

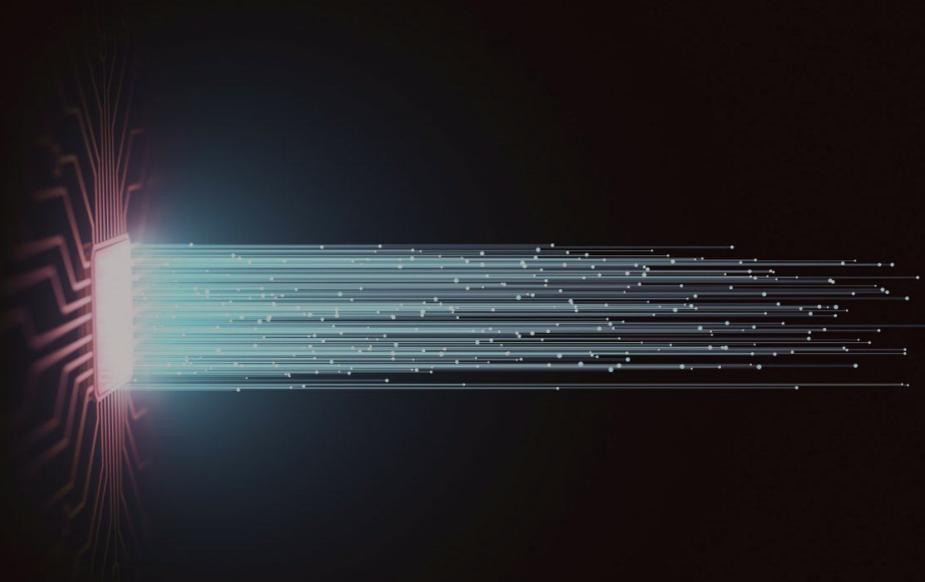


**Responsible
Investing in**

AI

**A GUIDEBOOK
FOR VCS**

Ravit Dotan, PhD



EXECUTIVE SUMMARY

CONTENTS

Part I: What is AI? What is AI ethics?

Part II: How AI ethics impacts materiality

Part III: How to conduct responsible AI due diligence

Part IV: Walkthrough: Due diligence case studies

Part V: How to support portfolio companies on responsible AI

Addendum: A flowchart of the responsible due diligence process

AI is booming. Most companies already use AI now or are planning to do so. In 2022, 77% of companies used and provided AI ([Accenture survey](#)). In 2020, VCs invested \$75 bn in AI, about 20% of all VC investments. In 2012, VCs only invested \$3 bn in AI in 2012, about 3% of all VC investments ([OECD report](#)). Showing a similar trend, in 2021, global corporate investments in AI reached \$176 bn, up from about \$5 bn in 2013. In particular, corporate M&A AI investment almost doubled between 2020 and 2021 ([Stanford's HAI report](#)).

AI can have severe unintended consequences. These influence companies' social impact and financial performance. To avoid pitfalls, investors need to understand the risks associated with AI investments.

This guidebook introduces investors to AI risks and provides guidance for investing in AI and data-heavy companies. You can find all the resources [here](#).

ACKNOWLEDGEMENTS

This guidebook was written by Ravit Dotan in collaboration with [VentureESG](#), it was published in December 2022. This material is based upon work supported in whole or in part by [the Notre Dame-IBM Tech Ethics Lab](#), [The Center for Philosophy of Science at the University of Pittsburgh](#), and the [Center for Governance and Markets at the University of Pittsburgh](#). Such support does not constitute an endorsement by the sponsors of the views expressed in this publication.

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

PART I

WHAT IS AI?

WHAT IS AI ETHICS?



WHAT IS AI? WHAT DOES IT DO?

One of the influential AI definitions is given by the draft of [the EU AI Act](#), the proposed bill that is expected to regulate AI in the EU. According to this definition:

Artificial intelligence system (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts (Article 3, November 2022 draft)

To get a sense of how AI works, see [this](#) 10-minute video.

AI is so widely adopted because it is versatile and effective. While AI is often associated with futuristic technologies such as self-driving cars or sentient robots, in reality, AI is primarily used in back-office operations that all companies need. For example, [a 2022 IBM survey](#) found that the top five AI use cases are automating IT, security and threat detection, automating business processes, marketing and sales, and business analytics.

WHAT IS AI ETHICS?

As the adoption of AI soars, the social and environmental costs of irresponsible AI become increasingly apparent. AI systems typically make decisions about and process the data collected from masses of people. When something goes wrong, it goes wrong at scale.

AI ethics is the field aimed at understanding and managing AI risks and opportunities for people, society, and the environment.

Prominent areas of attention include fairness and non-harm, data protection, transparency, explainability, and human autonomy and control. (This list and the definitions below are based on [a study led by Jessica Fjeld](#).)

Fairness and Non-harm

are about avoiding harm and avoiding replicating and exacerbating social inequalities. AI can conflict with fairness ideals when it wrongfully discriminates, which may happen when datasets and design choices are biased and when the AI's outputs impact minority groups differently.

For example, Northpointe's recidivism algorithm, COMPAS, predicted the likelihood that defendants would commit additional crimes. It was widely used in the US criminal justice system. In 2016, [ProPublica exposed](#) that the algorithm was biased against black people. Black people were twice as likely to be labeled "high risk" to re-offend but not go on to re-offend. White people were almost twice as likely to be labeled "low risk" to re-offend but go on to re-offend.

Data Protection includes protecting privacy, asking for consent before collecting and using data, allowing users to access, revise, and erase their data, and preventing data leaks.

For example, [Clearview](#) is a facial recognition startup. Given a photo of a person's face, their algorithm outputs other photos of that person found online and links to them. Clearview was accused and fined for violating data rights because they [scrapped billions of photos without their owners' consent](#) to create the algorithm.

