and metrics

- 30+ SOTA attacks (evasion, poisoning, extraction, inference)
- 25+ baseline defenses
- Modules for detection, metrics and certification

<https://art-demo.mybluemix.net/>

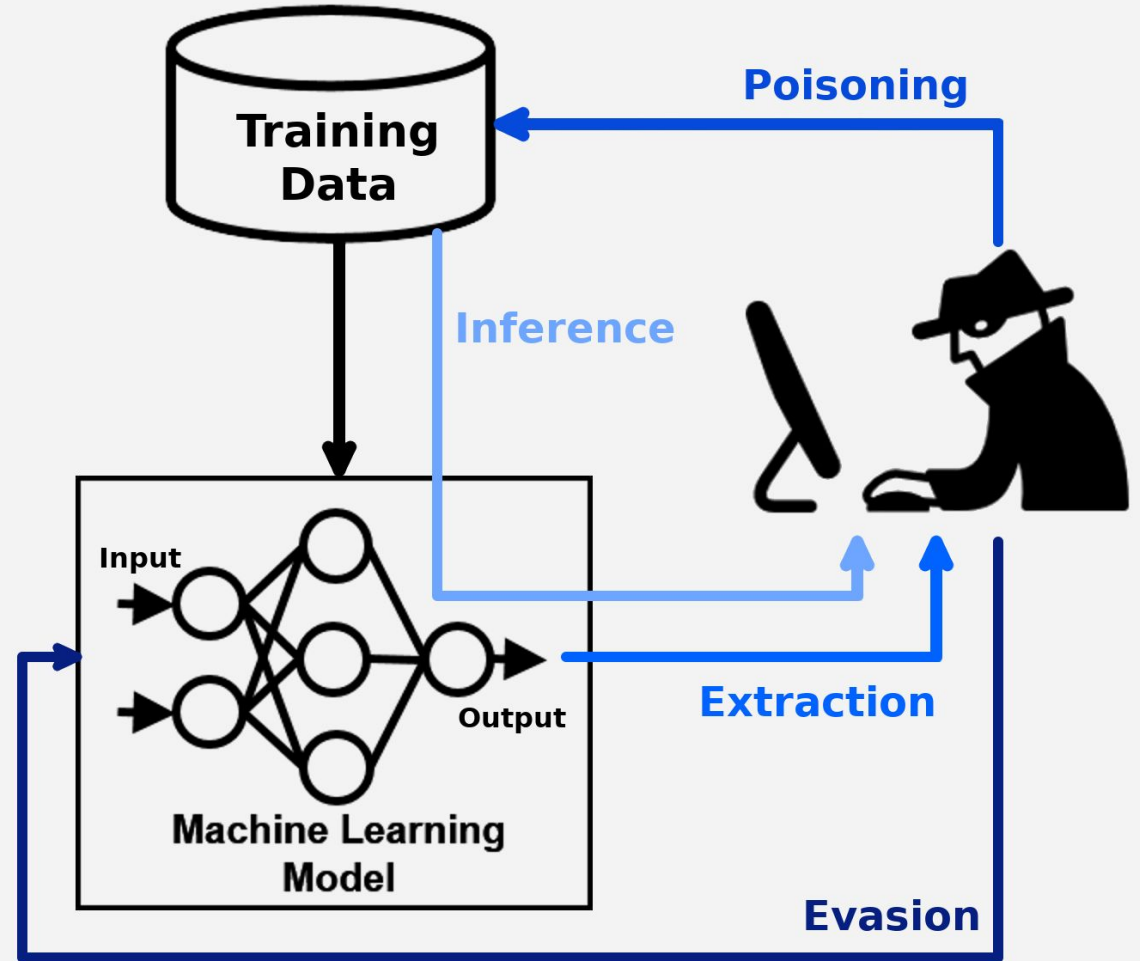


### Supported ML/DL frameworks:



# Adversarial Threats

- › Adversarial threats against machine learning models and applications have a wide variety of attack vectors.
  - › **Evasion:** Modifying input to influence model
  - › **Poisoning:** Modify training data to add backdoor
  - › **Extraction:** Steal a proprietary model
  - › **Inference:** Learn information on private data



# Real Adversarial Threats

## › Evasion.

- › Imperceptible modifications to medical images to influence classification.

## › Poisoning.

- › Imperceptible patterns in training data create backdoors that control models.

## › Extraction.

- › Theft of proprietary models through model queries.

## › Inference.

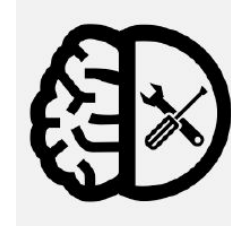
- › Derive properties of the model's training data up to identifying single data entries.

# Adversarial Threat Combinations

- › Combinations of adversarial threats become more effective than their sum.
  - › Extraction attacks enable stronger white-box evasion attacks
  - › Extraction attacks steal models that could leak more private information in inference attacks



# Adversarial Robustness Toolbox (ART)



› **ART is a Python library for machine learning security.**

 TensorFlow  Keras

 PYTORCH  mxnet

 scikit-learn

GPy

 dmlc  
XGBoost

LightGBM

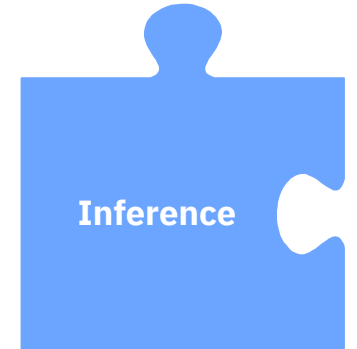
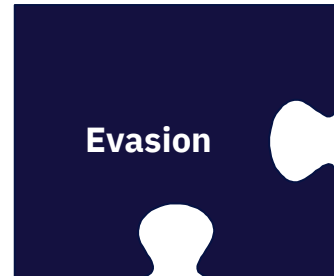
 CatBoost

- › [github.com/IBM/adversarial-robustness-toolbox](https://github.com/IBM/adversarial-robustness-toolbox)
- › 1500+ Stars (~500 in last 6 months)
- › providing tools to developers and researcher
- › Evaluating, Defending, Certifying and Verifying of machine learning models and applications
- › **All Tasks:** Classification, Object Detection, Generation, Encoding, Certification, etc.
- › **All Frameworks:** TensorFlow, Keras, PyTorch, MXNet, scikit-learn, XGBoost, LightGBM, CatBoost, GPy
- › **All Data:** images, tables, audio, video, etc.
- › Contributions and feedback are very welcome!

