

**sas innovate**  
on tour 2025

# Designing for responsible AI

Make being responsible, irresistible

#SASInnovate

Copyright © SAS Institute Inc. All rights reserved.



Maribel Costa  
Lead trustworthy AI designer



Sierra Shell  
Trustworthy AI designer

# Our pursuit of making being responsible, irresistible...

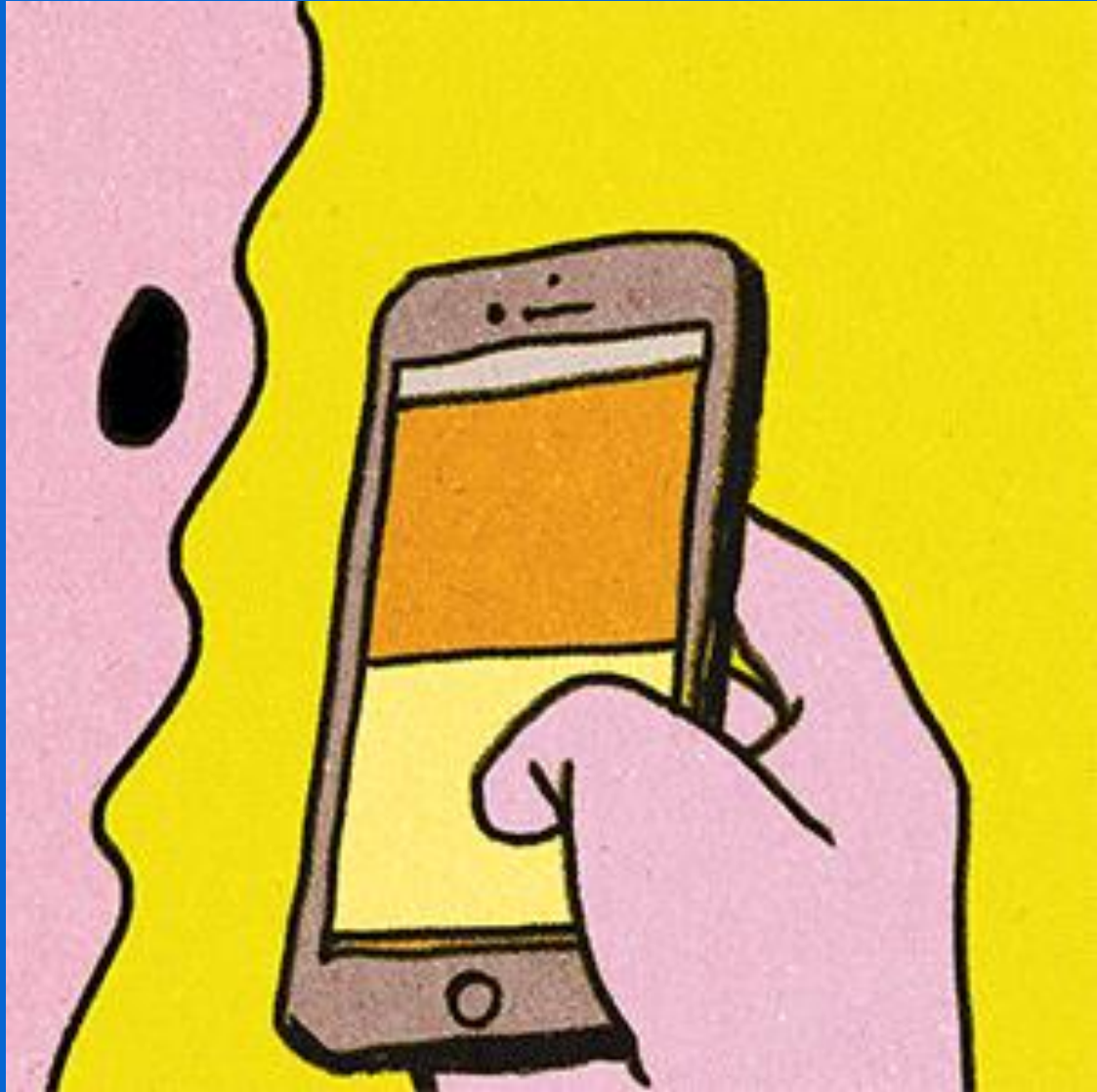
Design destroying the world with  
unintended consequences

3 Laws of New Technology

Reintroducing friction for responsible  
design

AI





Source: @specksofdust, Giphy



# Aza Raskin

Co-chairing member of the World Economic Forum's Global Council on the Future of AI

Co-founder of the Center for Humane Technology

Named FastCompany's Master of Design

Inventor of the infinite scroll



# 3 Laws of New Technology

From The Center for Humane Technology

[www.humanetech.com](http://www.humanetech.com)

# LAWS OF NEW TECHNOLOGY

1.