Transparency

means that the system makes oversight easy.

Transparent AI systems notify users when interacting with them, when they make decisions about them, and when an AI system generates an artifact (such as an image or a video).

For example, [Nightingale and Farid \(2022\)](#) found that AI can generate images of faces that are indistinguishable and more trustworthy than images of real faces. Without marking the articles as computer-generated, they could be used to spread disinformation.

Explainability

means that technical concepts and decisions are translated into comprehensible content. Explanations help developers and users alike to evaluate algorithmic decisions and take action if needed.

For example, in 2014, [Amazon developed an AI system to sort through resumes](#). While testing it, Amazon noticed that the AI disadvantaged women, but they didn't know why. An examination revealed a comprehensible explanation: The algorithm penalized resumes that contained the word "women's" or more than one women-only college. Having this explanation at hand, Amazon decided to shut down the project.

Human Autonomy and Control

mean that humans should have control over what the AI does. First, people should be able to intervene in the actions of AI systems. Second, end users should be able to appeal automated decisions made about them and refuse to be subject to automated decisions.

For example, in 2012, [the Dutch government started using an AI system to review applications for childcare benefits](#). Humans didn't properly review the AI's decisions and it ended up falsely accusing about 26,000 Dutch parents of fraud. The reasons were sometimes as minor as forgetting to sign a form. The entire Dutch cabinet resigned as a result of the scandal.

ADDITIONAL RESOURCES

To learn more about AI ethics, see these lists of [documentaries](#), [podcasts and books](#), [newsletters](#), and [free online courses](#).

PART II

HOW AI ETHICS IMPACTS MATERIALITY

OVERVIEW: AI ETHICS AND MATERIALITY

Attention to AI ethics is correlated with improved financial performance.

94% of executives said that they think AI ethics improves shareholders' return on investment in a survey by [The Economist Intelligence Unit](#)

Companies with the highest returns from their AI engage in risk-mitigation practices more often than others, according to a [2021 McKinsey report](#)

AI ethics impacts financial performance for five reasons:

- Compliance
- Product quality
- Product adoption
- Talent attraction
- Reputation

COMPLIANCE

AI regulation is here and more is coming:

- **Proposed AI-specific laws** - many jurisdictions are currently drafting AI regulation.
- **Existing AI-specific laws** - some jurisdictions have already passed AI laws.
- **Non-AI-specific laws** - AI is subject to non-AI-specific regulation.

Proactively protecting against AI harms when developing and deploying AI is important for compliance with existing laws and for future proofing in the face of regulatory uncertainty.

You can learn more about AI-related regulation efforts [here](#). Below are a few examples.

Proposed AI Laws

The EU AI Act is the most prominent effort to regulate AI, and it is expected to pass into law soon. The Act divides AI applications into risk categories and defines rules for each category. Read about it [here](#).

The Algorithmic Accountability Act is the most prominent federal AI regulation effort in the US. If passed, it would require companies to conduct impact assessments on the automated systems they sell and use. Read about it [here](#).

Existing AI Laws

China passed a law regulating AI-based recommendation services. It is already in effect. Read about it [here](#).

New York City passed a law prohibiting employers from using AI for employment decisions unless they comply with certain guidelines. The law will coming into effect on January 2023. Read about it [here](#).

Non-AI-specific laws

Privacy laws have been successfully used to sue AI companies. For example, [Clearview](#), which [used millions of images without consent](#), has been restricted and accrued millions of Euros in fines for privacy violations in countries all over the world, including [Australia](#), [Canada](#), [Greece](#), [Italy](#), and the [US](#).

Non-discrimination laws have been successfully used to sue AI companies. For example, the US Department of Justice sued Meta for violating the Fair Housing Act. The DOJ argued that Facebook's housing ads algorithm targets people partly relying on protected categories, such as race. The parties reached a [settlement](#) that requires Meta to change its housing ads algorithm.

PRODUCT QUALITY

Many agree that prioritizing AI ethics improves AI systems.

97% of executives said that AI ethics is important to product and service innovation in a survey by the Economist Intelligence Unit ([here](#)).

The Economist Intelligence Unit [identifies](#) the following connections between responsible AI practices and the quality of AI systems:

- Being proactive against the risk of wrongful discrimination includes ensuring that the AI works well for diverse users. The result is **better performance across more kinds of user profiles**.
- Protecting user rights involves ensuring transparency and accountability to end-users. Transparency and accountability help companies **understand their products better** and facilitate **more efficient feedback loops**.
- When a company protects users' privacy, users are likely to trust it more and share **more data** with it. Since AI systems generally perform better when they have more data, the result is AI that works better.

PRODUCT ADOPTION

AI ethics improves adoption among individual consumers and business consumers alike.

60% of executives said that their company had already refused to work with an AI vendor due to ethical concerns, in an Economist Intelligence Unit survey ([here](#)).

75% of executives said that their company would refuse work with an AI if ethical concerns arose, in an Economist Intelligence Unit survey ([here](#)).

59% of individuals said that they would purchase more products if a company's AI-enabled interaction is perceived to be ethical, in a Capgemini survey ([here](#)).

34% of individuals said that they would stop engaging with a company altogether if ethical concerns arose, in a Capgemini survey ([here](#)).

TALENT

AI ethics is important for talent attraction and retention.

Respondents of all age groups rated ethical leadership as one of the top two things they look for in employers, in a Gallup Survey ([here](#)).

For example, Meta struggled with recruitment **after the Cambridge Analytica scandal**. Recruiters in the company [indicated](#) that the **acceptance rates of job offers decreased from 85% to 35-55%**, and the effect was the most pronounced among engineers. In addition, job candidates asked tougher questions about privacy during the interview process.

REPUTATION

Negative AI impacts can tarnish the reputation of those who develop, use, or invest in that AI system. Here are two examples.