# The Tools of ART

## > ART 1.3

- **art.metrics**
  - Methods to quantify robustness
- **art.estimators**
  - Abstractions for models



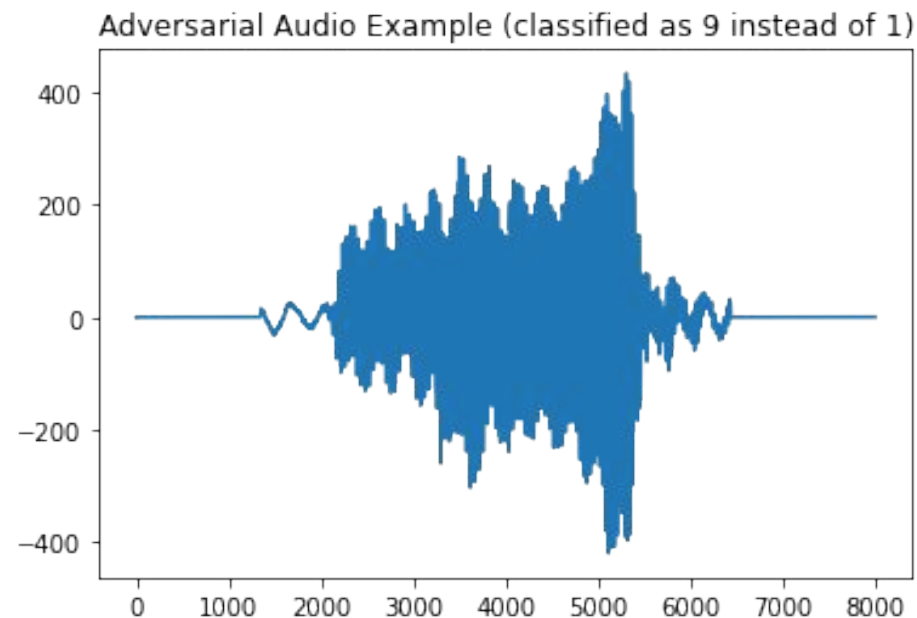
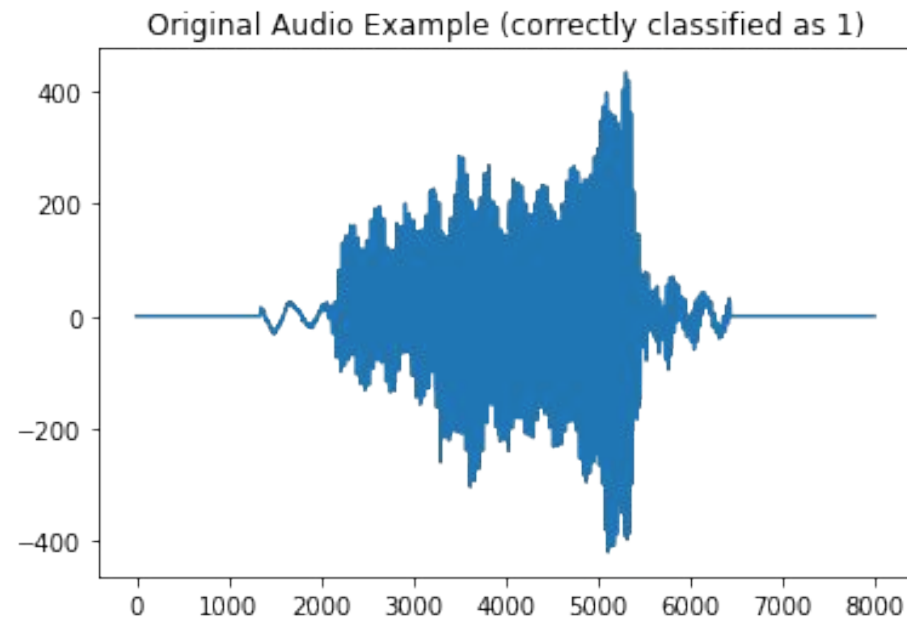
<p><b>art.attacks</b> examples</p>	<ul style="list-style-type: none"> <li>• 21 (+8)</li> <li>• White-box (e.g. FGSM, PGD, Carlini&amp;Wagner, ...)</li> <li>• Black-box (HopSkipJump, Boundary, ZOO, ...)</li> </ul>	<ul style="list-style-type: none"> <li>• 3 (+1)</li> <li>• Backdoor, Feature Collision, SVM, ...</li> </ul>	<ul style="list-style-type: none"> <li>• 3</li> <li>• FunctionallyEquivalent, KnockOffNets, CopyCat, ...</li> </ul>	<ul style="list-style-type: none"> <li>• 4 (+4)</li> <li>• Model Inversion (MIFace, ...)</li> <li>• Attribute Inference</li> </ul>
<p><b>art.defences</b> examples</p>	<ul style="list-style-type: none"> <li>• 15 (+4)</li> <li>• Adversarial Training (Madry, Fast is Better than Free, ...)</li> <li>• Preprocessing</li> <li>• Transformer</li> <li>• Detection</li> </ul>	<ul style="list-style-type: none"> <li>• 4 (+1)</li> <li>• Detection (Activation, Provenance, RONI, Spectral Signature, ...)</li> </ul>	<ul style="list-style-type: none"> <li>• 6</li> <li>• Postprocessing (Reverse Sigmoid, ...)</li> </ul>	<ul style="list-style-type: none"> <li>• DiffPrivLib</li> </ul>

# New Attacks and Defenses

- **Dpatch** (Liu et al., 2019)
  - Adversarial patches for object detectors
- **Shadow Attack** (Ghiasi et al., 2020)
  - Breaking/spoofing robustness certificates
- **Feature Adversaries** (Sabour et al., 2016)
  - Imitates feature representation of benign samples
- **Frame Saliency Attack** (Inkawhich et al., 2018)
  - Attack on action recognition systems
- **Wasserstein Attack** (Wong et al., 2019)
  - Large but naturally looking perturbations
- **Auto Attack** (Croce and Hein, 2020)
  - Multiple white- and black-box attacks optimized for achieving state-of-the-art robustness evaluation performance of leading experts completely automated
- **Auto-PGD** (Croce and Hein, 2020)
  - multiple attack losses and automated learning rate adjustment
- **Square Attack** (Croce and Hein, 2019)
  - very efficient black-box attack based on random search
- **DefenseGAN** (Samangouei et al., 2018), **InvGAN** (Lin et al., 2019)
  - Defense based on Generative Adversarial Networks (GAN)
- **MP3 compression, resampling** (Carlini et al., 2018)
- **MPEG compression, frame-wise JPEG and spatial smoothing**
- **Fast is Better than Free** (Wong et al., 2019)
  - Fastest adversarial training protocol

