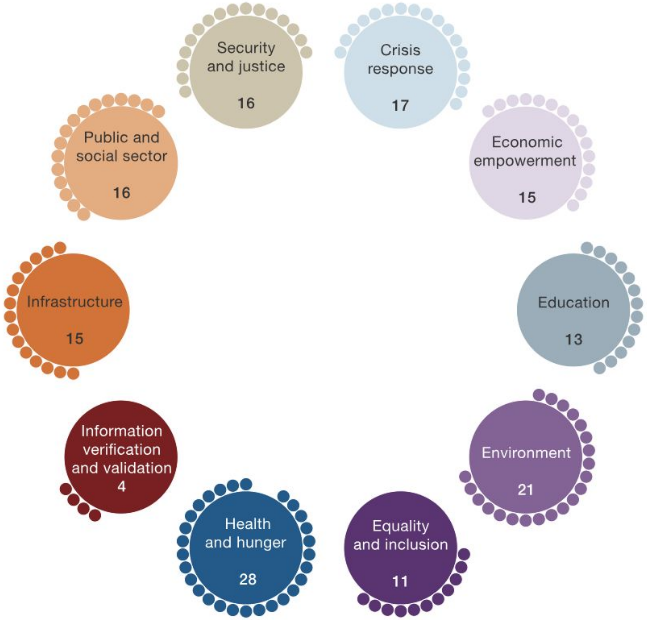
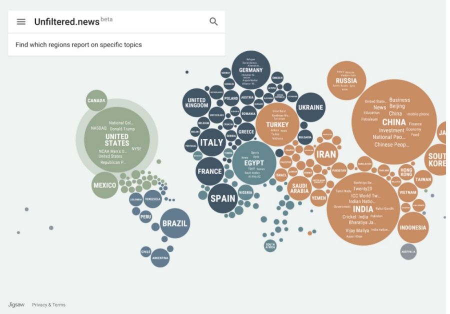
AI use cases per domain, number



Note: Our library of about 160 use cases is not comprehensive and will continue to evolve. This listing of the number of cases per domain should thus not be read as exhaustive.

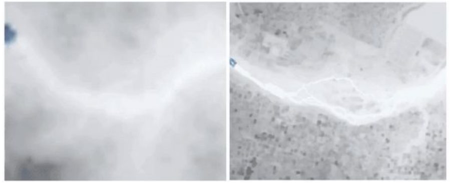
McKinsey&Company | **Source:** McKinsey Global Institute analysis

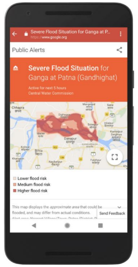

Unfiltered.news beta

Find which regions report on specific topics

'We keep seeing this correlation where you see spikes in attacks particularly at organizations that have really important information around things like elections or conflict in the world.'

—GEORGE CONRAD, PROJECT SHIELD




Severe Flood Situation for Ganga at P...
https://www.google.org

Public Alerts

Severe Flood Situation for Ganga at Patna (Gandhighat)

Active for next 5 hours
Central Water Commission

Lower flood risk
Medium flood risk
Higher flood risk

This map displays the *approximate area* that could be flooded, and may differ from actual conditions.

Send Feedback

Cell phones in Amazon trees alert rangers to illegal logging, record wildlife

Rainforest Connection is now leveraging machine learning to save rainforests

The New York Times

India Fights Diabetic Blindness With Help From A.I.

Google’s AI Principles

Google committed to seven principles which govern its development and deployment of AI. They state that technology should:

1. Be socially beneficial
2. Avoid creating or reinforcing unfair bias
3. Be built and tested for safety
4. Be accountable to people
5. Incorporate privacy design principles
6. Uphold high standards of scientific excellence
7. Be made available for uses that accord with these principles

AI Applications We Will Not Pursue

1

Technologies that cause or are likely to cause **overall harm**. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.

2

Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.

3

Technologies that gather or use information for **surveillance** violating internationally accepted norms.

4

Technologies whose purpose contravenes widely accepted principles of **international law and human rights**.