When you invent a new technology, you invent a new class of **responsibility**.

## 2.

If the technology confers **power**,  
you create a race.

3.

If we don't coordinate, the race will end  
in **tragedy**.

What role do designers play in taking responsibility for the “new” technology of AI?

How do we **coordinate?**

# How do we inspire **stewardship**?

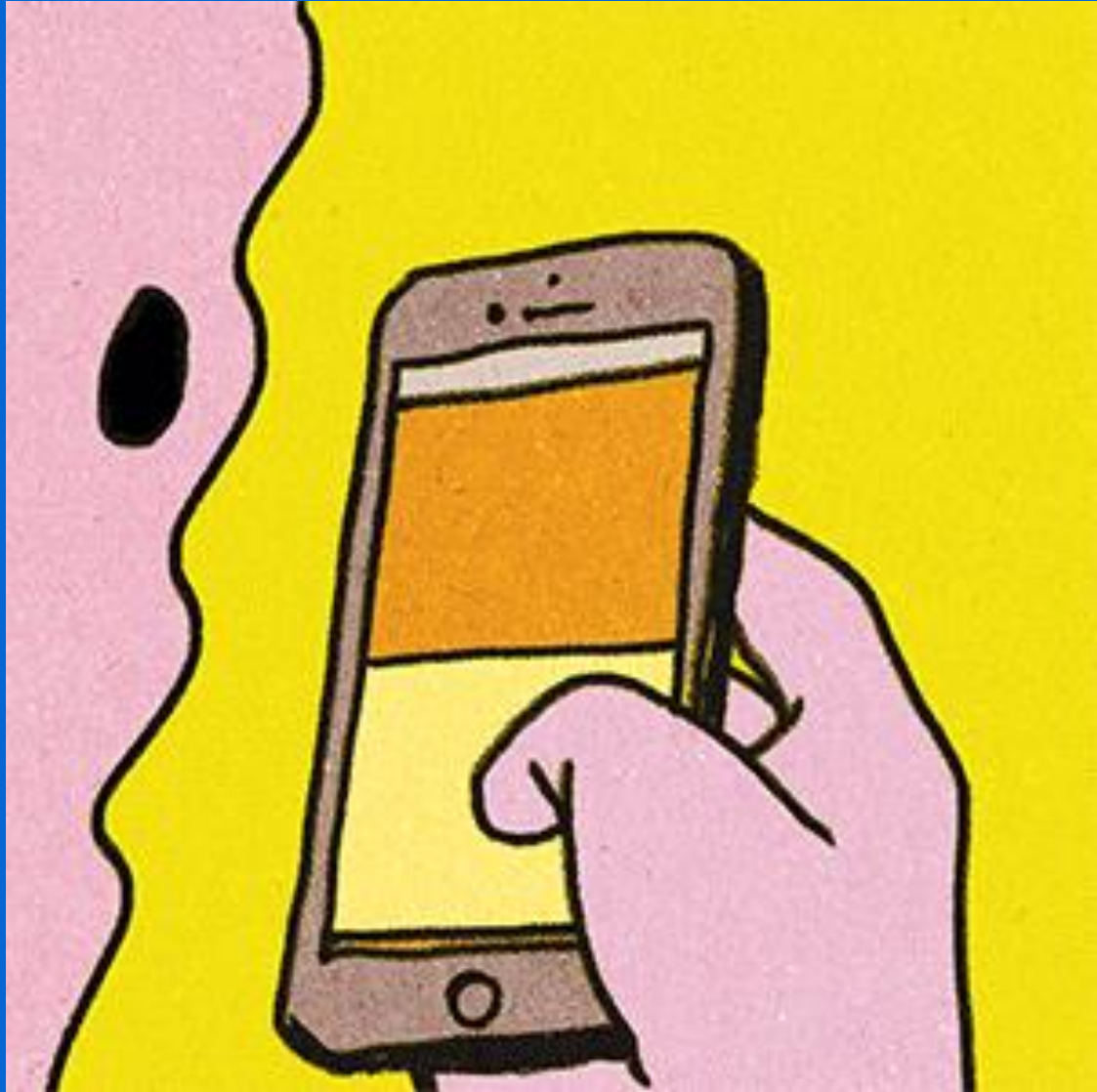
From Google AI Overview...