# ART Audio Example

- 
- **Speech Classification**
  - Application of ART to Speech classification
  - Dataset: Audio-MNIST, spoken digits [0-9] with multiple speakers
  - Baseline for evaluating defenses against evasion on audio data
  - Starting point for ART towards speech recognition and sequence-to-sequence models
  - [https://github.com/IBM/adversarial-robustness-toolbox/blob/master/notebooks/adversarial\\_audio\\_examples.ipynb](https://github.com/IBM/adversarial-robustness-toolbox/blob/master/notebooks/adversarial_audio_examples.ipynb)





# AI Fairness 360

↳ (AIF360)

<https://github.com/IBM/AIF360>

AIF360 toolkit is an open-source library to help detect and remove bias in machine learning models. **AIF360 translates algorithmic research from the lab into practice.** Applicable domains include finance, human capital management, healthcare, and education.

The AI Fairness 360 Python package includes a comprehensive set of metrics for datasets and models to test for biases, explanations for these metrics, and algorithms to mitigate bias in datasets and models.

## Toolbox

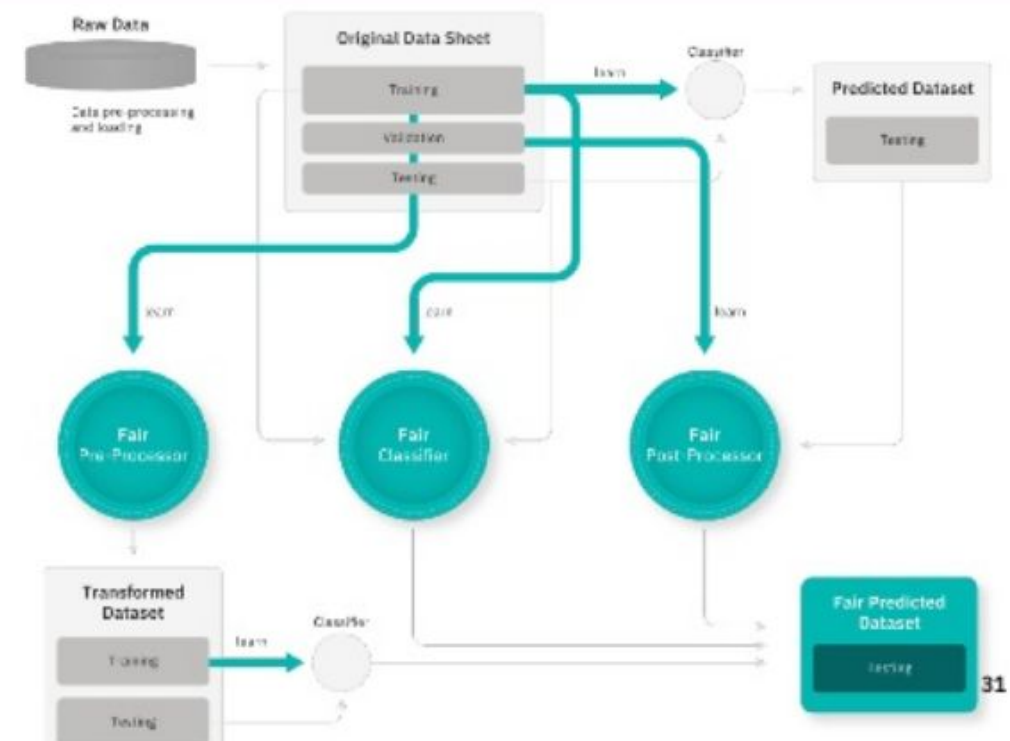
Fairness metrics (70+)

Fairness metric explanations

Bias mitigation algorithms (10+)

<http://aif360.mybluemix.net/>

# AIF360



# AI Explainability 360 ↳ (AIX360)

<https://github.com/IBM/AIX360>

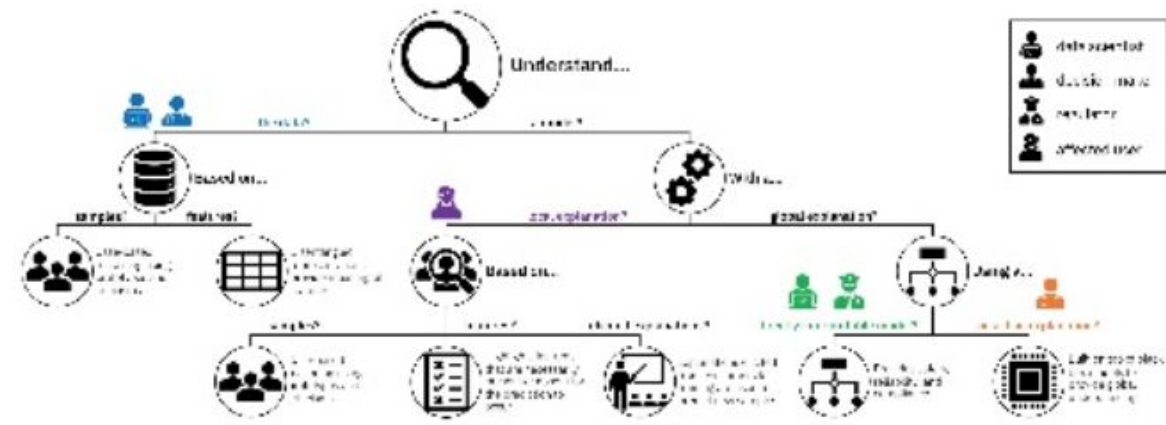
AIX360 toolkit is an open-source library to help explain AI and machine learning models and their predictions. This includes three classes of algorithms: local post-hoc, global post-hoc, and directly interpretable explainers for models that use image, text, and structured/tabular data.

The AI Explainability360 Python package includes a comprehensive set of explainers, both at global and local level.

