Google Cloud

AI Principles and Responsible AI

Introduction to Responsible AI in Practice

In this module, you learn to ...

01 Identify the **need for Responsible AI**

02 Recognize that **decisions** made at **all stages of a project** have an impact on Responsible AI

03 Understand **Google's AI Principles**

04 Explore **Responsible AI practices**

Topics

01 AI & Responsibility

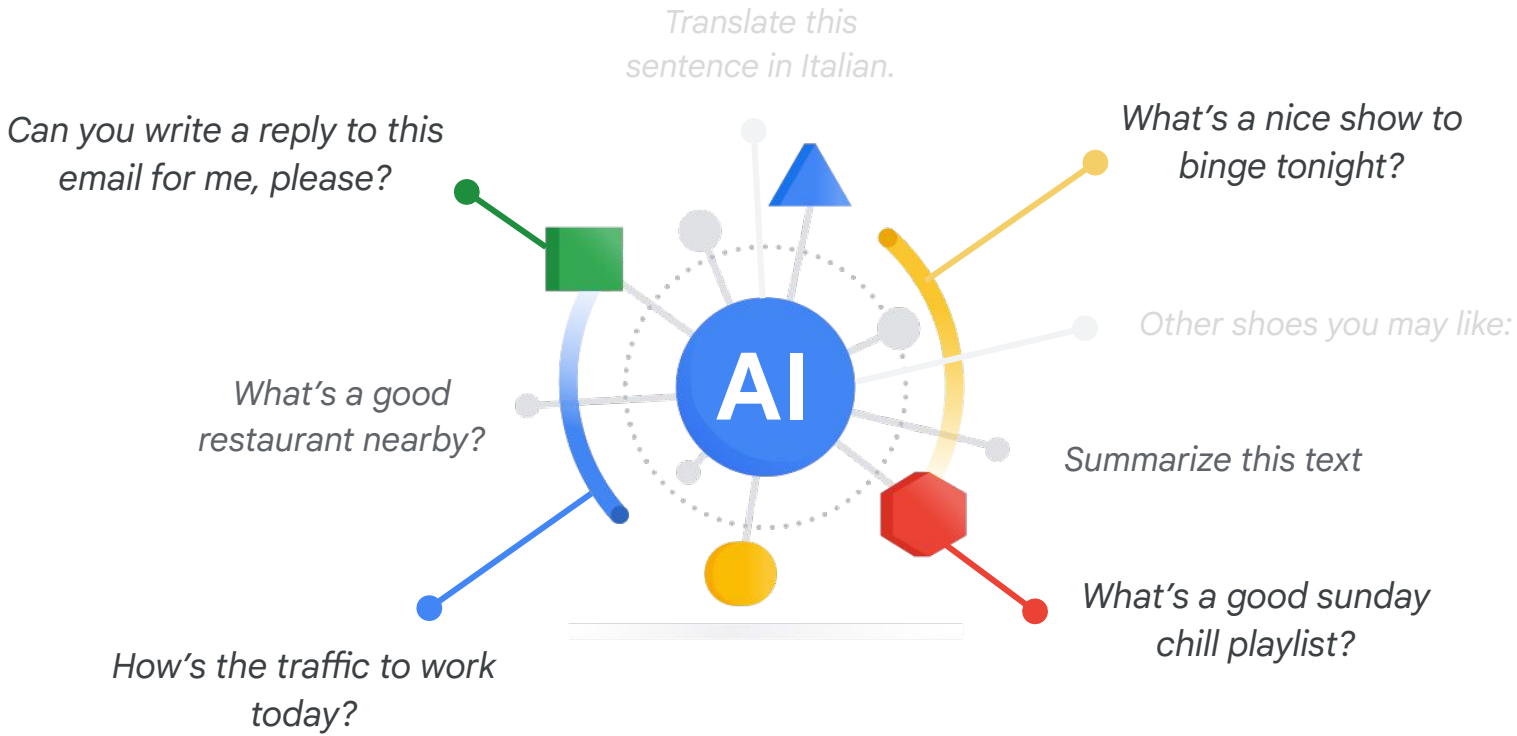
02 Google's AI Principles

03 Responsible AI Practices

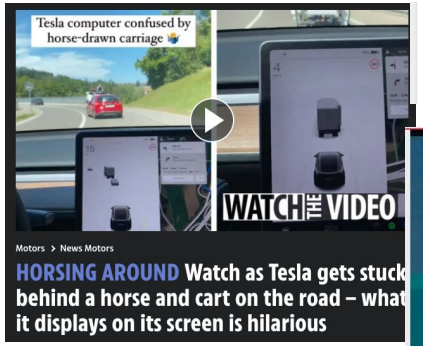
Topics

01	AI & Responsibility
02	Google’s AI Principles
03	Responsible AI Practices

AI is part of our daily lives



AI is not infallible



Zillow AI Goes Crazy. Causes $8 Billion Drop in Market Cap, a $304 Million Operating Loss, and 2,000+ Jobs