To act responsibly and be accountable for the long-term well-being and growth of an asset or creation on behalf of its beneficiaries.

# How do we inspire **stewardship**?

From Google AI Overview...

To act responsibly and be accountable for the long-term well-being and growth of an asset or creation on behalf of its beneficiaries.



Source: @specksofdust, Giphy



Source: @specksofdust, Giphy

# What is friction?

**Friction** is any added difficulty, effort, or delay encountered by a user.





## Cancel Subscription

Simply select one of the options below

**Contact support**

Current wait time: 29 days

Contact support

**Mail Us**

666 Avenue street,NY

Mail Us

# When does friction become friendly?

High-speed meets high-consequence

# Same function. Which is quicker?



# Same function. Which is quicker?



How do we utilize **friction** when designing AI features to ensure humans remain “in the loop”?

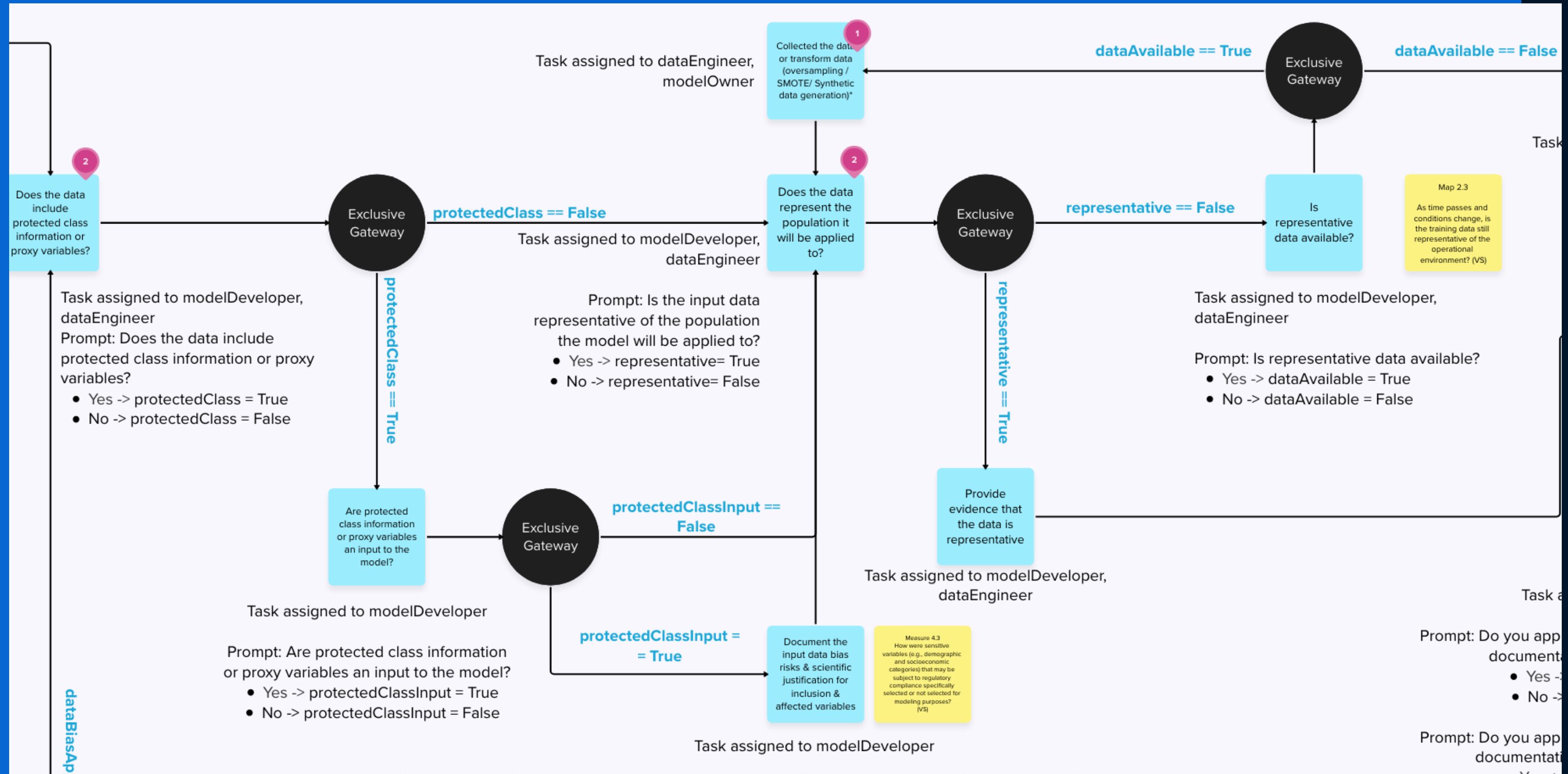
# USING INTENTIONAL FRICTION TO DESIGN RESPONSIBLE AI

1.

Guide the **process**, but not the decision.



# Trustworthy AI Lifecycle workflow



# USING INTENTIONAL FRICTION TO DESIGN RESPONSIBLE AI

## 2.

**Prime** the user for AI interaction.

# Google AI Overview

The screenshot shows a Google search interface with the query "What are strategies for responsible AI design?". The search results include an AI Overview section and several search results. The AI Overview section provides a summary of responsible AI design, highlighting key strategies such as using diverse, unbiased data, embedding fairness metrics, conducting regular audits, fostering diverse development teams, and creating clear accountability structures. Below the overview, there is a "Show more" button. The search results section includes a link to "Responsible AI by design: Building a framework of trust | The Enterprisers Project" and another link to "Building a responsible AI: How to manage the AI ethics debate - ISO".

**AI Overview**

Responsible AI design involves prioritizing fairness, ensuring transparency, implementing robust privacy and security measures, establishing strong governance, and promoting human-centeredness and safety. Strategies include using diverse, unbiased data, embedding fairness metrics into development, conducting regular audits, fostering diverse development teams, and creating clear accountability structures to mitigate risks and build trust.