### Toolbox

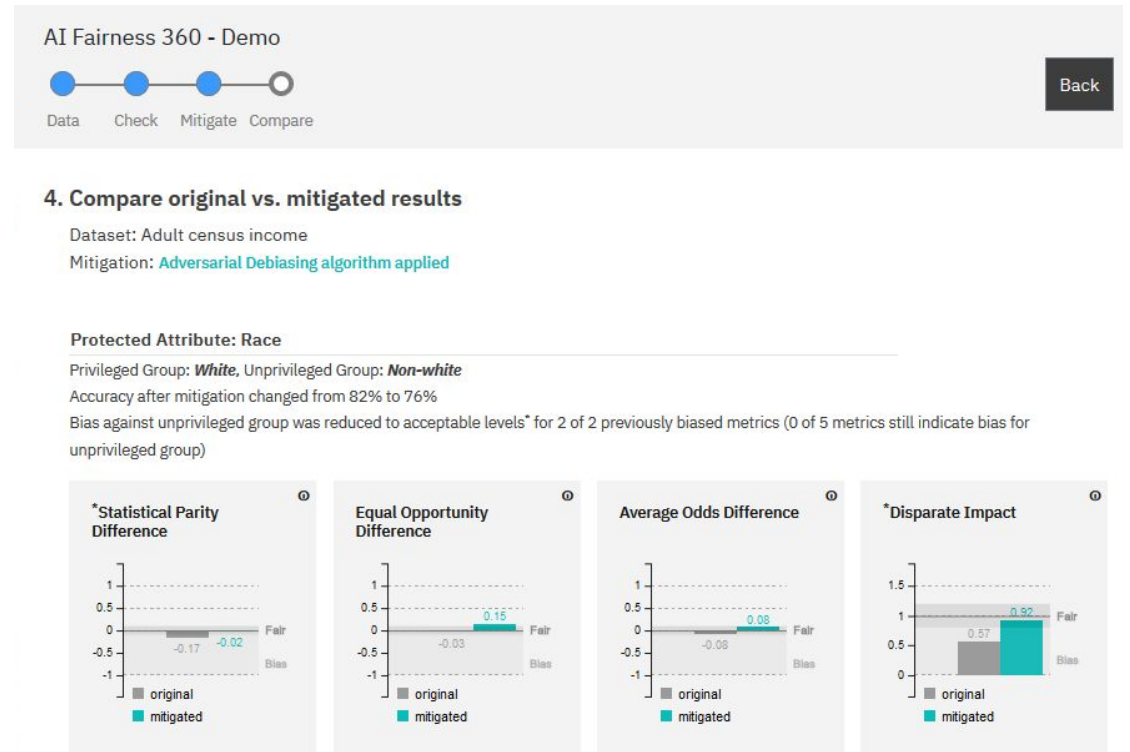
- Local post-hoc
- Global post-hoc
- Directly interpretable

<http://aix360.mybluemix.net>



Most comprehensive **open source** toolkit for detecting & mitigating bias in ML models:

- 70+ fairness metrics
- 10 bias mitigators
- Interactive demo illustrating 5 bias metrics and 4 bias mitigators
- extensive industry tutorials and notebooks



[aif360.mybluemix.net](http://aif360.mybluemix.net)

# AI Fairness 360

aif360.mybluemix.net

IBM Research Trusted AI

[Home](#)

[Demo](#)

[Resources](#)

[Events](#)

[Videos](#)

[Community](#)

## AI Fairness 360 Open Source Toolkit

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. Containing over 70 fairness metrics and 10 state-of-the-art bias mitigation algorithms developed by the research community, it is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education. We invite you to use it and improve it.

[API Docs](#)

[Get Code](#)

Not sure what to do first? Start here!

### Read More

Learn more about fairness and bias mitigation concepts, terminology, and tools before you begin.



### Try a Web Demo

Step through the process of checking and remediating bias in an interactive web demo that shows a sample of capabilities available in this toolkit.



### Watch Videos

Watch videos to learn more about AI Fairness 360.



### Read a paper

Read a paper describing how we designed AI Fairness 360.



### Use Tutorials

Step through a set of in-depth examples that introduces developers to code that checks and mitigates bias in different industry and application domains.



### Ask a Question

Join our AIF360 Slack Channel to ask questions, make comments and tell stories about how you use the toolkit.



### View Notebooks

Open a directory of Jupyter Notebooks in GitHub that provide working examples of bias detection and mitigation in sample datasets. Then share your own notebooks!



### Contribute

You can add new metrics and algorithms in GitHub. Share Jupyter notebooks showcasing how you have examined and mitigated bias in your machine learning application.



Learn how to put this toolkit to work for your application or industry problem. Try these tutorials.

### Credit Scoring

See how to detect and mitigate age bias in predictions of credit-worthiness using the German Credit dataset.



### Medical Expenditure

See how to detect and mitigate racial bias in a care management scenario using Medical Expenditure Panel Survey data.



**Designed to translate new research from the lab to industry practitioners: tutorials, education, glossary, resources.**

# Three categories of bias mitigation algorithms

**Pre-processing algorithm** – a bias mitigation algorithm that is applied to training data

**In-processing algorithm** – a bias mitigation algorithm that is applied to a model during its training

**Post-processing algorithm** – a bias mitigation algorithm that is applied to predicted labels

The choice among algorithm categories can partially be made based on the user persona's ability to intervene at different parts of a machine learning pipeline.

If the user is allowed to modify the training data, then pre-processing can be used.

If the user is allowed to change the learning algorithm, then in-processing can be used.

If the user can only treat the learned model as a black box without any ability to modify the training data or learning algorithm, then only post-processing can be used.