AI-Generated 'Seinfeld' Banned From Twitch After Making Transphobic Jokes

PRO CYBER NEWS

Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case

Scams using artificial intelligence are a new challenge for companies

SECURITY

Facial recognition tool led to mistaken arrest, lawyer says

Facial recognition systems have faced criticism because of their mass surveillance capabilities and because some studies have shown that the technology is far more likely to misidentify Black and other people of color than white people.

AI is built by people

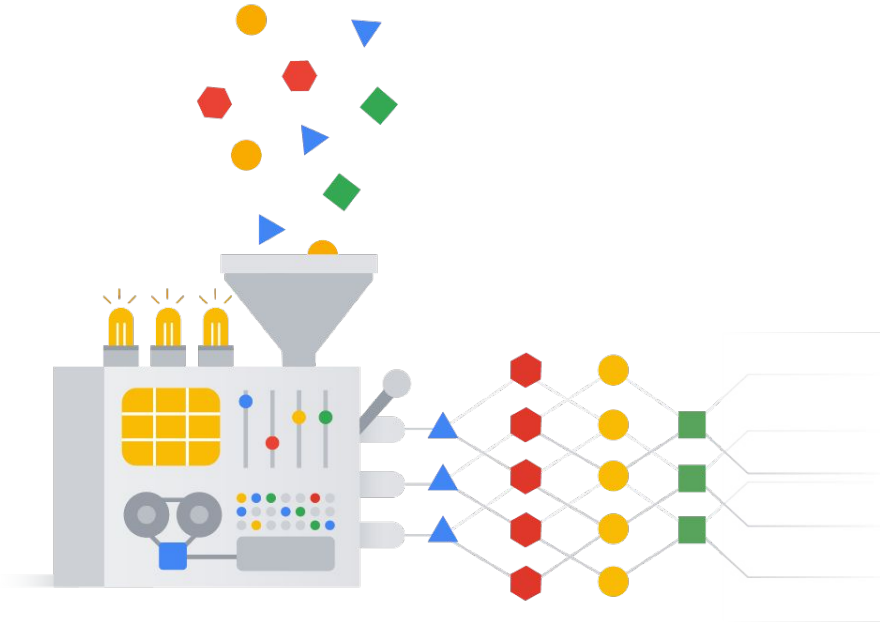
Collect data

Design

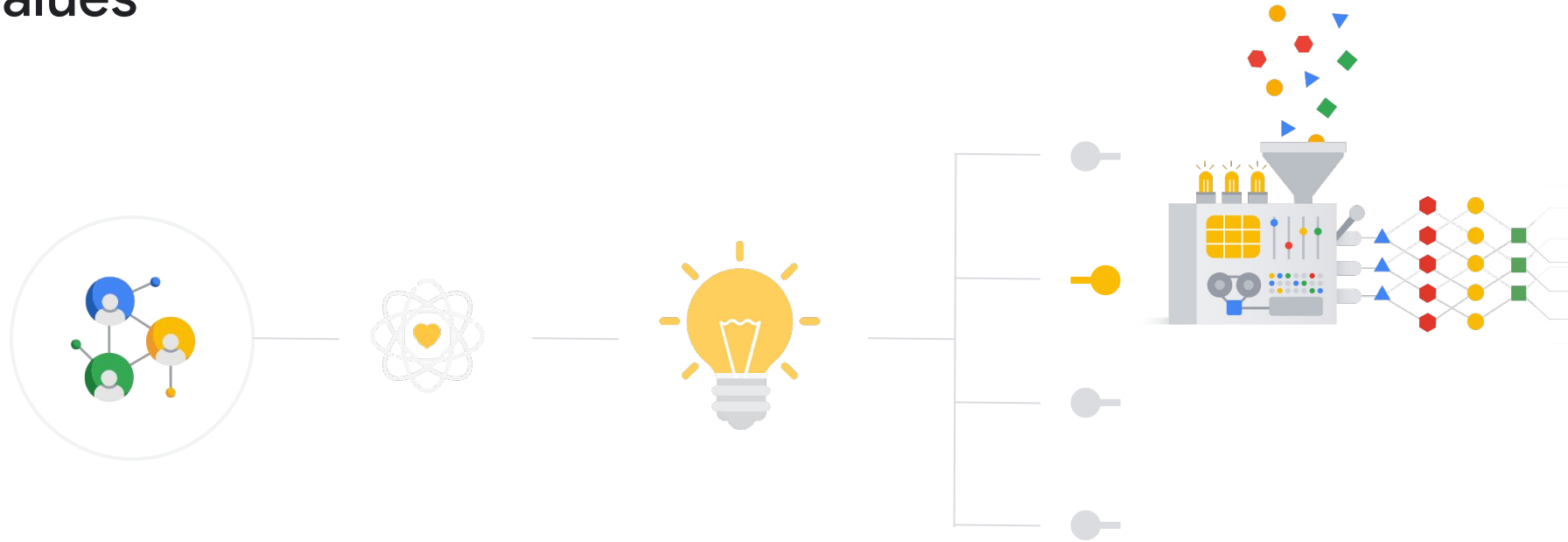
Build

Deploy

Apply



People make decisions on their own values