[Clearview](#) scraped billions of public photos without users' consent

The New York Times, in their exposé about Clearview, called it "[The Secret Company That Might End Privacy as We Know It](#)." Clearview was also publicly criticized for bias issues in its algorithm, to the extent that Forbes described it as a "[Threat to Black Communities](#)."

AnyVision (now called [Oosto](#)) was [accused](#) of selling its algorithm for surveilling Palestinians

Microsoft, which invested in AnyVision, suffered reputational damage. Forbes, for example, ran a headline that read, "[Microsoft Slammed For Investment in Israeli Facial Recognition 'Spying on Palestinians'](#)." In the end, [Microsoft divested](#) from AnyVision.

ADDITIONAL RESOURCES

To learn more about how AI ethics impacts materiality, see:

- [The Economist Intelligence Unit, "Staying ahead of the curve: The business case for responsible AI"](#)
- [Capgemini, "Why Addressing Ethical Questions in AI Will Benefit Organizations"](#)

PART III

HOW TO CONDUCT RESPONSIBLE AI DUE DILIGENCE



OVERVIEW: RESPONSIBLE AI DUE DILIGENCE

When investors evaluate companies that develop, use AI, or process big data, they should examine them through the lens of AI ethics. The reason is that AI ethics improves companies' social impact and financial performance, as discussed above.

Below is a workflow investors can use when engaging with AI and data-heavy companies. Investors can integrate it into their due diligence process to evaluate companies and determine how to support them on AI and data responsibility. This workflow has been tailored to early stage VCs, but it is suitable for other kinds of investors as well.

The due diligence process includes five steps:

STEP 1: Decide whether the company poses AI risks

STEP 2: Evaluate the risk of conflict with regulation and with your values

STEP 3: Evaluate the company's responsible AI maturity

STEP 4: Decide whether the company is eligible for investment

STEP 5: If you invest in the company, prepare to support it on data and AI ethics issues

You can find a flowchart describing this workflow in the addendum and [here](#).

STEP 1: DOES THE COMPANY POSE AI RISKS?

AI is a powerful technology that can give rise to severe unintended consequences. Three types of companies can cause these harms:

- Companies that **develop AI**
- Companies that **use AI**
- Companies that **process masses of data**

If the company you're diligencing falls into one of these buckets, it may cause AI harms and **you need to take the following steps in the workflow**. If a company is at an early stage and doesn't have a product yet, it is enough that it is expected to fall into one of these buckets.

STEP 2: CONFLICTS WITH REGULATION AND VALUES

AI companies are at risk of conflicting with existing and upcoming regulations. The most prominent is the EU AI Act, the bill that is expected to regulate AI in the EU. In addition, AI applications may conflict with your values. For example, investors [raised concerns](#) that facial recognition threatens human rights.

Investors can divide AI applications into four risk categories, based on the severity of the potential conflicts with regulation in relevant jurisdictions and with their values.

EXTREME RISK	AI applications that are or are likely to become illegal	E.g., AI Applications that are outlawed by the EU AI Act
	AI Applications in sectors that are excluded by your values	E.g., weapons, gambling
HIGH RISK	AI applications that are or likely to become heavily regulated	E.g., AI applications that are on the EU AI Act's "high risk" list
	AI Applications that are at high risk of conflicting with your values.	E.g., facial recognition
MODERATE RISK	AI applications that are or likely to become lightly regulated	E.g., AI Applications that are in the EU AI Act's "transparency obligations" list
	AI Applications that are at moderate risk of conflicting with your values	
MINIMAL RISK	AI applications that are not expected to be regulated	E.g., AI applications that are not regulated by the EU AI Act
	AI Applications that are aligned with your values.	

EU AI ACT RUNDOWN

The EU AI Act is the most prominent upcoming AI regulation. It divides AI applications into risk categories and defines rules for each, from prohibition to no restrictions. Similar to GDPR, it will apply to those who work in or engage with the EU and is likely to have an extensive global impact. The Act's website is [here](#).

Below is a summary of the applications in each of the risk categories.

To Be Prohibited

AI applications that

1. Subliminally distort people's behavior in ways that are likely to cause physical or psychological harm
2. Distort people's behavior by exploiting vulnerabilities related to age, disability, social, or economic situation in ways that are likely to cause physical or psychological harm
3. Evaluate or classify people on their personality or social behavior leading to unfavorable treatment that is either disproportionate or pertains to unrelated social contexts
4. Perform 'real-time' remote biometric identification in publicly accessible spaces by law enforcement authorities or on their behalf, unless certain exceptions apply

To Be Heavily Regulated

AI applications that

1. Are intended to be used in the following areas:
 - Biometric identification and categorisation of natural persons
 - Management and operation of critical infrastructure
 - Education and vocational training
 - Employment, workers management and access to self-employment
 - Access to and enjoyment of essential private services and public services and benefits
 - Law enforcement
 - Migration, asylum and border control management
 - Administration of justice and democratic processes
2. AI Applications that are a product or a safety component in a product that requires conformity assessment in the EU, e.g., toys, vehicles, aviation, and medical devices.

To Be Lightly Regulated

AI applications that

1. Interact with humans (e.g., chatbots)
2. Are used to detect emotions or determine association with (social) categories based on biometric data
3. Generate or manipulate content ('deep fakes')

STEP 3: RESPONSIBLE AI MATURITY

Investors should evaluate the company's efforts to mitigate AI risks. To do so, investors can focus on three dimensions of the company's activities: their knowledge, their workflows, and their oversight. Investors can evaluate the company's performance in each of these dimensions and then combine it to an overall assessment of the company's responsible AI maturity.

To evaluate the company's responsible AI knowledge, workflow, and oversight, investors may use the following questions and metrics:

KNOWLEDGE

To what extent does the company understand prominent AI ethics themes?

Risk Articulation

How well can the company articulate how prominent AI ethics risks relate to it?

