

Trusted AI Assessment

Lucas Institute for Venture Equity & Ethics



Markkula Center
for Applied Ethics
at Santa Clara University

Table of Contents

AI Use

1

- 1.1 Stage
- 1.2 Application
- 1.3 Geography
- 1.4 Industry

Governance

2

- 2.1 Laws & Regulations
- 2.2 RAI Expertise
- 2.3 Stakeholder Engagement
- 2.4 Culture and Values

Management

3

- 3.1 Data Rights
- 3.2 Data Privacy
- 3.3 Validity & Reliability
- 3.4 Security & Resilience
- 3.5 Safety
- 3.6 Fairness & Bias Mitigation
- 3.7 Explainability & Transparency
- 3.8 Environmental Impact

Monitoring

4

- 4.1 Monitoring and Reporting

Resources

5

- 5.1 Resources

1.0 AI Use Case

The AI Use section of the assessment tool is tailored to evaluate how companies utilize AI technologies at different stages of funding and development. The assessment evaluates the alignment of AI applications with the specific needs and risks at each developmental phase, ensuring that companies scale their AI operations responsibly. Additionally, the tool assesses whether AI products and services fall into high-risk categories under regulatory frameworks such as the EU AI Act and the US Executive Order on AI.

Company Summary

Company Name	
Stage of Funding	Seed
Primary Use of AI	AI-Enabled - Embedded Algorithm at the core.
Geographic Region	USA
Primary Industry Focus	Healthcare
Primary Business Functions	Customer experience enhancement Operational efficiency and automation Data analysis and insights Product or service innovation

High Risk Application Considerations Based on EU AI Act or US Executive Order

- The company has not indicated uses that are considered high-risk based on the EU AI Act or US Executive Order.

2.0 Governance

The governance section of the AI assessment tool is designed to evaluate whether companies have governance structures in place to manage and mitigate ethical risks associated with AI technologies.

2.1 Laws and Regulation

Question

Company Response

Best Practice

Has the company reviewed laws and regulations applicable to AI use and application?

Yes

Yes

Consider implementing the following for a company that has not yet identified applicable statutes and regulations.

Recommendations

Pre-Seed to Seed

Series A

Series B

Actions

- Develop an understanding of AI regulations for your industry, and start documenting.

- Conduct an analysis of specific regulations impacting your industry and geography.
- Consult with legal experts as needed.
- Create product development guidance to ensure adherence to AI laws and regulations.

- Develop an internal compliance team or hire advisors.
- Implement compliance monitoring tools.

KPIs

- Quarterly analysis of potential laws and regulations.
- Quarterly report on applicable laws and regulations with compliance roadmap.

- List of applicable laws and regulations along with a roadmap for compliance with ownership and accountability defined.
- Legal consultation completed.
- Product development guidelines to ensure adherence to AI laws and regulations.
- Monitoring and auditing processes in place to ensure adherence to AI laws and regulations.

- A compliance team is established or a legal advisor is contracted.
- Auditing and monitoring systems in place.

Resources

[AI Ethics News](#)
[Regulation of AI in the US](#)
[Stanford Human-Centered Artificial Intelligence: By the Numbers: Tracking The AI Executive Order](#)
[KPMG Decoding the EU AI Act, A Guide for Business Leaders](#)

[Stanford HELM: A holistic framework for evaluating foundation models](#)
[Assessment List for Trustworthy AI: Self Assessment](#)

[Credo AI Governance, Risk, and Compliance Platform](#)

2.2 Responsible AI Expertise

Question	Company Response	Best Practice
Does the company have advisors or team members with RAI expertise?	Yes	Yes

For a company that answered "No" to having advisors or team members with Responsible AI (RAI) expertise, here are some activities and performance indicators to consider by stage of company funding:

Recommendations	Pre-Seed to Seed	Series A	Series B
Actions	<ul style="list-style-type: none"> Build awareness of RAI networks through RAI-focused events and communities. Start networking to identify potential RAI advisors. 	<ul style="list-style-type: none"> Engage with RAI consultants for initial assessments and training plan. Develop a plan to integrate RAI expertise into the team. 	<ul style="list-style-type: none"> Establish a dedicated RAI advisory board. Implement RAI training program for staff and executives.
KPIs	<ul style="list-style-type: none"> Build a list of RAI experts and advisors. 	<ul style="list-style-type: none"> RAI expertise strategy and plan developed. Number of team members trained in RAI basics. 	<ul style="list-style-type: none"> RAI advisory board established with regular meeting intervals. Percentage of staff completing RAI training.
Resources	<ul style="list-style-type: none"> The AI Alliance AI Northeastern Advisory Board Services AI Ethics Lab 		