That's why you need Responsible AI

Every decision point requires consideration and evaluation to ensure that choices have been made responsibly

What is ethics for AI?

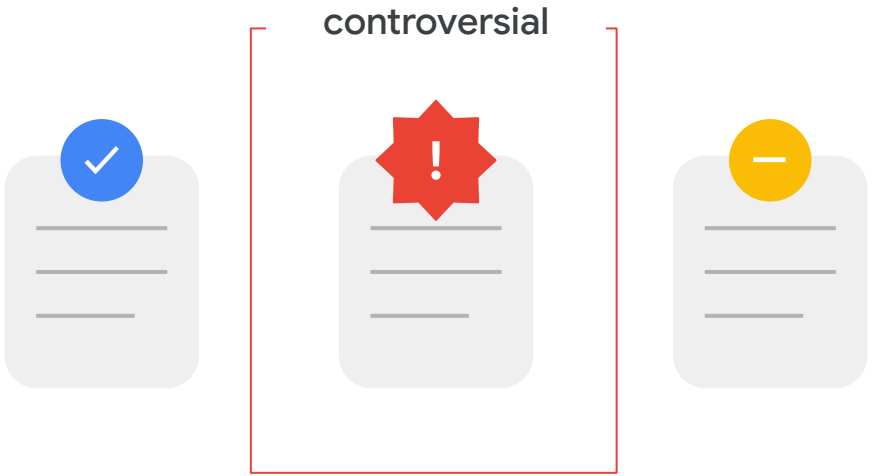
Ethics ≠ Law

Ethics ≠ Policy

What is ethics for AI?

- What we **ought** to do
- What others can **rightly** blame us for not doing
- What sustains our flourishing **together** in human society.

Focus on the outcomes, not intentions...



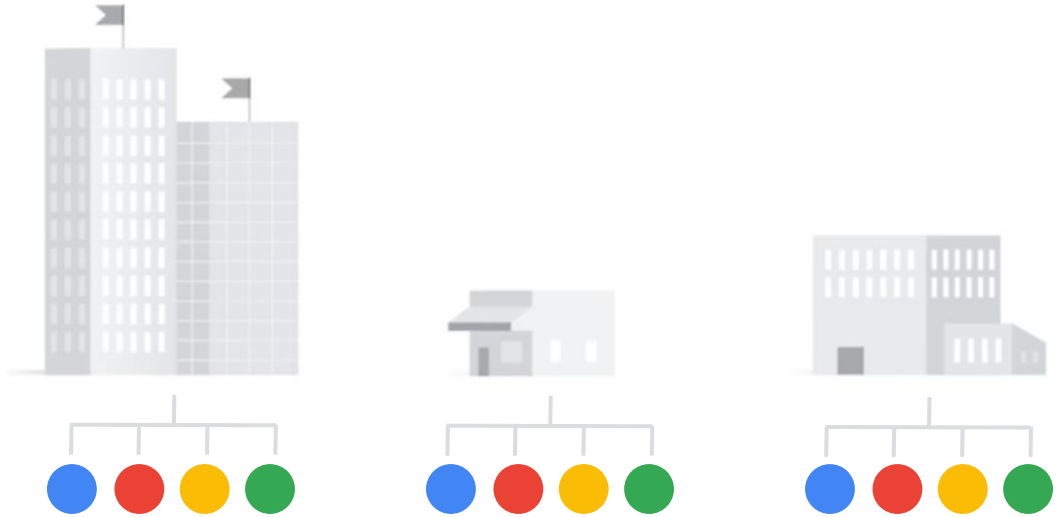
...to understand what Responsible AI is

Responsible AI is important because it's the **right thing to do** and it can guide AI design to be more **beneficial for people's lives.**

Responsible AI is....?