Here are key strategies for responsible AI design:

1. Focus on Fairness & Bias Mitigation

Show more

**What is responsible AI?**  
From sources across the web

Transparency Privacy Accountability

# Where's the warning?

## 1. Focus on Fairness & Bias Mitigation

### Use representative data:

Ensure training data is diverse and reflects the population the AI will serve to avoid underrepresentation and bias.

### Embed fairness metrics:

Incorporate fairness assessments and metrics into the AI development process to monitor outcomes across different demographic groups.

### Assemble diverse teams:

Create multidisciplinary teams with varied backgrounds to bring different perspectives and identify potential biases.

## 2. Ensure Transparency & Explainability

### Promote explainability:

Design AI models whose decision-making processes are understandable and interpretable to humans.

### Provide clear documentation:

Maintain comprehensive documentation of AI systems to help stakeholders understand their functions and limitations.

## 3. Prioritize Privacy & Security

- **Protect data:** Implement strong privacy-preserving techniques and measures to safeguard individual data used in AI systems.
- **Secure systems:** Build AI systems that are secure from malicious attacks and breaches.

## 4. Establish Strong Governance & Accountability

### Define clear roles and responsibilities:

Establish governance frameworks with mechanisms to hold entities accountable for AI-driven outcomes.

### Conduct regular audits:

Perform regular audits of deployed AI systems to assess their real-world performance and identify issues.

### Develop ethical guidelines:

Create comprehensive mitigation strategies and ethical policies that guide the entire AI lifecycle, from development to deployment.

## 5. Promote Human-Centered Design & Safety

### Design for human well-being:

Focus on human-centered design to ensure AI systems complement and empower users, rather than replace them.

### Ensure safety and reliability:

Implement features and processes to prevent harmful system outputs, ensure system reliability, and mitigate unintended consequences.

## 6. Implement Continuous Monitoring & Improvement

### Collect feedback:

Actively collect user feedback to identify system blind spots and areas for improvement.