2.3 Stakeholder Engagement

Question	Company Response	Best Practice
<p>Does the company engage with stakeholders to solicit feedback during designing, developing, or deploying AI products and services?</p>	<p>Yes</p>	<p>Yes</p>

For a company that answered "No" to stakeholder engagement, here are some activities and performance indicators to consider by stage of company funding:

Recommendations	Pre-Seed to Seed	Series A	Series B
<p>Actions</p>	<ul style="list-style-type: none"> Identify relevant internal and external stakeholders. Conduct feedback sessions with a small group of potential users or experts in your field to understand how the product or service impacts them. 	<ul style="list-style-type: none"> Develop feedback mechanisms with stakeholders, such as surveys or focus groups. Create a system for sharing feedback for product development and iterations. Document changes in products or services based on stakeholder feedback and learnings. 	<ul style="list-style-type: none"> Establish regular stakeholder advisory panels and integrate feedback loops into the product lifecycle for continuous improvement.
<p>KPIs</p>	<ul style="list-style-type: none"> Number of stakeholder feedback sessions or surveys conducted. Qualitative summary of feedback received. 	<ul style="list-style-type: none"> Percentage of product changes influenced by stakeholder feedback. 	<ul style="list-style-type: none"> Quantifiable impact of stakeholder feedback on product enhancements and customer satisfaction metrics.

2.4 Culture and Values

Question

Company Response

Best Practice

Does the company have a clearly defined set of values that guide its decision-making processes?

Yes

Yes

For a company that answered "No" to culture and values, here are some activities and performance indicators to consider by stage of company funding:

Recommendations

Pre-Seed to Seed

Series A

Series B

Actions

- Start internal discussions to articulate core values.
- Engage in team-building activities to align on a shared vision.
- Document values and vision.

- Formalize the company's core values through workshops and integrate them into company policies and employee onboarding.
- Develop values-based guidelines for RAI development using a tool such as Ethics Canvas.

- Embed the company's values in all aspects of the business, from hiring practices to product development.
- Evaluate decisions and strategies against these values.
- Share stories or case studies demonstrating values in action.
- Train staff on applying values in their day-to-day operations.
- Implement a formal review process to ensure AI practices adhere to values.
- Set up a reporting system to track value alignment in AI operations.
- Include the resources for scaling and maintaining an AI ethics program in the budget.

KPIs

- A document outlining company values.
- A values statement that guides the company's decision-making.

- Company values are published on the website and included in the employee handbook.
- Ethical AI framework that guides teams.

- Regular reviews of business practices for alignment with company values.
- Feedback mechanisms to assess adherence to values across the organization such as anonymous surveys and focus groups.
- Case studies of organizational value-driven decisions.
- Budget to support ethical AI training and processes.