Responsible AI requires an understanding of the possible issues, limitations, or unintended consequences.

Organizations are developing their own AI principles



Organizations are developing their own AI principles

Values-based AI is good for your business

Safer and more accountable products

Advanced technologies are most successful when everyone can benefit from them.

Earn and keep your customers' trust

Irresponsible AI loses customers' trust, then customers. Responsible AI delights customers.

A culture of responsible innovation

Ethics forms the foundation as you explore new, innovative ways to drive your mission forward.

Topics

01	AI & Responsibility
02	Google's AI Principles
03	Responsible AI Practices

Google commits that its AI applications are:

- Built for everyone
- Accountable and safe
- Respect privacy
- Driven by scientific excellence

Google's AI Principles

7

objectives to follow

4

areas **not** to pursue

Google's AI Principles

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

Google's AI Principles

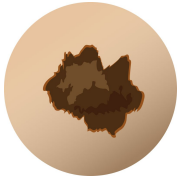
AI should:

Be socially beneficial

Google's AI Principles

1

Be socially beneficial



AI/ML models designed to **predict future development of melanomas** in patients

A recommendation engine to **suggest online skills training** for employees

A drone guidance system for **emergency aid airdrops** to disaster sites

Google's AI Principles

AI should:

Avoid creating or reinforcing unfair bias

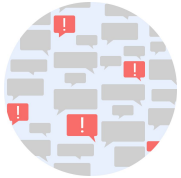
Google's AI Principles

2

Avoid creating or reinforcing unfair bias

Tech that makes or assists in **criminal justice decisions**

A hiring algorithm **ranks candidate application** relevance for recruiters



A machine learning-driven AI designed to **flag abusive, offensive, or hate speech**

Google's AI Principles

AI should:

Be built and tested for safety

Google's AI Principles

3

Be built and tested for safety

An ML Model that explores new **strategies and efficiencies in city power grid**

An AI agent that routes calls in an **emergency dispatch system**

A new ML model that **predicts jet engine failure**

Google's AI Principles

AI should:

Be accountable to people

Google's AI Principles

4

Be accountable to people

A recommendation system that makes fully automated decisions without consent, explanation, and right of appeal, such as **credit** and **insurance decisions**

An AI bot that **convincingly imitates a human agent**

A biometric ID system that is introduced **without a user's notice, consent, and ability to opt-out**

Google's AI Principles

AI should:

Incorporate privacy design principles

Google's AI Principles

5

Incorporate privacy design principles

A 'smart' refrigerator that **learns user purchasing habits**

A geolocation app that **predicts local foot traffic patterns**

A therapy app that **processes records of psychological issues**

Google's AI Principles

AI should:

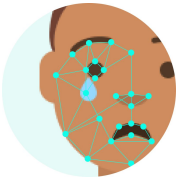
Uphold high standards of scientific excellence

Google's AI Principles

6

Uphold high standards of scientific excellence

An AI/ML app for **emotion detection**



An AI/ML app that **detects signs of clinical depression**

An AI/ML tool that **advances deepfake detection**

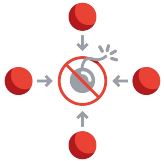
Google's AI Principles

AI should:

Be made available for uses that accord with these principles

Applications we will not pursue

likely to cause overall harm



technologies primarily intended to cause injury

surveillance violating internationally accepted norms

purpose contravenes international law and human rights

Google's AI Principles

We will **not** pursue:

Technologies likely to cause overall harm

Google's AI Principles

We will **not** pursue:

Weapons or technologies primarily intended to cause or facilitate injury

Google's AI Principles

We will **not** pursue:

Surveillance technology that violates internationally accepted norms

Google's AI Principles

We will **not** pursue:

Technologies whose purpose contravenes international law and human rights

Google's AI Principles

AI 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

We will not pursue:

1. Technologies likely to cause overall harm
2. Weapons or technologies primarily intended to cause or facilitate injury
3. Surveillance technology that violates internationally accepted norms
4. Technologies whose purpose contravenes international law and human rights

Topics

01 AI & Responsibility

02 Google's AI Principles

03 Responsible AI Practices

Responsible AI draws general best practices from software and quality engineering

6

ML-specific practices

Responsible AI Practices

When you develop AI, you should:

Use a human-centered design approach

Responsible AI Practices

1

Use a human-centered design approach

- Design features with appropriate disclosures built-in
- Consider augmentation and assistance
- Model potential adverse feedback early throughout
- Engage with a diverse set of users and use-case scenarios