Diverse input collection

To what extent does the company collect diverse input about prominent AI risks its technology pose?

Employee Education

How extensively does the company educate its employees about prominent AI ethics risks?

WORKFLOW

To what extent do the company's workflows mitigate the risk of conflicts with prominent AI ethics themes?

Strategy and Measures

Does the company have an AI ethics strategy, including clear metrics and standards?

Implemented Procedures

To what extent do the company's workflows include AI ethics practices, including all stages of their development life cycle?

Incentives

To what extent do the company's internal incentive structures support the execution of its AI ethics strategy and procedures?

OVERSIGHT

To what extent does the company's oversight support compliance with prominent AI ethics themes?

Internal Reporting

To what extent does the company report on its AI ethics progress to internal stakeholders, such as a senior AI ethics owner?

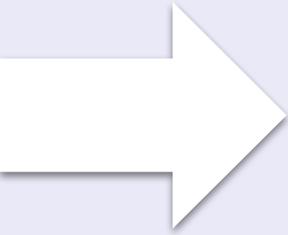
External Reporting

To what extent does the company report on its AI ethics progress to external stakeholders, such as its board?

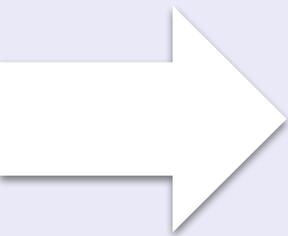
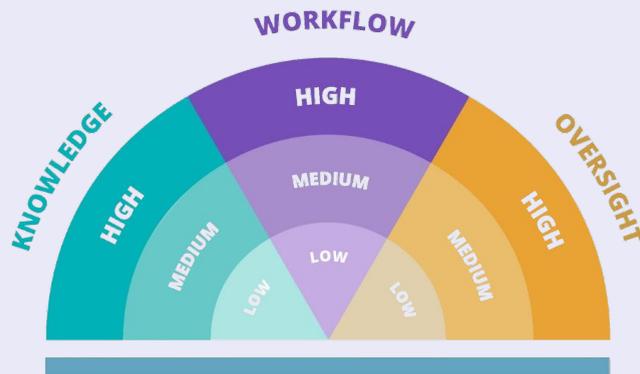
Periodical External Audits

How regularly and extensively does the company undergo external audits?

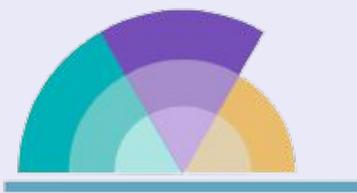
To Evaluate the company's responsible AI maturity



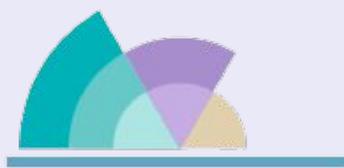
First, evaluate the company's performance in each dimension from low to high, using the suggested questions and metrics above.



Then, rank the company's overall responsible AI maturity from beginner to advanced, using the thresholds below.



ADVANCED



INTERMEDIATE



BEGINNER

At least:

- High knowledge
- High workflow
- Medium oversight

At least:

- High knowledge
- Medium workflow
- Low oversight

No minimum requirements

STEP 4: INVESTMENT ELIGIBILITY

Investors can use the information obtained in this due diligence process to shape their decisions on next steps, including whether to engage with the company, whether to require an external audit, and what AI ethics support to provide it. Recommend policies are as follows.

	Maturity Level (Step 3)	Risk Level (Step 2)
Do not invest	Any maturity level	Extreme
Require that they pass an external AI ethics audit as a condition for investing Unless the company is raising pre-seed (usually that's too early for an audit)	Beginner	Moderate
	Intermediate	High
Requires that they employ someone to be in charge of AI ethics issues	All except those that are both <ul style="list-style-type: none"> • Advanced maturity and • Moderate or minimal risk The form of employment will depend on the company, ranging from consultant to a full time employee.	

STEP 5: IF YOU INVEST, PROVIDE SUPPORT

After investing in any AI company, investors can and should help it improve its responsible AI maturity. That includes motivating the company and supporting education efforts.

Motivate

- Conduct periodic assessments of responsible AI maturity
- Put responsible AI issues on the board's agenda
- Include responsible AI expectations in your terms sheet or in a side letter

Educate

- Connect them with AI ethics experts and vendors
- Sponsor consultations with AI ethics experts
- Sponsor workshops for individual portfolio companies or groups of them

PART IV

WALKTHROUGH: DUE DILIGENCE CASE STUDIES



CASE STUDIES OVERVIEW

The case studies in this section are based on real companies. Some of the details in the companies descriptions have been changed to protect their privacy.

Three cases are discussed:

COMPANY 1 - a two-month old fintech company, raising its pre-seed round

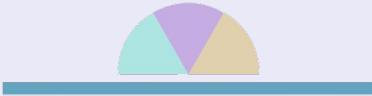
COMPANY 2 - a two-year old video generation and analysis company, raising its seed round

COMPANY 3 - a three-year old generative AI company, raising its A round

COMPANY 1: FINTECH, RAISING PRE-SEED

About the company:

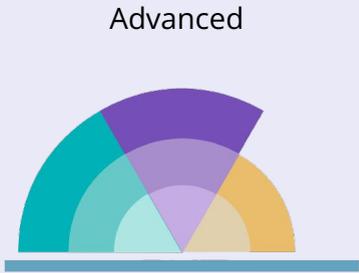
- Two-month old, consists only of founders, raising its pre-seed round
- Developing technology for fraud detection in financial services
- The founders are generally familiar with AI risks, but they think that these risks are only marginally applicable to their company. They don't plan to educate themselves or be proactive.