# AI Explainability 360

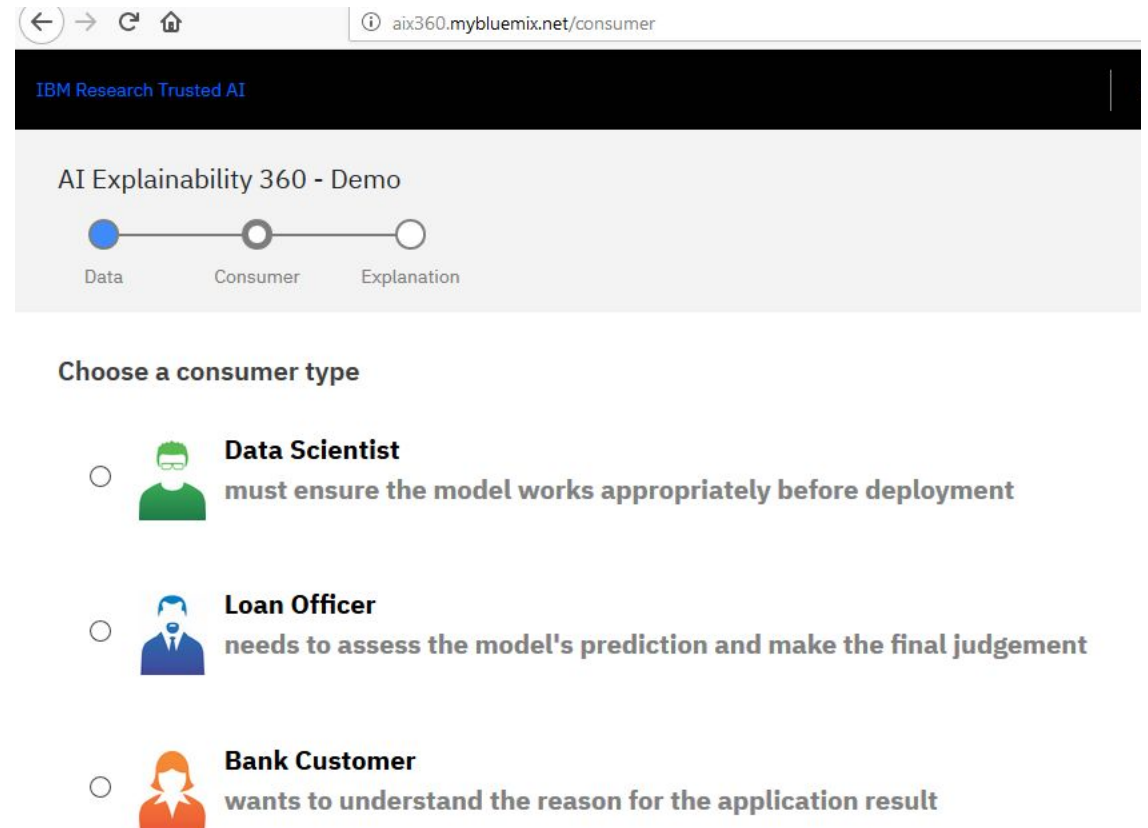
The most comprehensive **open source** toolkit for explaining ML models and data:

- 10 innovated algorithms to explain data and AI models + 2 metrics
- An interactive demo that provides a gentle introduction through a credit scoring application
- 13 tutorial notebooks covering use cases in finance, healthcare, lifestyle, retention, etc.
- documentation that guides the practitioner on choosing an appropriate explanation method.

***One Explanation Does Not Fit All:  
A Toolkit and Taxonomy of AI  
Explainability Techniques***

*by Arya et al.*

<https://arxiv.org/abs/1909.03012>



The screenshot shows a web browser window with the URL `aix360.mybluemix.net/consumer`. The page header is "IBM Research Trusted AI". The main content area is titled "AI Explainability 360 - Demo" and features a progress indicator with three steps: "Data" (selected), "Consumer", and "Explanation". Below this, the user is prompted to "Choose a consumer type" with three radio button options:

- Data Scientist**  
must ensure the model works appropriately before deployment
- Loan Officer**  
needs to assess the model's prediction and make the final judgement
- Bank Customer**  
wants to understand the reason for the application result

<http://aix360.mybluemix.net/>

# AI Explainability 360

aix360.mybluemix.net

90%



- [Home](#)
- [Demo](#)
- [Resources](#)
- [Events](#)
- [Videos](#)
- [Community](#)

## AI Explainability 360 Open Source Toolkit

This extensible open source toolkit can help you comprehend how machine learning models predict labels by various means throughout the AI application lifecycle. Containing eight state-of-the-art algorithms for interpretable machine learning as well as metrics for explainability, it is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education. We invite you to use it and improve it.

- [API Docs](#)
- [Get Code](#)

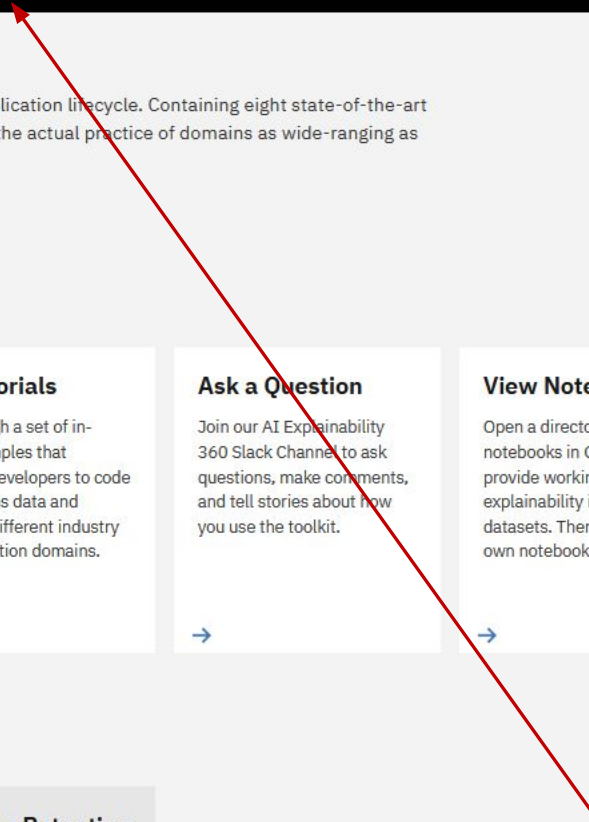
Not sure what to do first? Start here!

- Read More**  
Learn more about explainability concepts, terminology, and tools before you begin.  
→
- Try a Web Demo**  
Step through the process of explaining models to consumers with different personas in an interactive web demo that shows a sample of capabilities available in this toolkit.  
→
- Watch Videos**  
Watch videos to learn more about AI Explainability 360 toolkit.  
→
- Read a Paper**  
Read a paper describing how we designed AI Explainability 360 toolkit.  
→
- Use Tutorials**  
Step through a set of in-depth examples that introduce developers to code that explains data and models in different industry and application domains.  
→
- Ask a Question**  
Join our AI Explainability 360 Slack Channel to ask questions, make comments, and tell stories about how you use the toolkit.  
→
- View Notebooks**  
Open a directory of Jupyter notebooks in GitHub that provide working examples of explainability in sample datasets. Then share your own notebooks!  
→
- Contribute**  
You can add new algorithms and metrics in GitHub. Share Jupyter notebooks showcasing how you have enabled explanations in your machine learning application.  
→

Learn how to put this toolkit to work for your application or industry problem. Try these tutorials.