Responsible AI Practices

ai.google/education/responsible-ai-practices

Use a human-centered design approach — including consideration of appropriate disclosures; choice of outputs (eg: single vs multiple options); feedback from diverse users/use-cases

Identify multiple metrics to assess training and monitoring — including short and long term; false positives/negatives sliced across different subgroups

When possible, directly examine your raw data — including checks for mistakes; skewed training data; unrepresentative sampling

Understand the limitations of your dataset and model — eg: don’t use model built to detect correlation to imply causation; communicate limitations to users

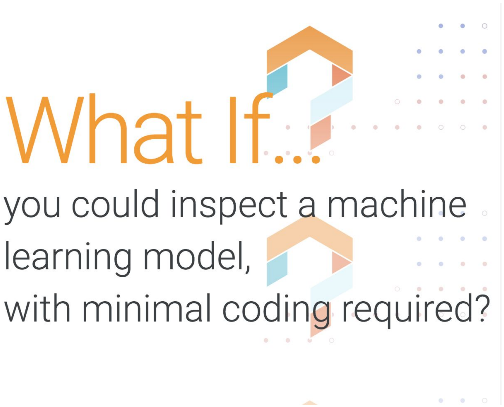
Test upfront and monitor after deployment — rigorous testing upfront; built-in quality checks; update model as needed, balancing short vs long term fixes

Tools

Facets — to aid understanding and analysis of machine learning datasets (ie. uncover bias across gender and race)

What if Tool — to explore model results without the need for writing code, providing a sense of which factors are most influential determining result.

Model and Data Cards — to reduce risk of models developed for one purpose being applied in contexts for which they are ill-suited.



What If...

you could inspect a machine learning model, with minimal coding required?

Google's AI Principles

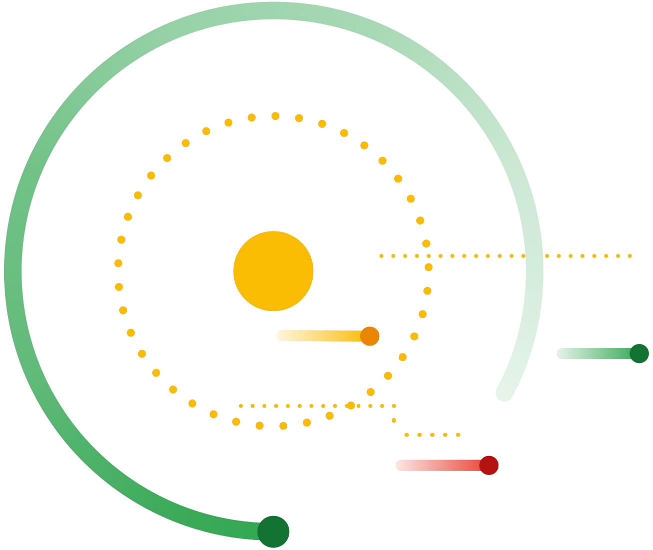
Google committed to seven principles which govern its development and deployment of AI. They state that technology should:

1. Be socially beneficial
2. Avoid creating or reinforcing unfair bias
3. Be built and tested for safety
4. Be accountable to people
5. Incorporate privacy design principles
6. Uphold high standards of scientific excellence
7. Be made available for uses that accord with these principles

Accountability Options

Suite of options for transparency & accountability

- Explainability
- Interpretability
- Auditability
- Testing and validation
- International Standards
- External engagement
- Contestability
- User Feedback



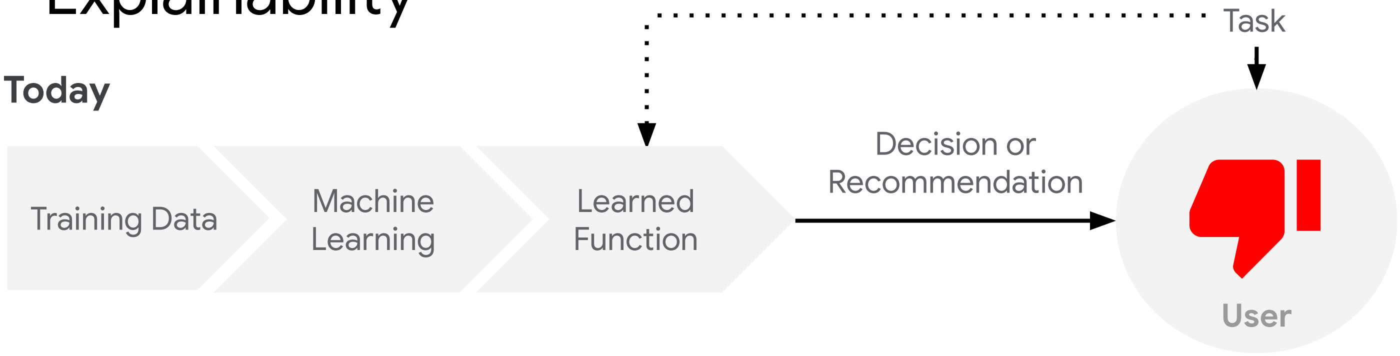
Explainability

Depends on:

- how **consequential** the decision/output is: e.g., medical diagnosis
- how much **trust** users have
- how **opaque** the decision-making process appears to be
- potential **legislation**: GDPR and beyond
- how **new/novel** the application is
- how much **agency** users have to contest output

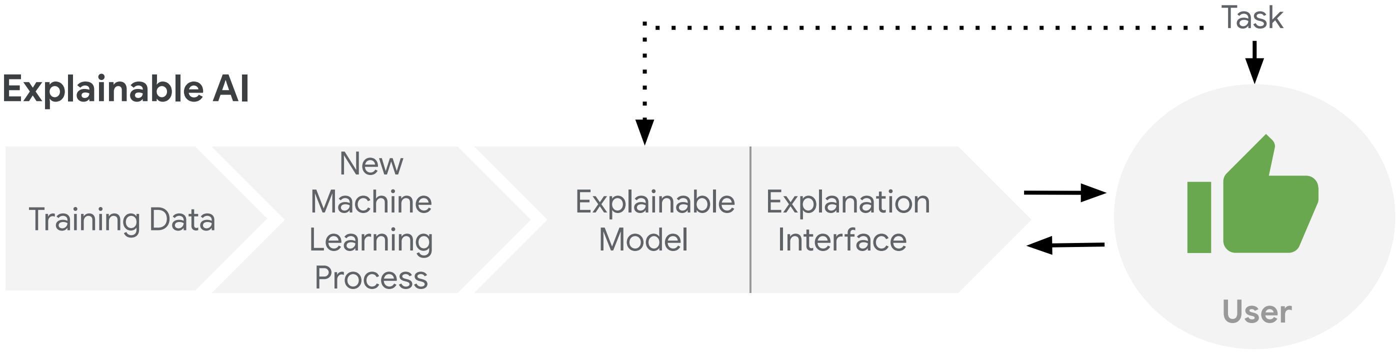
Explainability

Today



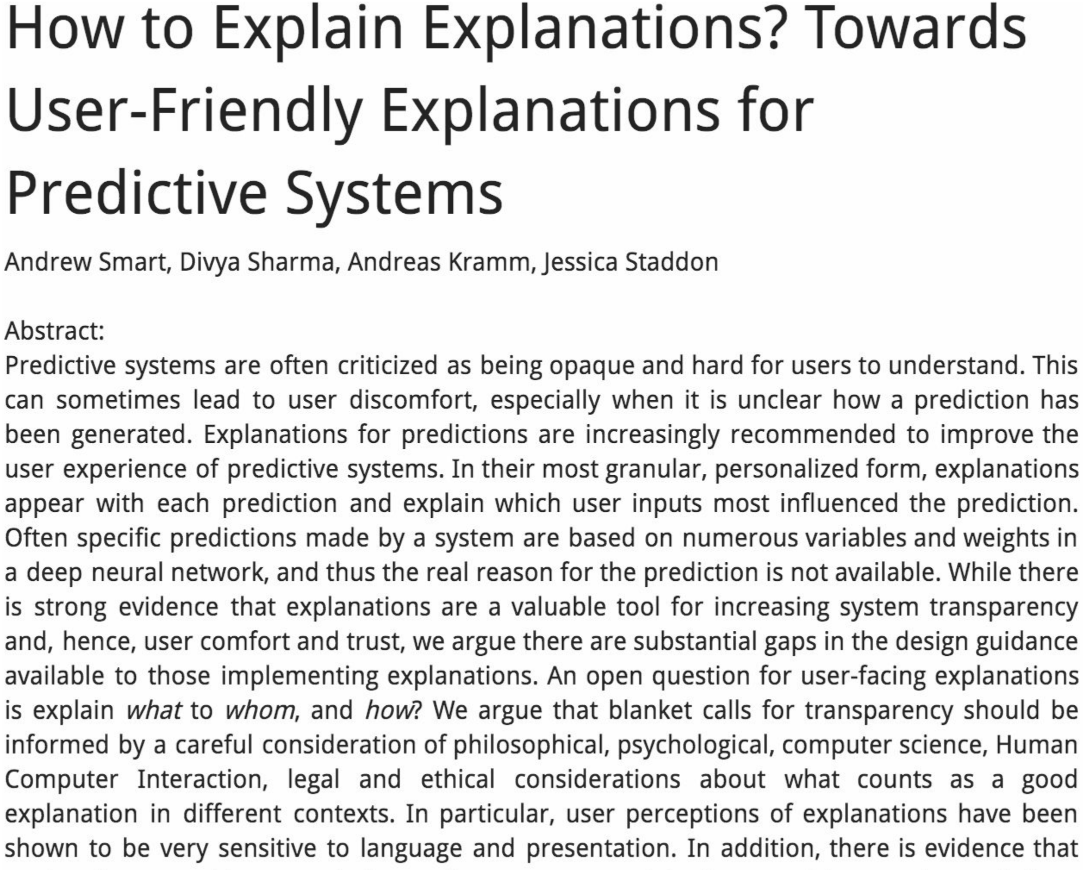
- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Explainable AI