Resources

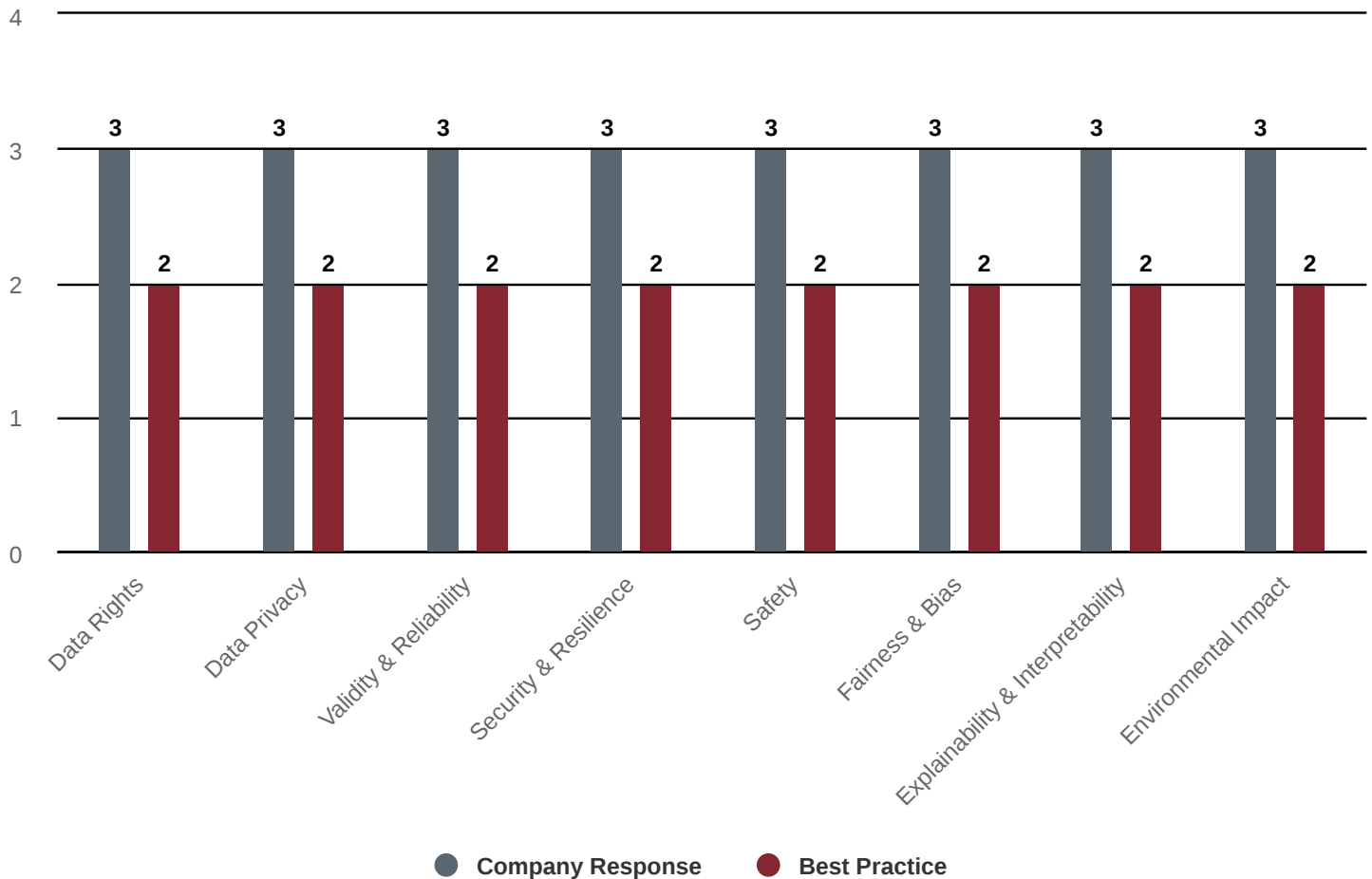
[ITEC Handbook from the Markkula Center for Applied Ethics at Santa Clara University](#)

[Ethics Canvas](#)
[Open Ethics](#)
[O'Reilly Data Ethics Checklist](#)

[An Ethical Checklist for Data Scientists](#)
[A Lean Approach to AI Governance](#)

Company Compared to Best Practices

This section of the assessment serves as a benchmark for companies and investors by detailing how a company's current AI activities align with recommended practices, highlighting areas of strength and opportunities for improvement. In the absence of formalized responsible AI standards, this comparison helps companies to identify and prioritize the next steps in their AI strategy. It facilitates informed decision-making, helping companies to develop actionable plans that enhance compliance, ethical alignment, and operational effectiveness in their company.



- **Level 1:** The company recognizes the importance of responsible AI (RAI) but has not yet implemented processes or practices.
- **Level 2:** The company is in the initial stages of determining RAI risks. Informal processes and guidelines have been initiated.
- **Level 3:** The company has implemented guidelines, processes, and systems for managing RAI. The team shares ownership and accountability, and ad hoc monitoring is in place.
- **Level 4:** The company has implemented formal policies, processes, and systems for managing RAI. There is oversight at the executive level. Accountability is defined. RAI is continuously monitored, managed, and updated. Training has been implemented.

3.0 RAI Management

This section evaluates critical aspects of ethical risk and mitigation that impact both operational integrity and stakeholder trust. For each of the areas, the assessment provides recommended actions tailored to enhance the company's AI practices, relevant performance indicators to track progress, and resources to help implement improvements. This approach helps companies and investors understand their current position relative to best practices, enabling them to make informed decisions on how to address gaps and build a more responsible and effective AI strategy.

3.1 Data Rights

Assessing how a company defines, enforces, and respects individuals and entities' data ownership and usage rights.