### Monitor for changes:

Continuously monitor AI systems to adapt to changing environments and new data, which includes retraining models as needed.

Dive deeper in AI Mode

### Responsible AI by design: Building a framework of trust | The Enterprisers Project

Dec 13, 2022

The Enterprisers Project

### Building a responsible AI: How to manage the AI ethics debate - ISO

Keeping up with AI best practice To keep your AI system trustworthy, it's important to focus on...

ISO

### The Decent Dozen: 12 Principles for Responsible AI by Design

Aug 20, 2024 — At their core, responsible AI considerations flow from fundamental human...

Infosys

Show all

# Waaaaaay down there...

**1. Focus on Fairness & Bias Mitigation**

**Use representative data:**  
Ensure training data is diverse and reflects the population the AI will serve to avoid underrepresentation and bias.

**Embed fairness metrics:**  
Incorporate fairness assessments and metrics into the AI development process to monitor outcomes across different demographic groups.

**Assemble diverse teams:**  
Create multidisciplinary teams with varied backgrounds to bring different perspectives and identify potential biases.

**2. Ensure Transparency & Explainability**

**Promote explainability:**  
Design AI models whose decision-making processes are understandable and interpretable to humans.

**Provide clear documentation:**  
Maintain comprehensive documentation of AI systems to help stakeholders understand their functions and limitations.

**3. Prioritize Privacy & Security**

- Protect data:** Implement strong privacy-preserving techniques and measures to safeguard individual data used in AI systems.
- Secure systems:** Build AI systems that are secure from malicious attacks and breaches.

**4. Establish Strong Governance & Accountability**

**Define clear roles and responsibilities:**  
Establish governance frameworks with mechanisms to hold entities accountable for AI-driven outcomes.

**Conduct regular audits:**  
Perform regular audits of deployed AI systems to assess their real-world performance and identify issues.

**Develop ethical guidelines:**  
Create comprehensive mitigation strategies and ethical policies that guide the entire AI lifecycle, from development to deployment.

**5. Promote Human-Centered Design & Safety**

**Design for human well-being:**  
Focus on human-centered design to ensure AI systems complement and empower users, rather than replace them.

**Ensure safety and reliability:**  
Implement features and processes to prevent harmful system outputs, ensure system reliability, and mitigate unintended consequences.

**6. Implement Continuous Monitoring & Improvement**

**Collect feedback:**  
Actively collect user feedback to identify system blind spots and areas for improvement.

**Monitor for changes:**  
Continuously monitor AI systems to adapt to changing environments and new data, which includes retraining models as needed.

**Responsible AI by design: Building a framework of trust | The Enterprisers Project**  
Dec 13, 2022  
The Enterprisers Project

**Building a responsible AI: How to manage the AI ethics debate - ISO**  
Keeping up with AI best practice To keep your AI system trustworthy, it's important to focus on...  
ISO

**The Decent Dozen: 12 Principles for Responsible AI by Design**  
Aug 20, 2024 — At their core, responsible AI considerations flow from fundamental human...  
Infosys

Show all

Dive deeper in AI Mode

AI responses may include mistakes. [Learn more](#)

📄 📌 📄 🗨

# How many people “learn more”?

## 1. Focus on Fairness & Bias Mitigation

### Use representative data:

Ensure training data is diverse and reflects the population the AI will serve to avoid underrepresentation and bias.

### Embed fairness metrics:

Incorporate fairness assessments and metrics into the AI development process to monitor outcomes across different demographic groups.

### Assemble diverse teams:

Create multidisciplinary teams with varied backgrounds to bring different perspectives and identify potential biases.

## 2. Ensure Transparency & Explainability

### Promote explainability:

Design AI models whose decision-making processes are understandable and interpretable to humans.

### Provide clear documentation:

Maintain comprehensive documentation of AI systems to help stakeholders understand their functions and limitations.

## 3. Prioritize Privacy & Security

- **Protect data:** Implement strong privacy-preserving techniques and measures to safeguard individual data used in AI systems.
- **Secure systems:** Build AI systems that are secure from malicious attacks and breaches.

## 4. Establish Strong Governance & Accountability

### Define clear roles and responsibilities:

Establish governance frameworks with mechanisms to hold entities accountable for AI-driven outcomes.