- **I understand why**
- **I understand why not**
- **I know when you succeed**
- **I know when you fail**
- **I know when to trust you**
- **I know when you erred**

Explainability



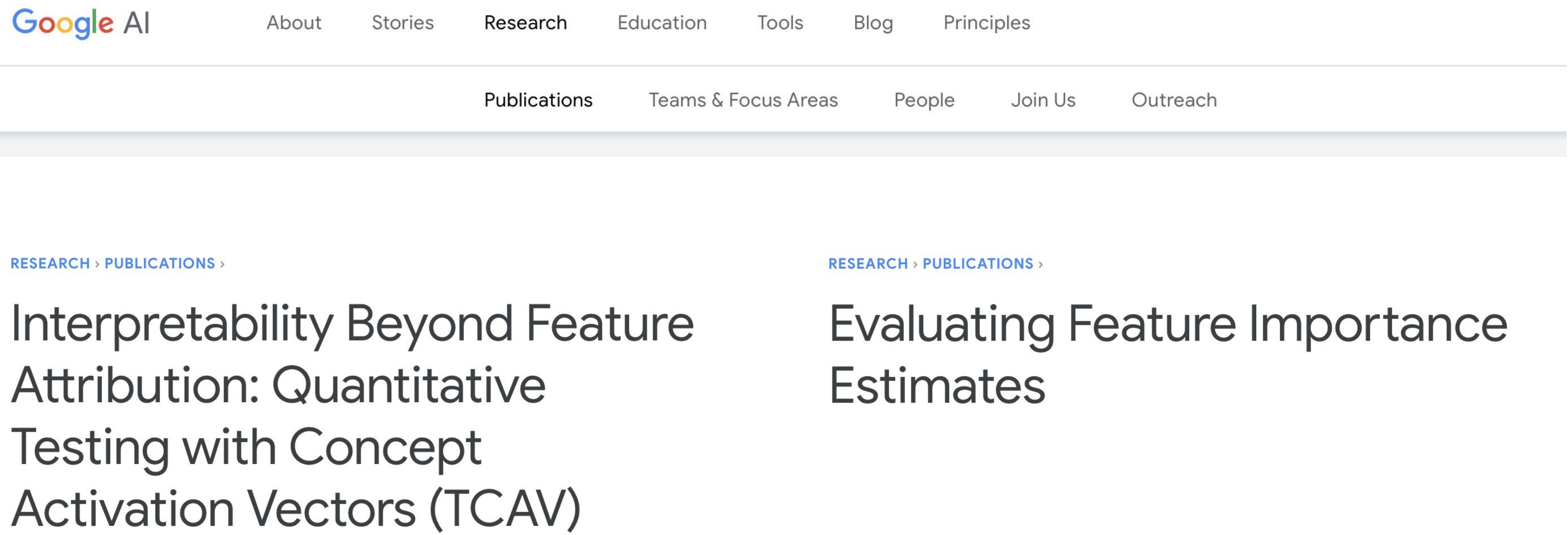
How to Explain Explanations? Towards User-Friendly Explanations for Predictive Systems

Andrew Smart, Divya Sharma, Andreas Kramm, Jessica Staddon

Abstract:

Predictive systems are often criticized as being opaque and hard for users to understand. This can sometimes lead to user discomfort, especially when it is unclear how a prediction has been generated. Explanations for predictions are increasingly recommended to improve the user experience of predictive systems. In their most granular, personalized form, explanations appear with each prediction and explain which user inputs most influenced the prediction. Often specific predictions made by a system are based on numerous variables and weights in a deep neural network, and thus the real reason for the prediction is not available. While there is strong evidence that explanations are a valuable tool for increasing system transparency and, hence, user comfort and trust, we argue there are substantial gaps in the design guidance available to those implementing explanations. An open question for user-facing explanations is explain *what* to *whom*, and *how*? We argue that blanket calls for transparency should be informed by a careful consideration of philosophical, psychological, computer science, Human Computer Interaction, legal and ethical considerations about what counts as a good explanation in different contexts. In particular, user perceptions of explanations have been shown to be very sensitive to language and presentation. In addition, there is evidence that