Question

Company Response

Best Practice

Please indicate the level of the company's RAI practices for managing Data Rights

Level 3
Intermediate RAI Practices

Level 2
Basic RAI Practices

Recommendations

Level 2

Level 3

Level 4

Actions

- | | | |
|---|---|--|
| <ul style="list-style-type: none"> • Document data sources and flows. • Identify data rights risks. • Develop a set of guidelines for responsible stewardship of data. | <ul style="list-style-type: none"> • Implement protocols for verifying data provenance and securing sourcing rights. • Implement data handling and consent processes. • Enhance employee training to include specific modules on data rights in AI applications. | <ul style="list-style-type: none"> • Regularly audit data sourcing and provenance practices to ensure compliance with evolving AI ethics and laws. • Develop partnerships or certifications that reinforce responsible data sourcing standards. • Monitor and report instances of data misuse and user feedback. • Provide data rights training for all relevant staff. • Policy reviews and updates. |
|---|---|--|

KPIs

- | | | |
|--|---|--|
| <ul style="list-style-type: none"> • Overview of data rights risks and mitigation plan and actions. • Data flow diagrams and source documentation. | <ul style="list-style-type: none"> • Number of data sources with verified provenance. • The frequency of training sessions. | <ul style="list-style-type: none"> • Records of policy reviews and updates. • Compliance reports. • Training completed. • Data misuse and user feedback reports. |
|--|---|--|

3.2 Data Privacy

The measures are in place to protect personal and sensitive information from unauthorized access or disclosure.

Question	Company Response		Best Practice
<p>Please indicate the level of the company's RAI practices for managing Data Privacy</p>	<p>Level 3 Intermediate RAI Practices</p>		<p>Level 2 Basic RAI Practices</p>
Recommendations	Level 2	Level 3	Level 4
Actions	<ul style="list-style-type: none"> • Post a customer/user privacy policy. • Begin implementing user consent mechanisms for data collection and use. • Start adopting data minimization practices. • Conduct a data privacy risk assessment. 	<ul style="list-style-type: none"> • Implement encryption technologies for data at rest and in transit. • Explore differential privacy techniques • Start incorporating federated learning for model training. 	<ul style="list-style-type: none"> • Integrate advanced privacy-enhancing technologies like homomorphic encryption and secure multi-party computation. • Establish comprehensive data breach and incident reporting protocols.
KPIs	<ul style="list-style-type: none"> • Data privacy risk assessment findings. • Privacy policy published. • Percentage of data privacy processes compliant with user consent protocols. 	<ul style="list-style-type: none"> • Confirmation of GDPR, CCPA, and CPRA compliance where applicable. • List of data privacy encryption technologies implemented. • Initial projects utilizing differential privacy or federated learning. 	<ul style="list-style-type: none"> • Number of systems using homomorphic encryption or secure multi-party computation. • Documented incident response plan. • Data breach reporting rate. • Data privacy policy views.

3.3 Validity and Reliability

The policies, practices, and activities are in place to manage the accuracy, consistency, and trustworthiness of AI-generated outcomes and predictions.

Question	Company Response		Best Practice
<p>Please indicate the level of the company's RAI practices for managing Validity and Reliability</p>	<p>Level 3 Intermediate RAI Practices</p>		<p>Level 2 Basic RAI Practices</p>
Recommendations	Level 2	Level 3	Level 4
Actions	<ul style="list-style-type: none"> Establish baseline reliability performance metrics and tracking. Initiate validation tests to compare against baselines. Establish a process for collecting feedback on system performance from users. 	<ul style="list-style-type: none"> Implement a testing framework for new and updated models. Conduct frequent reviews of model reliability. Update validation criteria as feedback indicates. 	<ul style="list-style-type: none"> Deploy a real-time performance monitoring dashboard Engage third-party auditors for system validation Establish a system for integrating user and stakeholder feedback into continuous improvement cycles.
KPIs	<ul style="list-style-type: none"> Baseline reliability metrics. Validation test results. User feedback on system performance. 	<ul style="list-style-type: none"> Testing framework documentation (such as model cards). Reports on system reliability metrics. Regular review of model/system outcomes. 	<ul style="list-style-type: none"> Analytics dashboard in use. Third-party validation certificates. Records of feedback integration and system improvements.

3.4 Security and Resilience

The policies, practices, and activities in place to manage AI systems' robustness against cyber threats and their ability to recover from attacks.