<p>STEP 1: Exposure to AI Risks</p>	<p>Exposed to AI risks</p>	<p>The company is expected to process big data. In addition, given that they need to analyze that data, there is a good chance that they will either develop or buy AI.</p>
<p>STEP 2: Level of Regulatory and Ethical Risk</p>		<p>The company is active in the financial sector, which is highly regulated. Without attention to AI ethics issues, the company may struggle to comply with regulation. E.g.:</p> <ul style="list-style-type: none"> • AI may illegally discriminate • AI may fail to provide legally-required explanations (e.g. on loan decisions) <p>In addition, AI in finance can create extensive unintended social harm. Depending on your values as investors, it could be enough to put the company in the "high risk" category.</p>
<p>STEP 3: Responsible AI Maturity</p>		<p>Knowledge level: Low They are familiar with AI harms, but minimize the relevance to their company. In addition, they are not educating themselves about AI ethics issues.</p> <p>Workflow & Oversight levels: Low AI ethics is not on the company's agenda.</p>
<p>STEP 4: Investment Eligibility</p>	<p>Eligible if they agree to work with an AI ethics expert</p>	<p>The AI ethics expert should educate the founders on AI ethics, how to identify AI ethics issues in their company, and how to mitigate risks of AI-related harm. Ideally, the expert would also guide and track responsible AI progress. At the very least, the company should meet with the expert periodically. The meetings may be group meetings with other portfolio companies.</p>
<p>STEP 5: If you Invest, Support</p>	<p style="text-align: center;">Educate</p> <p style="text-align: center;">Introduce the company to AI ethics experts Sponsor on-going support from an AI ethics expert Sponsor participation in an AI ethics workshop to jump-start their journey</p> <p style="text-align: center;">Motivate</p> <p style="text-align: center;">Ask for annual responsible AI progress reports</p>	

COMPANY 2: VIDEO ANALYSIS, RAISING SEED

About the company:

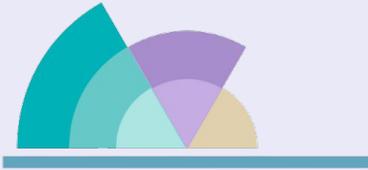
- Two years old, consists of 20 employees, raising its seed round
- Developing AI to assist in hiring decisions: job candidates interact with an AI-generated person over video, and the AI analyzes and evaluates the interaction.
- The founders are highly conversant in AI ethics risks and regulation. In addition, they have plans for compliance, mitigation of unintended bias, and data protection. For example, the company plans to undergo external risk assessment audits.

STEP 1: Exposure to AI Risks	Exposed to AI risks	The company's main business is developing AI
STEP 2: Level of Regulatory and Ethical Risk		<p>The company's main use case is employment systems, which is in the EU AI Act's "high risk" category. In addition, if they plan to be active in New York City, they would be subject to the city's law on AI for employment decisions, which will come into effect in January 2023.</p> <p>In addition, AI in employment can create extensive unintended social harm. Depending on your values as investors, it could be enough to put the company in the "high risk" category.</p>
STEP 3: Responsible AI Maturity		<p>Knowledge level: High The founders are excellent at articulating AI risks and the relevance of these risks to their company. They also educate their employees about it. The company not collect much feedback from diverse stakeholders, however.</p> <p>Workflow level: High They have a plan and measures for the mitigation of unintended bias. These are systematically integrated into the company's procedures and KPIs.</p> <p>Oversight level: Medium The company plans to undergo external audits and have AI ethics owner. However they don't report on AI ethics progress externally.</p>
STEP 4: Investment Eligibility	Eligible	The company's AI ethics owner is one of the founders. Match them with an AI ethics expert that can provide more support.
STEP 5: If you Invest, Support	<p>Motivate</p> <p>Make sure the board regularly discusses responsible AI progress</p> <p>Ask for annual responsible AI progress reports. In particular, monitor external audits to make sure they do them and pass.</p>	

COMPANY 3: GENERATIVE AI, RAISING A SERIES

About the company:

- Three years old, consists of 40 employees, raising its A round
- Developing AI to manipulate photos and generate photos from text
- The co-founders and other key actors in the company can speak to AI risks the company poses. They educate employees, discuss AI ethics with stakeholders, and have an AI ethics strategy and ongoing projects. They have a fractional AI ethics owner. However, they don't have AI ethics measures, e.g., they don't measure fairness in their algorithms and datasets. They also don't report on their responsible AI progress externally.

<p>STEP 1: Exposure to AI Risks</p>	<p>Exposed to AI risks</p>	<p>The company's main business is developing AI</p>
<p>STEP 2: Level of Regulatory and Ethical Risk</p>	<p>Moderate</p>	<p>Generative AI is included in the EU AI Act's "transparency obligations" list. They are subject to data-related regulation, such as GDPR and copyright regulation. These may constrain their data collection and usage.</p> <p>In addition, generative AI can be used to spread misinformation, which may create extensive unintended social harm. Depending on your values as investor, you could even class it as "high risk".</p>
<p>STEP 3: Responsible AI Maturity</p>	<p>Intermediate</p> 	<p>Knowledge level: High They can articulate risks, they collect AI ethics feedback, and they educate their employees.</p> <p>Workflow level: Medium The have a plan and some projects. But they have few implemented practices, and they don't use AI ethics metrics or incentives.</p> <p>Oversight level: Low They have an AI ethics owner, but they don't report on progress to anyone else.</p>
<p>STEP 4: Investment Eligibility</p>	<p>Eligible</p>	<p>The company is already supported by an AI ethics expert.</p>
<p>STEP 5: If you Invest, Support</p>	<p>Motivate</p> <p>Make sure the board regularly discusses responsible AI progress</p> <p>Ask for annual responsible AI progress reports</p>	

* This case study is based on the company discussed in the next part. However, the example above does not represent that company, as significant details have been changed.