Responsible AI Practices

1

Use a human-centered design approach

Design features with appropriate disclosures built-in

Consider augmentation and assistance

Model potential adverse feedback early throughout

Engage with a diverse set of users and use-case scenarios

Responsible AI Practices

Use a human-centered design approach

Design features with appropriate disclosures built-in

Consider augmentation and assistance

Model potential adverse feedback early throughout

Engage with a diverse set of users and use-case scenarios

Responsible AI Practices

Use a human-centered design approach

Design features with appropriate disclosures built-in

Consider augmentation and assistance

Model potential adverse feedback early throughout

Engage with a diverse set of users and use-case scenarios

Responsible AI Practices

1

Use a human-centered design approach

Design features with appropriate disclosures built-in

Consider augmentation and assistance

Model potential adverse feedback early throughout

Engage with a diverse set of users and use-case scenarios

Responsible AI Practices

When you develop AI, you should:

Identify multiple metrics to assess training and monitoring

Responsible AI Practices

Identify multiple metrics to assess training and monitoring

Define metrics from user feedback, system performance, short-term and long-term product health, and performance across data slices

Ensure that your metrics are appropriate for the context and goals of your system

Responsible AI Practices

Identify multiple metrics to assess training and monitoring

Define metrics from user feedback, system performance, short-term and long-term product health, and performance across data slices

Ensure that your metrics are appropriate for the context and goals of your system

Responsible AI Practices

Identify multiple metrics to assess training and monitoring

Define metrics from user feedback, system performance, short-term and long-term product health, and performance across data slices

Ensure that your metrics are appropriate for the context and goals of your system

Responsible AI Practices

When you develop AI, you should:

Directly examine your raw data

Responsible AI Practices

Directly examine your raw data

- Data should be accurate
- Data and data samples should be representative
- Training-serving skew shouldn't happen
- Data and model should be simple
- Features should be predictive of the label
- Data should have no / minimal bias

Responsible AI Practices

Directly examine your raw data

Data should be accurate

Data and data samples should be representative

Training-serving skew shouldn't happen

Data and model should be simple

Features should be predictive of the label

Data should have no / minimal bias

Responsible AI Practices

Directly examine your raw data

Data should be accurate

Data and data samples should be representative

Training-serving skew shouldn't happen

Data and model should be simple

Features should be predictive of the label

Data should have no / minimal bias

Responsible AI Practices

Directly examine your raw data

Data should be accurate

Data and data samples should be representative

Training-serving skew shouldn't happen

Data and model should be simple

Features should be predictive of the label

Data should have no / minimal bias

Responsible AI Practices

Directly examine your raw data

Data should be accurate

Data and data samples should be representative

Training-serving skew shouldn't happen

Data and model should be simple

Features should be predictive of the label

Data should have no / minimal bias

Responsible AI Practices

Directly examine your raw data

Data should be accurate

Data and data samples should be representative

Training-serving skew shouldn't happen

Data and model should be simple

Features should be predictive of the label

Data should have no / minimal bias

Responsible AI Practices

Directly examine your raw data

Data should be accurate

Data and data samples should be representative

Training-serving skew shouldn't happen

Data and model should be simple

Features should be predictive of the label

Data should have no / minimal bias

Responsible AI Practices

When you develop AI, you should:

Understand the limitations of your dataset and model

Responsible AI Practices

Understand the limitations of your dataset and model

Don't mistake correlation for causation

Communicate the scope and coverage of the training set

Communicate limitations to users where possible

Responsible AI Practices

Understand the limitations of your dataset and model

Don't mistake correlation for causation

Communicate the scope and coverage of the training set

Communicate limitations to users where possible

Responsible AI Practices

Understand the limitations of your dataset and model

Don't mistake correlation for causation

Communicate the scope and coverage of the training set

Communicate limitations to users where possible

Responsible AI Practices

Understand the limitations of your dataset and model

Don't mistake correlation for causation

Communicate the scope and coverage of the training set

Communicate limitations to users where possible

Responsible AI Practices

When you develop AI, you should:

Test, Test, Test

Responsible AI Practices

Test, Test, Test

- Conduct rigorous unit tests
- Conduct integration tests
- Detect input drift
- Use a gold standard dataset
- Conduct iterative user testing
- Apply the quality engineering principle of poka-yoke