### Conduct regular audits:

Perform regular audits of deployed AI systems to assess their real-world performance and identify issues.

### Develop ethical guidelines:

Create comprehensive mitigation strategies and ethical policies that guide the entire AI lifecycle, from development to deployment.

## 5. Promote Human-Centered Design & Safety

### Design for human well-being:

Focus on human-centered design to ensure AI systems complement and empower users, rather than replace them.

### Ensure safety and reliability:

Implement features and processes to prevent harmful system outputs, ensure system reliability, and mitigate unintended consequences.

## 6. Implement Continuous Monitoring & Improvement

### Collect feedback:

Actively collect user feedback to identify system blind spots and areas for improvement.

### Monitor for changes:

Continuously monitor AI systems to adapt to changing environments and new data, which includes retraining models as needed.

Responsible AI by design: Building a framework of trust | The Enterprisers Project  
Dec 13, 2022

The Enterprisers Project

Building a responsible AI: How to manage the AI ethics debate - ISO

Keeping up with AI best practice To keep your AI system trustworthy, it's important to focus on...

ISO

The Decent Dozen: 12 Principles for Responsible AI by Design

Aug 20, 2024 — At their core, responsible AI considerations flow from fundamental human...

Infosys

Show all

## Monitor for changes:

Continuously monitor AI systems to adapt to changing environments and new data, which includes retraining models as needed.

Dive deeper in AI Mode

AI responses may include mistakes. [Learn more](#)



Dive deeper in AI Mode

AI responses may include mistakes. [Learn more](#)

# If you got nothing to hide, stand by your AI

The screenshot displays the SAS AI Governance Navigator interface. At the top, the navigation bar includes 'Home', 'Tasks', 'Policies', 'Connections', and 'Automated Resume Screening'. The main header reads 'Automated Resume Screening' with a sub-header 'Parses and evaluates resumes to support fair and efficient hiring.' Below this, a secondary navigation bar shows 'Overview', 'Policies', 'Development', 'Change Log', 'Assessments', 'Monitoring', 'Data Governance', and 'Findings 0'. On the left sidebar, a circular progress indicator shows '26% Assessment Completeness' with '317 fields remaining'. Below this, a list of categories includes 'Risk Management', 'Data Governance', 'Transparency', 'Human Oversight', 'Accuracy', 'Robustness', and 'Cybersecurity'. The main content area is titled 'Assessments' and 'Risk Management'. It shows 'Missing Risk Management fields' with a count of 8 and 'Risk Management fields requiring approval' with a count of 51. A 'Filter' section includes a search bar and checkboxes for 'Hide completed fields' and 'Hide non-required fields'. Under 'Policy', there are buttons for 'SAS GenAI Policy' and 'SAS AI Security Policy'. The assessment content is organized into numbered questions: \*1. What is the intended use of the LLM? (To assist HR by summarizing resumes and cover letters and generating "fit scores" for administrative job applicants. The goal is to speed up candidate screening while supporting fair hiring. Approved by Lena Caldwell on 04/25/2025 at 03:23:01 PM); \*2. Who are the end users of the LLM? (Human Resources staff and hiring managers who review AI-generated candidate summaries and rankings. Approved by Lena Caldwell on 04/25/2025 at 03:35:45 PM); \*3. What decisions or actions will be influenced by the LLM's outputs? (The LLM will influence which candidates are shortlisted for interviews and how they are initially perceived by reviewers. Completed by Lena Caldwell on 04/25/2025 at 03:38:22 PM); \*4. What are the potential consequences if the model behaves incorrectly or unpredictably? (Qualified candidates may be unfairly excluded, and biased outputs could lead to discriminatory hiring. This poses reputational, legal, and ethical risks for the organization. Completed by Lena Caldwell on 04/25/2025 at 03:50:12 PM); and \*5. Do you have safeguards in place to mitigate harms or offensive outputs? (Yes—AI outputs are reviewed by humans, bias audits are conducted regularly, and candidates are informed about AI involvement with an option to appeal. The system also highlights which resume elements influenced each score. Edited by Lena Caldwell on 04/25/2025 at 04:02:28 PM; pg. 4). The fifth question's response is highlighted in a yellow box with 'Accuracy to source: 56% (edited)' and 'Reject' and 'Approve' buttons.