PART V

HOW TO SUPPORT PORTFOLIO COMPANIES ON RESPONSIBLE AI



OVERVIEW: RESPONSIBLE AI SUPPORT

Investors are in a great position to help companies develop and use AI more responsibly.

To support companies in increasing their responsible AI maturity, investors can **educate and motivate** companies to improve in the three organizational dimensions discussed in the previous section:

- The company's **knowledge** of AI ethics
- How the company integrates that knowledge into its **workflows**
- The company's AI ethics **oversight**

In this section, you will find:

- Examples of things investors can do to improve portfolio companies' responsible AI maturity.
- A case study: responsible AI practices at the startup [Bria.ai](https://bria.ai).

[Bria.ai](https://bria.ai) is a startup developing generative AI. Bria's platform generates, changes, and repurposes images and videos. The company is between seed and series A.



The author of this guidebook is leading the AI ethics efforts at Bria.

KNOWLEDGE

To help organizations improve their understanding of AI ethics, investors can focus on **employee education**, **diverse input collection**, and **risk articulation**.

Risk Articulation

Encourage companies to articulate their understanding of the AI risks that their technology poses.

For example, ask the founders to explain how prominent AI ethics themes pertain to their business. Good focal points include issues around fairness and non-harm, transparency and explainability, data protection, and autonomy and human control. The output of the articulation can come in the form of a dedicated AI risks deck, an AI ethics mission statement, or a list of principles they are committed to. To support the company in its efforts to articulate risks, consider sponsoring consultations with experts and responsible AI workshops for groups of portfolio companies.

At Bria, AI risks are articulated in the company's code of conduct. It was one of the outputs of the employee-wide workshop that mentioned above.

Free and Non-profit

Resources for Responsible AI Education

- Lists of responsible AI [documentaries](#), [podcasts and books](#), [newsletters](#), and [free online courses](#)
- Toolkits: [The EthicalOS Toolkit](#), [Markkula Center Resources](#)
- Self-assessment questionnaires: [Kosa's Responsible AI Self-Assessment](#), [Digital Dubai's AI System Ethics Self-Assessment Tool](#), [Canada's Algorithmic Impact Assessment Tool](#)

You can find more [here](#).

Diverse Input Collection

Encourage companies to collect diverse input about the AI risks their technology poses.

For example, ask **pre-seed companies** to include AI ethics in customer discovery processes and purposefully seek conversations with diverse stakeholders.

Ask **more established companies** to conduct research by deploying surveys, convening focus groups, and building feedback channels into their platform to engage with diverse stakeholders.

At Bria, Sales, Marketing, and Customer Success participate in feedback collection on AI ethics issues. The Responsible AI owner meets regularly with the heads of these departments to discuss ethical concerns coming from prospective customers. The goal is to understand the concerns better and find ways to address them.

Questions to Help Build Feedback Channels

- How can the company gain access to diverse perspectives from all stakeholders?
- How can the company cultivate the literacy that different stakeholders need to engage with AI ethics?
- How can the company proactively seek ethics feedback about the AI systems it develops and uses?
- How does the company use stakeholder feedback to improve its AI system?

Employee Education

Encourage companies to educate themselves about AI ethics.

For example, ask **pre-seed** founders to develop a plan for self-education.

Ask **more established companies** to include AI ethics in onboarding and ongoing training. In addition, connect them with consultants and services specializing in AI ethics education.

At Bria, the AI ethics journey started with a company-wide workshop to discuss:

- Introduction to AI ethics
- Best and worst case scenarios - participants discussed two questions: What is the worst-case scenario for Bria regarding AI ethics? What would it mean for Bria to excel in AI ethics?
- What should Bria do to increase its AI responsibility? Participants discussed strategies for increasing AI ethics knowledge, workflow integration, and oversight.

For new employees, Bria added a responsible AI component to the onboarding process. In addition, Bria holds periodic meetings to discuss AI ethics issues.

WORKFLOW

To help organizations improve their implementation of AI ethics, investors can focus on AI ethics **strategy and measures**, **implemented procedures**, and **incentives**.

Strategy and Measures

Encourage companies to develop responsible AI strategy and measures.

For example, ask **pre-seed companies** to set red lines to determine which features the company will not develop or buy. Consider sponsoring consultations with experts.

Ask **more established companies** to measure responsible AI performance and set minimum thresholds. For example, many companies need to decide how to measure and set thresholds for fairness.

At Bria, the company created the following high-level plan to guide the company's responsible AI (also available on Bria's website, [here](#)):

- Develop our knowledge and implementation of responsible AI on an ongoing basis. Our focus topics include bias, privacy, explainability, transparency, and copyright issues.
- Include responsible AI in the workflows of all departments, including R&D, sales, marketing, HR, and creative.
- Prioritize proactive prevention of bias - in our datasets, measurement tools, and output
- Build guardrails for responsible AI, privacy, and copyrights:
 - Do not generate synthetic videos which speak ("talking heads")
 - Mark the images we generate as synthetic in the file properties
- Give free access to nonprofit and academic users who can use it to promote goals that are aligned with our values, including
 - Democratizing creativity
 - Mitigating risks that deep fakes pose to people and society
 - Encouraging diversity and inclusion in visuals
- Make a meaningful contribution to the conversation about responsible AI in the sector of synthetic imaging.

Free and Non-profit

Resources for Responsible AI Workflows

[IBM's AI Fairness 360](#), [IBM's AI Explainability 360](#), [Meta's Casual Conversations](#) dataset for diverse videos and images, [the Data Nutrition Project](#) for creating dataset transparency, [Fairlearn](#) for improving fairness. You can find more [here](#).

Implemented Procedures

Encourage companies to implement responsible AI procedures in their workflows in all relevant departments, such as R&D, Sales, Marketing, Customer Success, Procurement, and HR.