- Credit Approval**  
See how to explain credit approval models using the FICO Explainable Machine Learning Challenge dataset.  
→
- Medical Expenditure**  
See how to create interpretable machine learning models in a care management scenario using Medical Expenditure Panel Survey data.  
→
- Dermoscopy**  
See how to explain dermoscopic image datasets used to train machine learning models that help physicians diagnose skin diseases.  
→
- Health and Nutrition Survey**  
See how to quickly understand the National Health and Nutrition Examination Survey datasets to hasten research in epidemiology and health policy.  
→
- Proactive Retention**  
See how to explain predictions of a model that recommends employees for retention actions from a synthesized human resources dataset.  
→

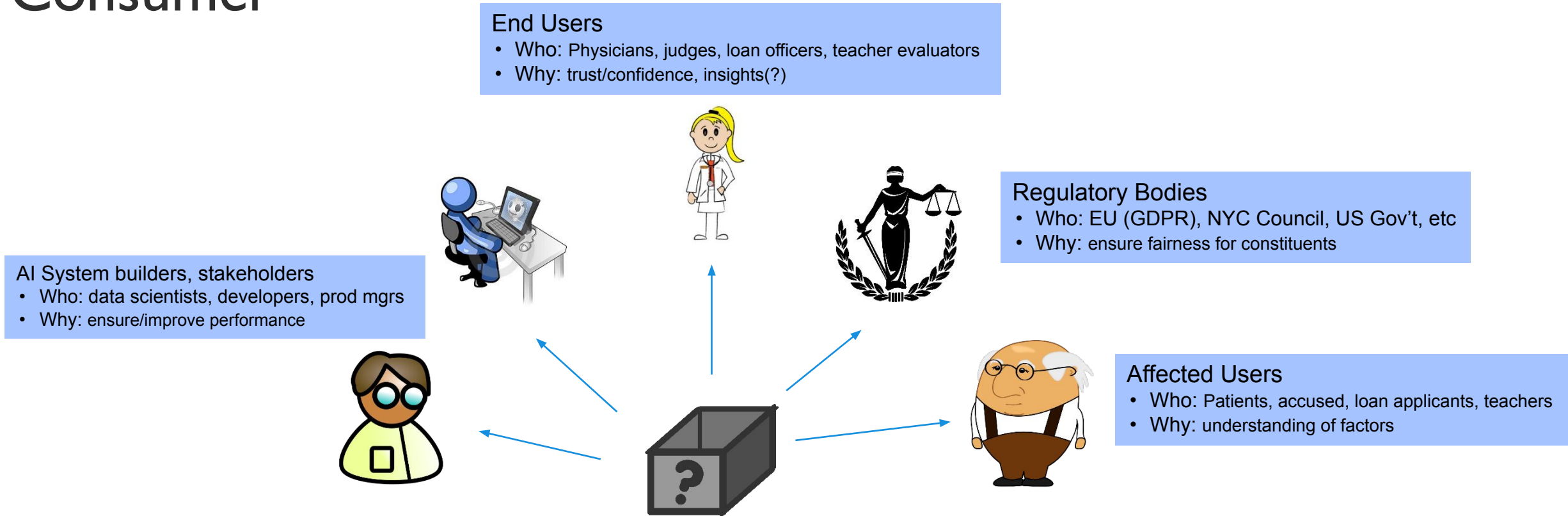
**Designed to translate new research from the lab to industry practitioners: tutorials, education, glossary, resources.**







# Meaningful Explanations Depend on the Explanation Consumer



Must match the **complexity capability** of the consumer  
Must match the **domain knowledge** of the consumer

*“We couldn’t explain the model to them because they didn’t have the training in machine learning.” Nautilus, Sept 2016*

# Review of the proposal on GH

<https://github.com/animeshsingh/proposing-projects/blob/trusted-ai/proposals/trusted-ai.adoc>

# Open Discussion

AI Fairness 360

Adversarial Robustness 360

AI Explainability 360

# TAC Vote on Project Proposal: IBM Trusted AI Projects

## **Proposed Resolution:**

The TAC approves the IBM Trusted AI Projects (AI Fairness 360, Adversarial Robustness 360, and AI Explainability 360) as an Incubation projects of the LF AI Foundation

# Next Steps

LF AI staff will work with IBM on onboarding the project leading to the announcement on the projects joining LF AI

Explore potential integrations between this project and other LF AI projects

Integrate the projects with LF AI operations

# Guest Presentation: Montreal AI Ethics Institute

## Abhishek Gupta

The logo for MAIEI (The Montreal AI Ethics Institute) is displayed on a black square background. The letters 'MAIEI' are rendered in a white, sans-serif font. The letter 'E' is stylized with a blue horizontal bar at its top and bottom, and a blue vertical bar on its right side, giving it a digital or circuit-like appearance.

# The Montreal AI Ethics Institute

*Presentation to The LF AI Foundation at The Linux Foundation  
June 18, 2020*

Abhishek Gupta, Founder  
[abhishek@montrealaiethics.ai](mailto:abhishek@montrealaiethics.ai)



<https://montrealaiethics.ai/>



@mtlaiethics



<https://aiethics.substack.com>

# About MAIEI

The logo for MAIEI is displayed on a black square background. The letters 'MAIEI' are in a white, sans-serif font. The letter 'E' is stylized with a blue-to-cyan gradient and a slight shadow effect.

MAIEI

---

The Montreal AI Ethics Institute (MAIEI) is an international, non-profit research institute dedicated to **defining humanity's place in a world increasingly characterized and driven by algorithms.**

We do this by creating tangible and applied technical and policy research in the ethical, safe, and inclusive development of AI.





# About MAIEI

The logo for MAIEI is displayed on a black square background. The letters 'MAIEI' are in white, with the 'E' being a stylized blue character with three horizontal bars.