This block provides a close-up of the AI-generated response for question 5. The question is: '\*5. Do you have safeguards in place to mitigate harms or offensive outputs?' The response is: 'Yes—AI outputs are reviewed by humans, bias audits are conducted regularly, and candidates are informed about AI involvement with an option to appeal. The system also highlights which resume elements influenced each score.' The response is marked as 'Generated by AI' and 'Edited by Lena Caldwell on 04/25/2025 at 04:02:28 PM; pg. 4'. Below the response, a yellow box displays 'Accuracy to source: 56% (edited)' and two buttons: 'Reject' and 'Approve'.

# USING INTENTIONAL FRICTION TO DESIGN RESPONSIBLE AI

## 3.

Identify **pause points** and build in features accordingly.

# Areas of high consequence

Policy Source: ⓘ

Internal  
 External

Governing Body: ⓘ

Policy Provider: ⓘ

### Risk Management

Policy Classification Options:

After entering policy classification(s), re-order them from **most severe** to **least severe**. Select the Flag icon to indicate the highest level that is permitted.

- Sample policy classification 1 for wireframe purposes
- Sample policy classification 2 for wireframe purposes
- Sample policy classification 3 for wireframe purposes

Add policy classification option

Not applicable

Policy classification order saved.

Penalty for Non-Compliance:

Select the policy classification with the highest level of priority to continue.

Previous Next Save and close Cancel

## ⚠ Your bill may increase

You're adding another 5 admins to your selected product, if they don't already have access to this product your bill may increase.

Cancel

Continue



## Delete (Testing) Prototype

Warning: this cannot be undone.

- All share links will be inaccessible
- All prototype comments and history will be destroyed
- All screens and hotspots will be deleted

YES, DELETE PROTOTYPE

CANCEL, KEEP PROTOTYPE

# Areas of high consequence

Governance Navigator

Organization Assessments

Provide documentation about your organization by completing responses to assessment questions.

Last updated: Tuesday, June 4, 2025

Opened items (6)

Total completion

Percentage complete	Missing responses	Incomplete responses	Total questions
25%	76	12	116

Filtered missing responses: 62

Filtered incomplete responses: 10

Filtered questions: 100

Filter

Search

Show all

- Completed responses
- Missing responses
- Incomplete responses
- Required questions
- Unrequired questions

Policy:

- SAS AI Security Policy
- SAS GenAI Policy
- Trustworthy AI Life Cycle Workflow
- XYZ AI Law

SAS GenAI Policy (6)

- \*1. Does your quality management system include systems and procedures for record-keeping of all relevant documentation?  
Yes, our organization follows NIST best practices, including
- \*12. Does the organization have risk management frameworks in place?
- \*31. Did you implement sensible human oversight measures?
- \*33. Does your quality management system include an accountability framework setting out responsibilities?
- \*36. Is there any exception which makes your usage exempt from the policy?
- \*50. Have you taken measures to ensure, to your best extent, a sufficient level of AI literacy of your staff and other persons dealing with the operation and use of AI systems on your behalf?

Trustworthy AI Life Cycle Workflow (31)

- \*1. Does the organization have risk management frameworks in place?

SAS GenAI Policy (6)

- \*1. Does your quality management system include systems and procedures for record-keeping of all relevant documentation?  
Yes, our organization follows NIST best practices, including
- \*12. Does the organization have risk management frameworks in place?
- \*31. Did you implement sensible human oversight measures?

Hover

**Friendly friction.**



## SAS Product Design wants to...

- Guide the process, but not the decision.
- Prime the user for AI interaction.
- Identify pause points and build in features accordingly.



## Because...

- When you invent a new technology, you invent a new class of responsibility
- If the technology confers power, you create a race.
- If you do not coordinate, that race will end in tragedy.

“Make being responsible,  
irresistible.”  
—Reggie Townsend, VP Data Ethics Practice at SAS

Thanks Stuart McGee, wherever you are.

Takk, ha en god dag.

“3 Laws of New Technology” by the Center for Humane Technology:  
[www.humanetech.com/podcast/the-three-rules-of-humane-tech](http://www.humanetech.com/podcast/the-three-rules-of-humane-tech)

**sas innovate**  
on tour 2025

#SASInnovate

Copyright © SAS Institute Inc. All rights reserved.