For example, in **pre-seed companies**, if the company's workflows consist of following a to-do list on a shared doc, make sure that AI ethics has a prominent place on that to-do list. Consider sponsoring consultations with experts.

In **more established companies**, the rule of thumb is that the company's workflows, whatever they are, should have systematic and holistic responsible AI facets.

At Bria, workflows include:

- The Responsible AI owner regularly meets with the heads of all the departments, including R&D, HR, Marketing, Sales, and Customer Success. The goal is to identify responsible AI growth opportunities, execute plans, and ensure that things stay on track.
- The company is currently working on implementing responsible AI solutions around transparency, copyrights, and bias mitigation.
- In R&D, new features and updated existing features undergo ethics review.
- In HR, the company has added responsible AI to its code of conduct and is added responsible AI components into onboarding and ongoing training.
- In Marketing, Sales, and Customer Support, the company is identifying AI ethics pain points that various stakeholders bring up.

Incentives

Encourage companies to align their incentive structure with their responsible AI goals, including having responsible AI OKRs and KPIs.

For example, ask **pre-seed companies** to set goals around AI ethics education and around the diversity of people they do customer discovery with.

In **more established companies**, ask for concrete goals around responsible AI metrics, such as fairness, explainability, and transparency.

At Bria, the company is working on defining goals for fairness and stakeholder feedback.

OVERSIGHT

To help organizations improve their AI ethics oversight, investors can focus on AI ethics **internal and external reporting and on external audits.**

Internal Reporting

Encourage companies to report their responsible AI progress to internal stakeholders, such as a responsible AI owner and their employees.

For Example, ask **pre-seed companies** to discuss responsible AI progress at periodic company-wide meetings.

Ask **more established companies** to put someone in charge of responsible AI and include responsible AI progress in internal progress reports.

At Bria, there is a person who is leading responsible AI efforts. Among other things, that person tracks the progress of all responsible AI projects, and reports about it directly to the CEO and to the heads of all departments.

External Reporting

Encourage companies to report their responsible AI progress to external stakeholders, such as yourself as an investor, their board, shareholders, end-users, and the public.

For Example, ask **pre-seed companies** to report on responsible AI progress to their board and to you as an investor. Consider conducting periodic responsible AI assessments and including responsible AI expectations in your terms sheets or in a side letter.

Ask **more established companies** to include responsible AI in external progress reports, such as reports that are available to their board, shareholders, end-users, and the public. Consider conducting periodic responsible AI assessments and including responsible AI expectations in your terms sheets or in a side letter.

At Bria, the company presents its responsible AI strategy and progress at conferences, the company's website, and documents such as this one.

External Auditing

Encourage companies to undergo external audits for responsible AI issues relevant to the company, such as fairness, transparency, explainability, and human control.

For Example, ask **pre-seed companies** to find out which kinds of audits are relevant to them and what they would need to do in the future to pass.

Ask **more established companies** to periodically undergo external responsible AI audits, and consider sponsoring the audits. Companies are often ready for audits from the seed stage and on.

At Bria, the type of responsible AI audits that is especially relevant is a fairness audits, to evaluate bias in the datasets and various design choices.

Free and Non-profit

Resources for Responsible AI Oversight

- [AI for Good](#) - Conducting AI ethics audits
- [EthicsGrade](#) - An ESG company that covers tech risks. You can find free rankings of tech companies, including AI ethics aspects, on their website.
- [ForHumanity](#) - Conducting AI ethics audits and certifying auditors
- [International Digital Accountability Council](#) - Conducting technical and policy investigations, the reports are public.
- [Internet Safety Labs](#) - Conducting safety audits
- [Ranking Digital Rights](#) - Evaluating digital platforms and telecom companies on how well they uphold human rights

You can find more [here](#).

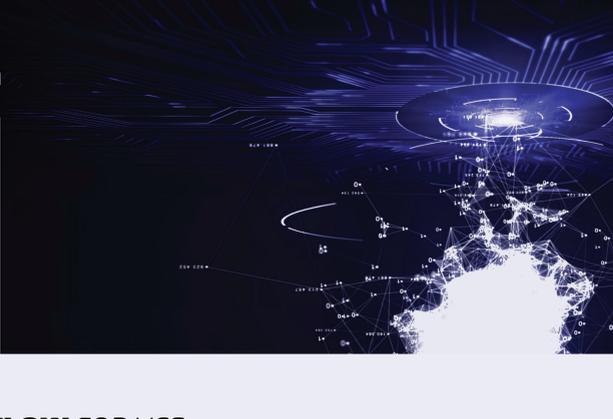
ADDENDUM RESPONSIBLE AI DUE DILIGENCE FLOWCHART



For your convenience,
below you can find a flowchart describing
the steps for conducting
a responsible AI due diligence.

The full flowchart and a summary of it
are available [here](#).

AI



A DUE DILIGENCE WORKFLOW FOR VCS

When investors evaluate companies that develop AI, use AI, or process big data, they should examine them through the lens of AI ethics. The reason is that **AI ethics improves companies' social impact and financial performance.**

Below is a **workflow investors can use when engaging with AI and data-heavy companies.** Investors can integrate it into their due diligence process to evaluate companies and determine how to support them on AI and data responsibility. This workflow has been tailored to early stage VCs, but it is suitable for other kinds of investors as well.

ADDITIONAL RESOURCES

In the full **guidebook**, you will find information about:

- What is AI and AI ethics?
- The business case for AI ethics
- How to conduct a responsible AI due diligence (a verbal explanation of the workflow)
- How to support portfolio companies on responsible AI issues

STEP 1:

Decide whether the company **POSES AI RISKS**

AI is a powerful technology that can have severe unintended negative consequences for people, society, and the environment. Companies are at risk of creating these harms if they develop AI, use AI, or process big data. If a company is at an early stage, it is enough that the company is expected to fall into one of these buckets.

You can read more about AI harms in [the guidebook](#).