Interpretability



RESEARCH › PUBLICATIONS ›

Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)

RESEARCH › PUBLICATIONS ›

Evaluating Feature Importance Estimates

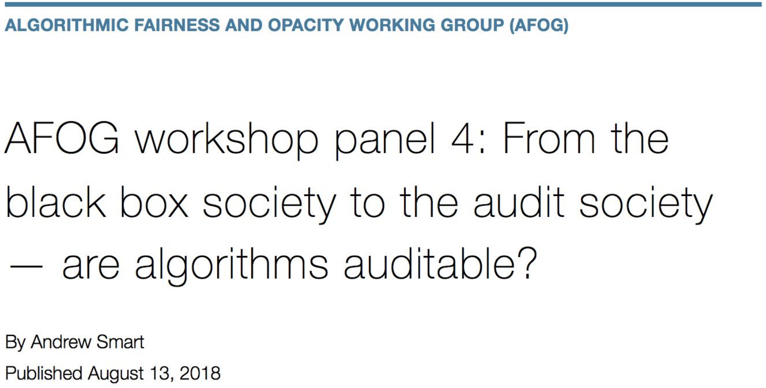
Auditability

- A systematic and independent examination of **product development processes, documents, and product performance** to determine whether the product's use case, performance, and development are in accordance our principles
- **Risk assessment** prior to launch

Algorithmic Fairness and Opacity Working Group

Our Partners

Center for Technology, Society & Policy
UNIVERSITY OF CALIFORNIA, BERKELEY



ALGORITHMIC FAIRNESS AND OPACITY WORKING GROUP (AFOG)

AFOG workshop panel 4: From the black box society to the audit society — are algorithms auditable?

By Andrew Smart
Published August 13, 2018

Transparency Options

Suite of options for transparency & accountability

- Explainability
- Interpretability
- Auditability
- Testing and validation
- International Standards
- External engagement
- Contestability
- User Feedback

Limitations and concerns of full algorithmic transparency

- Can be used by nefarious actors
- Does not lead to understanding
- Need to show not “what” is happening, but why
- Privacy and transparency tensions
- Algorithms learn from training data. Not reflected in source code
- Performance tradeoffs
- Competitive concerns

Key areas for clarification

1 Explainability standards

Fairness appraisal

Safety considerations

Human-AI collaboration

Liability frameworks

- Assemble a collection of best practice explanations along with commentary on their praiseworthy characteristics to provide practical inspiration.
- Provide guidelines for hypothetical use cases so industry can calibrate how to balance the benefits of using complex AI systems against the practical constraints that different standards of explainability impose.
- Articulate minimum acceptable standards in different industry sectors and application contexts.

Key areas for clarification

- Explainability standards
- 2 Fairness appraisal
- Safety considerations
- Human-AI collaboration
- Liability frameworks

- Articulate frameworks to balance competing goals and definitions of fairness.
- Clarify the relative prioritization of competing factors in some common hypothetical situations, even if this will likely differ across cultures and geographies.

Key areas for clarification

Explainability standards

Fairness appraisal

3 Safety considerations

Human-AI collaboration

Liability frameworks

- Outline basic workflows and standards of documentation for specific application contexts that are sufficient to show due diligence in carrying out safety checks.
- Establish safety certification marks to signify that a service has been assessed as passing specified tests for critical applications.

Key areas for clarification

- Explainability standards
- Fairness appraisal
- Safety considerations
- 4 Human-AI collaboration
- Liability frameworks

- Determine contexts when decision-making should not be fully automated by an AI system, but rather would require a meaningful "human in the loop".
- Assess different approaches to enabling human review and supervision of AI systems.

Key areas for clarification

Explainability standards

Fairness appraisal

Safety considerations

Human-AI collaboration

5 Liability frameworks

- Evaluate potential weaknesses in existing liability rules and explore complementary rules for specific high-risk applications.
- Consider sector-specific safe harbor frameworks and liability caps in domains where there is a worry that liability laws may otherwise discourage societally beneficial innovation.
- Explore insurance alternatives for settings in which traditional liability rules are inadequate or unworkable.

Resources

ai.google/education/

PAIR - ai.google/research/teams/brain/pair