---

Our goal is to **build public competence and understanding of the societal impacts of AI** and to equip and empower diverse stakeholders to actively engage in the shaping of technical and policy measures in the development and deployment of AI systems.

We are a **digital-first civil society organization** that brings together a diversity of individuals from different disciplines, areas of expertise, and geographic regions.



# Our Approach

We bridge the gap between technical and policy experience with real-world impact

## Biased Towards Action

While there are numerous organizations that have been working on sets of principles, frameworks, and other theoretical guidelines, the missing piece that is now starting to surface is the bridging of the gap between the proposed technical and policy measures and operationalizing them. We are firmly biased towards action and our work with partner organizations has not only created positive change but also one that is sustainable and transformative..

## Frameworks into Practice

Our team combines deep technical, policy, and design expertise with years of experience working with organizations from across the world in putting these frameworks into practice. Given our global network of interdisciplinary researchers and practitioners coupled with an in-depth and all-encompassing view of the cutting-edge responsible AI landscape, we are uniquely positioned to bring about a quick turnover of research into applied measures.

## Responsible AI as the Norm

Organizational change starts with people and people require knowledge that is presented in bite-sized, accessible chunks. Our experience in delivering content that meets these criteria has a proven track record of success. We can leverage our combined expertise to create bespoke experiences that will equip and empower individuals with the necessary skills to confidently lead their organization into a future where responsible AI becomes the norm rather than the exception.



# Our Approach

We bridge the gap between technical and policy experience with real-world impact

## Empowering Local Champions

We are creating local champions in the form of informed and engaged citizens who are able to take this knowledge of applied AI ethics to their communities and organizations, thus scaling the impact that we have as a single organization. As an example, a former research intern that worked with MAIEI in 2019 is now the Head of AI Ethics Policy for the Joint Artificial Intelligence Center, Department of Defense, US Government.

## Truly Inclusive, Global Participation

Our programs are truly inclusive and eliminate barriers for people from all parts of the world, including the Global South, who are typically not able to access similar programs because of financial constraints, visa troubles, family commitments, and so on. By being digital-first, we are able to bring together perspectives that are otherwise inaccessible where the emphasis is oftentimes on heavy credentials and traditional backgrounds.

## Open Source and Open Access

Open source and open access models are embedded into everything we do. This includes deeply researched content for governments and other public entities, made available for all researchers and practitioners so that they can build on our work rather than having to reinvent the wheel.



# Our Programs



## Public Policy Consultations

... through our **Meetups** with the following partners and organizations, in person and online, national and global in scale:

- Australian Human Rights Commission
- European Commission
- G7 Multi Stakeholder Conference on Artificial Intelligence
- Government of Scotland
- Office of the Privacy Commissioner of Canada (OPCC)
- Partnership on AI
- Prime Minister's Office of New Zealand
- World Economic Forum; and others.



## Research Projects

... **Topics** include::

- Comprehensiveness of Archives: A modern AI-enabled approach to building comprehensive shared cultural heritage
- Exploring the uncanny valley of climate change misinformation
- Folding IN the margins: Building inclusive AI systems using indigenous data
- SECure - Social and Environmental Certificate for AI systems
- Participatory Design as a mechanism for building trustworthy AI
- Participatory Design to build better contact- and proximity-tracing apps trust: the critical pillar of society; and others.



# Our Programs



## Learning Communities

... on **Slack** and **Zoom** that meet regularly every two weeks on the following topics:

1. Complex Systems Theory
2. Disinformation
3. Labor Impacts of AI
4. Machine Learning Security
5. Privacy



## Curriculum Design

... designing and delivering **education training programs** for law and policy on the ethical, social and regulatory implications of an organization's AI strategy.



## Inclusive Community Building

... leveraging important values such as diversity and inclusion and offering access to a **community of over 2600+ members** with whom we have hosted over 45 workshops.



# Examples of Past Meetups



**The Future of Education  
(Part 2)**  
(Nov 15, 2018)

Host: **Microsoft**  
Attendees: 178



**Discrimination in the  
Systems – Gender, Race,  
Class, & Power**  
(November 29, 2019)

Host: **McGill University**  
Attendees: 468



**AI Ethics: Public Consultation On  
Scotland’s AI Strategy**  
(May 4, 2020)

Event held online  
Attendees: 61



**AI Ethics: Publication Norms for  
Responsible AI (Part 1)**  
(May 13, 2020)

Event held online  
Co-host: **Partnership on AI**  
Attendees: 147



# The Meetups

...and why they are so effective



### Working Together

They offer participants a way to learn and work together, and leverage diverse and global expertise



### Solution-oriented

They allow participants to pool together insights and look for impactful solutions



### Meaningful Feedback

For organizations that have CFPs out, the meetups offer an opportunity to receive more comprehensive feedback compared to the responses that they might get from single individuals



### Inclusion & Diversity

They bring people together from all over the world and give them access to the Ethics of AI ecosystem



# Partners We've Worked With

## Government

- Australian Human Rights Commission
- European Commission
- G7 Multi Stakeholder Conference on Artificial Intelligence
- Government of Scotland
- Office of the Privacy Commissioner of Canada (OPCC)
- Prime Minister's Office, New Zealand
- Treasury Board Secretariat, Canada

## Corporations

- ABB
- ARUP
- Deloitte
- Element AI
- Espace CDPQ
- Expedia
- EY
- Fasken
- Lightspeed
- Maluuba
- Microsoft
- OVH
- PwC
- SAP
- Shopify
- Stradigi AI

## Nonprofits

- Acorn Aspirations
- AI Global
- Alberta Machine Intelligence Institute (AMII)
- DEFCON AI Village
- **LF AI Foundation at The Linux Foundation**
- Mechanism Design for Social Good (MD4SG)
- ML Retrospectives
- Montreal International
- Montreal Neurological Institute (MNI)
- Montreal NewTech
- Mozilla Foundation
- NeurIPS
- UpstartED

## Academia

- College Ste-Marcelline
- Concordia University / District 3
- Dawson College
- Goethe University / Frankfurt Big Data Lab
- McGill University / Dobson Centre for Entrepreneurship, School of Continuing Studies, Building 21
- MILA
- Northwest Commission on Colleges and Universities (NWCCU)
- OCAD U
- Oxford Internet Institute
- Université de Montréal

## Multilateral Organizations

- The Banff Forum
- International Network for Government Science Advice
- Partnership on AI
- United Nations / AI for Good Global Summit
- World Economic Forum





# Meet The Team

We are applied researchers and practitioners with technical and policy experience in operationalizing responsible AI.