STEP 2:

Evaluate **REGULATORY & VALUES CONFLICT RISKS**

AI companies are at risk of conflicting with existing and upcoming regulations. The most prominent is the EU AI Act, the bill that is expected to regulate AI in the EU. In addition, AI applications may conflict with your values. For example, investors raised concerns that **facial recognition may threaten human rights.**

EXTREME

Regulatory Conflict Risk

Risk of conflict with AI-related regulation
E.g. the EU AI Act

Prohibited by law

AI applications that are or are likely to become illegal
E.g., The EU AI Act outlaws applications that

1. Subliminally distort people's behavior in ways that are likely to cause physical or psychological harm
2. Distort people's behavior by exploiting vulnerabilities related to age, disability, social, or economic situation in ways that are likely to cause physical or psychological harm
3. Evaluate or classify people on their personality or social behavior leading to unfavorable treatment that is either inappropriate or pertains to unrelated social contexts
4. Perform 'real-time' remote biometric identification in publicly accessible spaces by law enforcement authorities or on their behalf, unless certain weapons apply

AI applications that are

Excluded by your values

E.g., weapons, gambling

HIGH

Highly regulated

Applications that are or likely to become heavily regulated
E.g., The EU AI Act heavily regulates applications that

1. Are intended to be used in the following areas:
 - Biometrics identification and categorization of natural persons
 - Management and operation of critical infrastructure
 - Education and vocational training
 - Employment, workers management and access to self-employment
 - Access to and enjoyment of essential private services and public services and benefits
 - Law enforcement
 - Migration, asylum and border control management
 - Administration of justice and democratic processes
2. Are a product or a safety component in a product that requires conformity assessment in the EU, e.g., toys, vehicles, aviation, and medical devices.

AI applications that are at

High risk of conflicting with your values

E.g., facial recognition may threaten human rights

MODERATE

Lightly regulated

Applications that are or likely to become lightly regulated
E.g., the EU AI Act lightly regulates applications that

1. Interact with humans (e.g., chatbots)
2. Used to detect emotions or determine association with (social) categories based on biometric data
3. Generate or manipulate content (deep fakes)

AI applications that are at

Moderate risk of conflicting with your values

MINIMAL

Not regulated

AI applications that are not expected to be regulated

AI applications that are

Aligned with your values

Don't Invest

Investment Eligible
Jump to step 5

STEP 3:

Evaluate **RESPONSIBLE AI MATURITY**

To evaluate the degree to which the company mitigates data and AI risks, investors can focus on three dimensions of the company's activities: their knowledge, their workflows, and their oversight. Evaluate the company's performance in each of these dimensions and then combine it to an overall assessment of the company's responsible AI maturity.

You can find case studies in [the guidebook](#).

KNOWLEDGE

To what extent does the company understand prominent AI ethics themes?

Recommended metrics:

Risk Articulation
How well can the company articulate how prominent AI ethics risks relate to it?

Diverse Input Collection
To what extent does the company collect diverse input about prominent AI risks its technology pose?

Employee Education
How extensively does the company educate its employees about prominent AI ethics risks?

WORKFLOW

To what extent do the company's workflows mitigate the risk of conflicts with prominent AI ethics themes?

Recommended metrics:

Strategy and Measures
Does the company have an AI ethics strategy, including clear metrics and standards?

Implemented Procedures
To what extent do the company's workflows include AI ethics practices, including all stages of their development life cycle?

Incentives
To what extent do the company's internal incentive structures support the execution of its AI ethics strategy and procedures?

OVERSIGHT

To what extent does the company's oversight support compliance with prominent AI ethics themes?

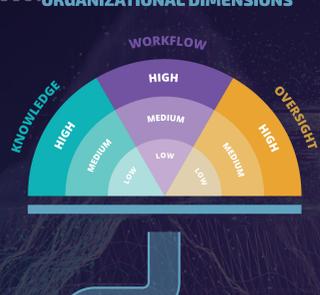
Recommended metrics:

Internal Reporting
To what extent does the company report on its AI ethics progress to internal stakeholders, such as a senior AI ethics owner?

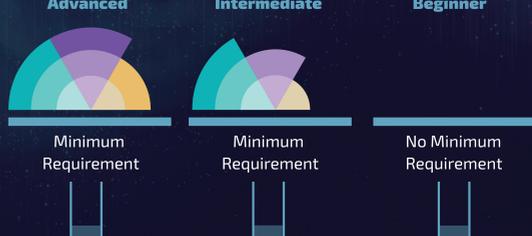
External Reporting
To what extent does the company report on its AI ethics progress to external stakeholders, such as its board?

Periodical External Audits
How regularly and extensively does the company undergo external audits?

Evaluate Key ORGANIZATIONAL DIMENSIONS



Evaluate the company's **OVERALL RESPONSIBLE AI MATURITY**



STEP 4:

Determine **INVESTMENT ELIGIBILITY** by combining:

The maturity evaluations from step 3
AND
The risks evaluations from step 2

Investors can use the information obtained in this due diligence process to shape their decision on next steps, including refusing to engage with the company, providing it with AI ethics support, and requiring an external audit.

You can find case studies in [the guidebook](#).



STEP 5:

Provide on going **RESPONSIBLE AI SUPPORT**

After investing in a company, investors should help the company improve its responsible AI maturity.

You can find case studies in [the guidebook](#).

Motivate
Ways to motivate companies to increase their responsible AI maturity include:

- Conducting periodic assessments of responsible AI maturity
- Putting responsible AI issues on the board's agenda
- Including responsible AI expectations in your terms sheet or in a side letter

Educate
Ways to educate portfolio companies on responsible AI issues include:

- Connecting them with AI ethics experts and vendors
- Sponsoring consultations with AI ethics experts
- Sponsoring workshops for individual portfolio companies or groups of them