Responsible AI Practices

Test, Test, Test

Conduct rigorous unit tests

Conduct integration tests

Detect input drift

Use a gold standard dataset

Conduct iterative user testing

Apply the quality engineering principle of poka-yoke

Responsible AI Practices

Test, Test, Test

Conduct rigorous unit tests

Conduct integration tests

Detect input drift

Use a gold standard dataset

Conduct iterative user testing

Apply the quality engineering principle of poka-yoke

Responsible AI Practices

Test, Test, Test

Conduct rigorous unit tests

Conduct integration tests

Detect input drift

Use a gold standard dataset

Conduct iterative user testing

Apply the quality engineering principle of poka-yoke

Responsible AI Practices

Test, Test, Test

Conduct rigorous unit tests

Conduct integration tests

Detect input drift

Use a gold standard dataset

Conduct iterative user testing

Apply the quality engineering principle of poka-yoke

Responsible AI Practices

Test, Test, Test

Conduct rigorous unit tests

Conduct integration tests

Detect input drift

Use a gold standard dataset

Conduct iterative user testing

Apply the quality engineering principle of poka-yoke

Responsible AI Practices

Test, Test, Test

Conduct rigorous unit tests

Conduct integration tests

Detect input drift

Use a gold standard dataset

Conduct iterative user testing

Apply the quality engineering principle of poka-yoke*

** A poka-yoke (Japanese, "mistake-proofing" or "error prevention") is any mechanism in a process that helps an equipment operator avoid (yokeru) mistakes (poka) and defects by preventing, correcting, or drawing attention to human errors as they occur.*

Responsible AI Practices

When you develop AI, you should:

Continue to monitor and update the system after deployment

Responsible AI Practices

Continue to monitor and update the system after deployment

Be ready for issues to occur

Consider both short- and long-term solutions to issues

Analyze the candidate model before deployment

Responsible AI Practices

Continue to monitor and update the system after deployment

Be ready for issues to occur

Consider both short- and long-term solutions to issues

Analyze the candidate model before deployment

Responsible AI Practices

Continue to monitor and update the system after deployment

Be ready for issues to occur

Consider both short- and long-term solutions to issues

Analyze the candidate model before deployment

Responsible AI Practices

Continue to monitor and update the system after deployment

Be ready for issues to occur

Consider both short- and long-term solutions to issues

Analyze the candidate model before deployment

Responsible AI Practices

When you develop AI, you should:

1. Use a human-centered design approach
2. Identify multiple metrics to assess training and monitoring
3. When possible, directly examine your raw data
4. Understand the limitations of your dataset and model
5. Test, Test, Test
6. Continue to monitor and update the system after deployment

Let’s recap...

In this module, you learned to ...

01 Responsible AI is the **right to do**, and it can guide AI to be more **beneficial for people's lives**

02 Google AI Principles strive towards AI that is **built for everyone, accountable and safe, respects privacy,** and is **driven by scientific excellence**

03 Responsible AI practices are applied **throughout the lifecycle**, and as early as possible

04 Responsible AI requires **governance processes**, a culture of **transparency** and **support**, and **continuous conversations**

Quiz question (1/4)

Which of the below is one of Google's 7 AI principles?

A: AI should create unfair bias.

B: AI should uphold high standards of operational excellence.

C: AI should uphold high standards of scientific excellence.

D: AI should gather or use information for surveillance.

Quiz question (1/4)

Which of the below is one of Google's 7 AI principles?

A: AI should create unfair bias.

B: AI should uphold high standards of operational excellence.

C: AI should uphold high standards of scientific excellence.

D: AI should gather or use information for surveillance.

Quiz question (2/4)

Which of these is correct with regard to applying responsible AI practices?

A: Decisions made at all stages in a project make an impact on responsible AI.

B: Decisions made at an early stage in a project do not make an impact on responsible AI.

C: Decisions made at a late stage in a project do not make an impact on responsible AI.

D: Only decisions made by the project owner at any stage in a project make an impact on responsible AI.

Quiz question (2/4)

Which of these is correct with regard to applying responsible AI practices?

A: Decisions made at all stages in a project make an impact on responsible AI.

B: Decisions made at an early stage in a project do not make an impact on responsible AI.

C: Decisions made at a late stage in a project do not make an impact on responsible AI.

D: Only decisions made by the project owner at any stage in a project make an impact on responsible AI.

Quiz question (3/4)

Organizations are developing their own AI principles that reflect their mission and values. What are the common themes among these principles?

A: A consistent set of ideas about transparency, fairness, and equity.

B: A consistent set of ideas about transparency, fairness, accountability, and privacy.

C: A consistent set of ideas about transparency, fairness, and diversity.

D: A consistent set of ideas about fairness, accountability, and inclusion.

Quiz question (3/4)

Organizations are developing their own AI principles that reflect their mission and values. What are the common themes among these principles?

A: A consistent set of ideas about transparency, fairness, and equity.

B: A consistent set of ideas about transparency, fairness, accountability, and privacy.