Question	Company Response		Best Practice
<p>Please indicate the level of the company's RAI practices for managing Validity and Reliability</p>	<p>Level 3 Intermediate RAI Practices</p>		<p>Level 2 Basic RAI Practices</p>
Recommendations	Level 2	Level 3	Level 4
<p>Actions</p>	<ul style="list-style-type: none"> Identify critical assets/vulnerabilities. Implement basic cybersecurity (encryption, access controls). Develop an initial incident response plan. 	<ul style="list-style-type: none"> Implement fundamental cybersecurity practices, such as encryption and access controls. Develop an incident response plan for quick action in case of breaches. Establish business continuity plans specific to AI system disruptions. 	<ul style="list-style-type: none"> Implement fundamental cybersecurity practices, such as encryption and access controls. Develop an incident response plan for quick action in case of breaches. Establish business continuity plans specific to AI system disruptions.
<p>KPIs</p>	<ul style="list-style-type: none"> List of critical assets and vulnerabilities. Implementation of cybersecurity measures. Incident response plan. 	<ul style="list-style-type: none"> Business continuity plan. Security risk assessment report Cybersecurity measures implemented. 	<ul style="list-style-type: none"> List of advanced threat detection systems in operation. Third-party audit reports confirming security and resilience standards.

3.5 Safety

The policies, practices, and activities are in place to ensure that AI applications do not harm users or society and comply with safety standards.

Question	Company Response			Best Practice
<p>Please indicate the level of the company's RAI practices for managing Validity and Reliability</p>	<p>Level 3 Intermediate RAI Practices</p>		<p>Level 2 Basic RAI Practices</p>	
Recommendations	Level 2	Level 3	Level 4	
Actions	<ul style="list-style-type: none"> Perform a preliminary safety risk assessment. Develop safety protocols based on identified risks. 	<ul style="list-style-type: none"> Establish a dedicated safety officer for continuous risk analysis and mitigation. Implement safety testing protocols for all AI deployments. Create incident response plans with action steps and responsibilities. 	<ul style="list-style-type: none"> Implement real-time safety monitoring systems. Engage in third-party safety audits and certifications. Foster a safety-first culture with regular drills, updates, and feedback loops for system users and stakeholders. 	
KPIs	<ul style="list-style-type: none"> Safety risk assessment report. Safety protocols implemented. 	<ul style="list-style-type: none"> Dedicated staff responsible for safety oversight. Documentation of safety testing protocols. Incident response plans and records or incidents or drills. 	<ul style="list-style-type: none"> Deployment of real-time safety monitoring systems. Third-party audit results and safety certifications. 	

3.6 Fairness and Bias

The policies, practices, and activities in place to identify, prevent, and correct biases in AI algorithms and datasets.

Question

Company Response

Best Practice

Please indicate the level of the company's RAI practices for managing Bias and Fairness

Level 3
Intermediate RAI Practices

Level 2
Basic RAI Practices

Recommendations

Level 2

Level 3

Level 4

Actions

- Engage diverse stakeholders early in the design process for feedback.
- Conduct preliminary bias audits.

- Regularly review and refine AI models for fairness.
- Document and address bias-related incidents.

- Formalize continuous bias monitoring mechanisms.
- Implement bias and fairness training.

KPIs

- Completion of bias audits and documentation of audit results.

- Fairness review cycles and update logs.
- Incident response and resolution records.

- Bias monitoring system implementation.
- Fairness training completion rates.

3.7 Explainability and Transparency

The transparency of AI decision-making processes and the ability for users to understand and trust AI outputs.

Question	Company Response		Best Practice
Please indicate the level of the company's RAI practices for managing Explainability and Interpretability	Level 3 Intermediate RAI Practices		Level 2 Basic RAI Practices
Recommendations	Level 2	Level 3	Level 4
Actions	<ul style="list-style-type: none">• Create user-friendly materials explaining AI outputs and chain of thought reasoning.	<ul style="list-style-type: none">• Create user-friendly materials explaining AI outputs and chain of thought reasoning.	<ul style="list-style-type: none">• Implement user feedback mechanisms to refine explainability approaches.• Regularly update AI models and explanations based on evolving standards.
KPIs	<ul style="list-style-type: none">• User research on explainability needs.	<ul style="list-style-type: none">• Availability of explanation materials for AI system outputs.	<ul style="list-style-type: none">• Records of user feedback and updates to AI models and explanations.

3.8 Environmental Impact

The policies, practices, and activities in place to manage the ecological impact of AI operations and initiatives to minimize carbon footprint and promote sustainability.