**Abhishek Gupta**  
Founder



**Renjie Butalid**  
Co-founder



**Tania DeGasperis**  
Associate



**Mo Akif**  
Director of Communications



**Marianna Ganapini**  
Researcher



**Camylle Lanteigne**  
Researcher



**Allison Cohen**  
Researcher



**Victoria Heath**  
Researcher



The logo for MAIEI is displayed within a black square. The letters 'MAIEI' are rendered in a white, sans-serif font. The letter 'E' is stylized with a blue horizontal bar on its right side.

Let's work together!

*Contact:*

Abhishek Gupta, Founder

[abhishek@montrealaiethics.ai](mailto:abhishek@montrealaiethics.ai)



<https://montrealaiethics.ai/>



@mtlaiethics






<https://aiethics.substack.com>

# LF AI General Updates











# Project Updates

# LF AI Hosted Projects

## Graduation

 <b>Acumos</b> ★ 10 LF Artificial Intelligence Foundation	 <b>Angel</b> ★ 5,825 Angel-ML LF Artificial Intelligence Foundation	 <b>ONNX</b> ★ 8,424 LF Artificial Intelligence Foundation	
--	---	---	--

## Incubation

 <b>Adlik</b> ★ 102 LF Artificial Intelligence Foundation	 <b>EDL</b> Elastic Deep Learning ★ 76 LF Artificial Intelligence Foundation	 <b>ForestFlow</b> ★ 16 LF Artificial Intelligence Foundation	 <b>Horovod</b> ★ 9,297 LF Artificial Intelligence Foundation	 <b>Ludwig</b> ★ 6,706 Uber MCap: \$60.39B
 <b>MARQUEZ</b> ★ 285 The We Company Funding: \$19.55B	 <b>Milvus</b> ★ 3,372 LF Artificial Intelligence Foundation	 <b>NNStreamer</b> ★ 212 LF Artificial Intelligence Foundation	 <b>Pyro</b> ★ 6,241 LF Artificial Intelligence Foundation	 <b>sparklyr</b> ★ 734 LF Artificial Intelligence Foundation

# Upcoming Releases

For links to details on upcoming releases for LF AI hosted projects visit the [Technical Project Releases wiki](#)

Project releases will be announced via a blog post and promoted on LF AI [Twitter](#) and/or [LinkedIn](#) social channels

If you are an LF AI hosted project and would like LF AI to promote your release, reach out to [pr@lfai.foundation](mailto:pr@lfai.foundation) to coordinate in advance (min 2 wks) of your expected release date. Please email [pr@lfai.foundation](mailto:pr@lfai.foundation) for more details and/or questions.

# Outreach Committee

# LF AI PR/Comms

- › Please follow LF AI on [Twitter](#) & [LinkedIn](#) and help amplify news via your social networks - Please retweet and share!
  - › Also watch for news updates via the tac-general mail list
  - › View recent announcement on the [LF AI Blog](#)
- › Open call to publish project/committee updates or other relevant content on the [LF AI Blog](#)
- › To discuss more details on participation or upcoming announcements, please email [pr@lfai.foundation](mailto:pr@lfai.foundation)



# Events

- › Upcoming Events
  - › Visit the [LF AI Events Calendar](#) or the [LF AI 2020 Events wiki](#) for a list of all events
  - › To participate visit the [LF AI 2020 Events wiki page](#) or email [info@lfai.foundation](mailto:info@lfai.foundation)
- › Please consider holding virtual events
  - › To discuss participation, please email [events@lfai.foundation](mailto:events@lfai.foundation)

# Call to Participate in Ongoing Efforts

# Trusted AI

- › **Leadership:**  
Animesh Singh (IBM), Souad Ouali (Orange), and Jeff Cao (Tencent)
- › **Goal:** Create policies, guidelines, tooling and use cases by industry
- › **Github:**  
<https://github.com/lfai/trusted-ai>
- › **Wiki:**  
<https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee>
- › **To participate:**  
<https://lists.lfai.foundation/g/trustedai-committee/>
- › **Next call:** Bi-weekly on Thursdays at 7am PT, subscribe to group calendar on wiki  
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091895>

# ML Workflow & Interop

- › **Leadership:**  
Huang “Howard” Zhipeng (Huawei)
- › **Goal:**  
Define an ML Workflow and promote cross project integration
- › **Wiki:**  
<https://wiki.lfai.foundation/display/DL/ML+Workflow+Committee>
- › **To participate:**  
<https://lists.lfai.foundation/g/mlworkflow-committee>
- › **Next call:** Every 4 weeks on Thursdays at 7:00 am PT, subscribe to group calendar on wiki  
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=18481242>

# Upcoming TAC Meetings

# Upcoming TAC Meetings

- › **July 2:** Cancelled due to OSS NA
- › **July 19:** [Mindspore Presentation](#)

Please send agenda topic requests to [tac-general@lists.lfai.foundation](mailto:tac-general@lists.lfai.foundation)

# TAC Meeting Details

- › To subscribe to the TAC Group Calendar, visit the wiki: <https://wiki.lfai.foundation/x/XQB2>
- › Join from PC, Mac, Linux, iOS or Android: <https://zoom.us/j/430697670>
- › Or iPhone one-tap:
  - › US: +16465588656,,430697670# or +16699006833,,430697670#
- › Or Telephone:
  - › Dial(for higher quality, dial a number based on your current location):
  - › US: +1 646 558 8656 or +1 669 900 6833 or +1 855 880 1246 (Toll Free) or +1 877 369 0926 (Toll Free)
- › Meeting ID: 430 697 670
- › International numbers available: <https://zoom.us/u/achYtcw7uN>

# Open Discussion

# Legal Notices

- › The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.
- › Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.
- › Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.
- › TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.
- › Facebook and the "f" logo are trademarks of Facebook or its affiliates.
- › LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.
- › YouTube and the YouTube icon are trademarks of YouTube or its affiliates.
- › All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.
- › The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>, each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.
- › Please email [legal@linuxfoundation.org](mailto:legal@linuxfoundation.org) with any questions about The Linux Foundation's policies or the notices set forth on this slide.