C: A consistent set of ideas about transparency, fairness, and diversity.

D: A consistent set of ideas about fairness, accountability, and inclusion.

Quiz question (4/4)

Why is responsible AI practice important to an organization?

A: Responsible AI practice can help drive revenue.

B: Responsible AI practice can help improve operational efficiency.

C: Responsible AI practice can help build trust with customers and stakeholders.

D: Responsible AI practice can improve communication efficiency.

Quiz question (4/4)

Why is responsible AI practice important to an organization?

A: Responsible AI practice can help drive revenue.

B: Responsible AI practice can help improve operational efficiency.

C: Responsible AI practice can help build trust with customers and stakeholders.

D: Responsible AI practice can improve communication efficiency.

Appendix

In this module, you learn to ...

01 Identify the **need for Responsible AI**

02 Recognize that **decisions** made at **all stages of a project** have an impact on Responsible AI

03 Understand **Google's AI Principles**

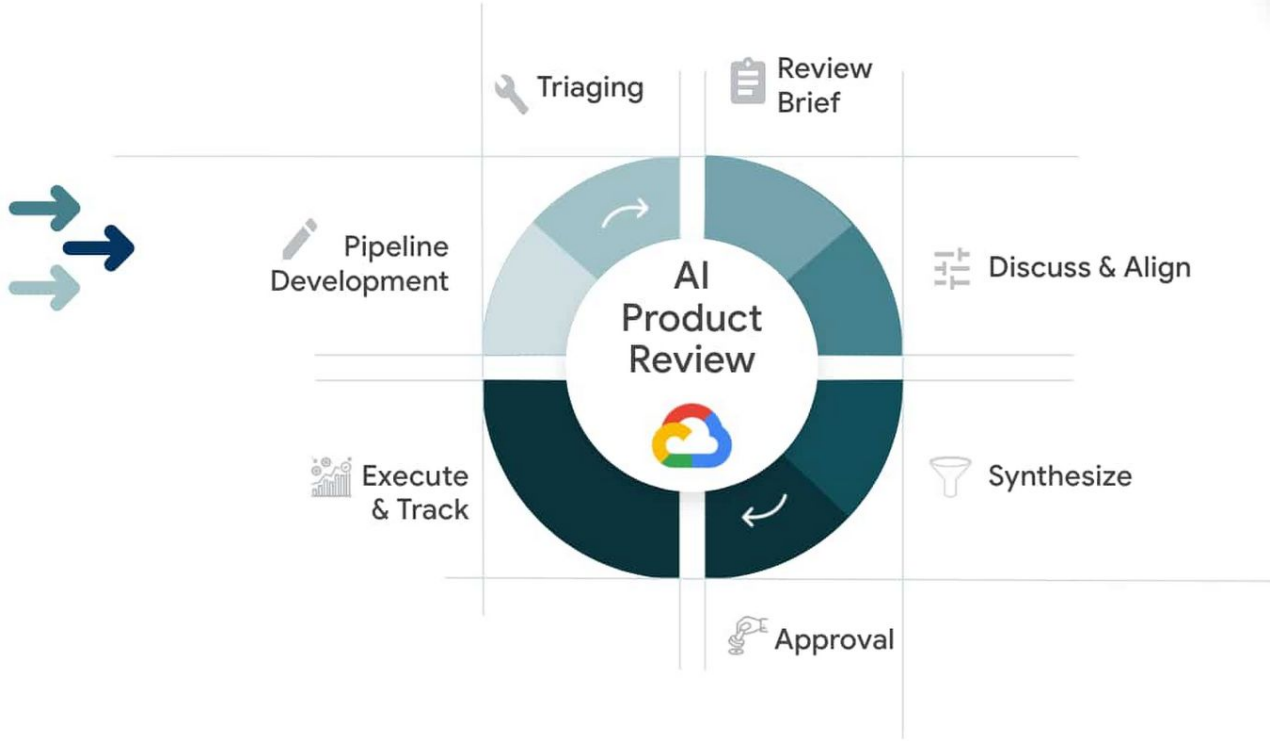
04 Explore **Responsible AI practices**

05 Discover **hands-on pro tips** for Responsible AI

Topics

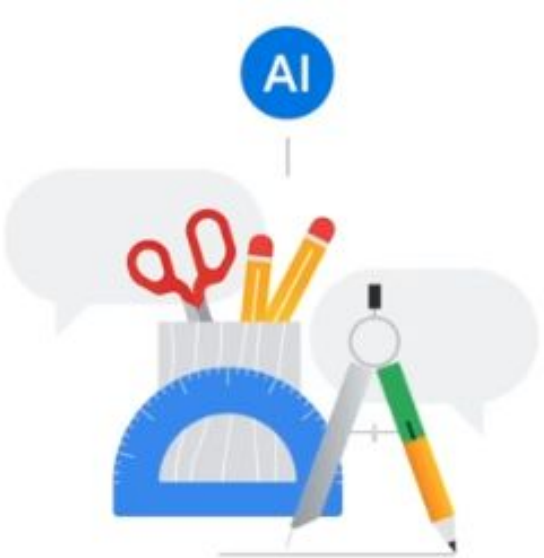
01	AI & Responsibility
02	Google's AI Principles
03	Responsible AI Practices
04	**Responsible AI Pro Tips**