Question	Company Response			Best Practice
<p>Please indicate the level of the company's RAI practices for managing Validity and Reliability</p>	<p>Level 3 Intermediate RAI Practices</p>			<p>Level 2 Basic RAI Practices</p>
	<p>Level 3 Intermediate RAI Practices</p>			
Recommendations	Level 2	Level 3	Level 4	
Actions	<ul style="list-style-type: none"> • Adopt energy-efficient practices in AI model training and deployment. • Monitor and optimize energy consumption in AI operations. • Publicly report environmental impacts and sustainability efforts. 	<ul style="list-style-type: none"> • Adopt energy-efficient practices in AI model training and deployment. • Monitor and optimize energy consumption in AI operations. • Publicly report environmental impacts and sustainability efforts. 	<ul style="list-style-type: none"> • Adopt energy-efficient practices in AI model training and deployment. • Monitor and optimize energy consumption in AI operations. • Publicly report environmental impacts and sustainability efforts. 	
KPIs	<ul style="list-style-type: none"> • List of sustainability criteria considered in AI projects. 	<ul style="list-style-type: none"> • List of energy-efficient practices implemented. 	<ul style="list-style-type: none"> • Reports on energy consumption and optimization efforts. 	

4.0 Monitoring and Reporting

Question	Company Response		Best Practice
<p>Please indicate the level of the company's RAI practices for managing Validity and Reliability</p>	<p>Level 3 Intermediate RAI Practices</p>		<p>Level 2 Basic RAI Practices</p>
Recommendations	Level 2	Level 3	Level 4
Actions	<ul style="list-style-type: none"> List of RAI policies and processes. Detailed RAI audit and update schedules. Deployment of monitoring dashboards. Documentation of auditing processes and actions taken. Production of internal and external reports and disclosures. 	<ul style="list-style-type: none"> List of RAI policies and processes. Detailed RAI audit and update schedules. Deployment of monitoring dashboards. Documentation of auditing processes and actions taken. Production of internal and external reports and disclosures. 	<ul style="list-style-type: none"> List of RAI policies and processes. Detailed RAI audit and update schedules. Deployment of monitoring dashboards. Documentation of auditing processes and actions taken. Production of internal and external reports and disclosures.
KPIs	<ul style="list-style-type: none"> List of RAI policies and processes. Detailed RAI audit and update schedules. 	<ul style="list-style-type: none"> Deployment of monitoring dashboards. Documentation of auditing processes and actions taken. Internal and external reports and disclosures. 	<ul style="list-style-type: none"> Deployment of monitoring dashboards. Documentation of auditing processes and actions taken. Internal and external reports and disclosures.

5.0 Additional Resources

This section offers a curated selection of tools, guidelines, and expert insights designed to assist companies in designing and implementing responsible AI practices. This repository includes access to white papers, case studies, regulatory updates, and best practice manuals developed by leading AI ethics researchers and practitioners. Additionally, it features interactive webinars, workshops, and training modules tailored to various aspects of AI ethics, such as data privacy, bias mitigation, and sustainability. These resources provide practical guidance and actionable steps that help companies not only comply with current standards but also excel in their responsible AI initiatives, fostering innovation while ensuring ethical integrity and public trust.

- **Practices**

- [Google Responsible AI Practices](#)

- **Principles**

- [Google Responsible AI Principles](#)[IBM Ethical AI Principles](#)

- **Research**

- [Google Research Responsible AI Whitepapers](#)[Partnership on AI Resources](#)

- **Toolkits**

- [Google People and AI Tools](#)[Thoughtworks Responsible Tech Playbook](#)[IBM 360 Fairness Toolkit](#)

- [Microsoft Responsible AI Tools and Practices](#)

- [An Ethical Toolkit for Engineering Design/Practice by the Markkula Center for Applied Ethics](#)

- [Design Ethically](#)[Ethical Design Toolkit](#)[LIME](#)

- **Courses & Webinars**

- [Certified Ethical Emerging Technologist Professional Certificate](#)[Ethics of AI: Safeguarding Humanity by MIT](#)

- [AI Ethics: Global Perspectives Series](#)



For more info

Tracy Barba

Director

Lucas Institute for Venture Equity and Ethics

Markkula Center for Applied Ethics

Email tbarba@scu.edu

[www.https://www.scu.edu/ethics/rai/](https://www.scu.edu/ethics/rai/)