Decisions around AI are made through a series of **assessments** and **reviews**



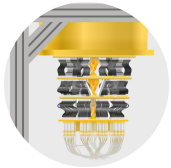
When **assessing** an application or a product, factors to consider are:

- Primary purpose and use
- Nature and Uniqueness
- Scale
- Nature of involvement
- ...and more!



When **assessing** an application or a product, factors to consider are:

- Primary purpose and use
- Nature and Uniqueness
- Scale
- Nature of involvement
- ...and more!



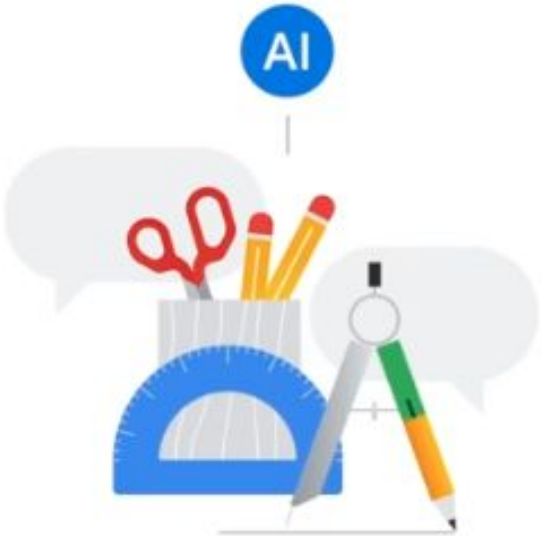
A quantum computing breakthrough that accelerates AI may require special evaluation because of its scale and its nature and uniqueness.

A custom ML model built for a **government customer** might require special evaluation because of the nature of Google's involvement and the technology's primary purpose and use.

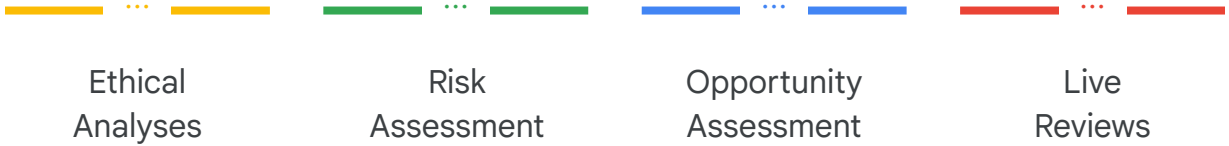
Federated learning that advances privacy

After **assessing** an application or a product, mitigation strategies may be:

- Narrow the scope of the product
- Create educational materials
- Define and release best practices
- Implement policy or terms of service
- Don’t move forward with the product
- ...and more!



Early in the development process, two diverse bodies provide their **review**



The review processes instill **rigor** and **consistency** in our approach across product areas and geographies.

Reviews succeed in an environment of free discussion and psychological safety

Not everyone will agree with every decision made on how products should be designed responsibly.

AI Principles rarely give us direct answers to our questions on how to build our products.

What should you keep in mind when developing Responsible AI practices?

Pro Tip #1:

No **ethics checklist.**

What should you keep in mind when developing Responsible AI practices?

Pro Tip #2:

Responsibility **by design.**

What should you keep in mind when developing Responsible AI practices?

Pro Tip #3:

Diversity of input.

What should you keep in mind when developing Responsible AI practices?

Pro Tip #4:

A culture of **support**.

What should you keep in mind when developing Responsible AI practices?

Pro Tip #5:

Transparency.

What should you keep in mind when developing Responsible AI practices?

Pro Tip #6:

A **humble** approach.

What should you keep in mind when developing Responsible AI practices?

Pro Tip #7:

The work is **not (easily) measurable.**

Responsible AI Pro Tips

When you apply Responsible AI, remember:

#1 No ethics checklist.

#2 Responsibility by design.

#3 Diversity of input.

#4 A culture of support.

#5 Transparency.

#6 A humble approach.

#7 The work is not (easily) measurable.