ALTIMETER®
@PROPHET

RESEARCH REPORT

Innovation + Trust

The Foundation of Responsible Artificial Intelligence

BY SUSAN ETLINGER
JULY 9, 2019

TABLE OF CONTENTS

During the past year, we've seen a dramatic increase in public conversation about the ethics of Artificial Intelligence (AI). Concerns about the impact of biased algorithms, lack of transparency, and how AI technologies are being and should be used have consistently garnered headlines in major news outlets around the world, part of a "techlash" against the large technology platform companies, such as Facebook, Google, and Amazon.

In response, some of these companies have instituted principles, advisory boards, checklists, and other tactics to prove that they are thoughtful and responsible stewards of not only data and algorithms, but of our broader experiences in the digital and physical world. But while news coverage of the techlash has raised awareness of ethical and experiential issues of AI, it generally hasn't addressed the implications for the organizations that buy, build, and implement these technologies to power digital experiences.

This report, based on interviews with leaders from business, academia, and non-governmental organizations defines ethical issues resulting from the development and deployment of AI, lays out the approaches being tested and used, and proposes an approach to begin to address them. While this is a highly complex field and no single strategy is appropriate for all industries or companies, this report may be used as a guide to better understand the unique implications of AI, begin to socialize them within your organization, and build organizational capability to foster both trust and innovation among customers, employees, shareholders, partners, and the general public.

RESPONSIBLE AI IS FOUNDATIONAL TO TRUST

The rapid commercialization of AI is only a few years old, but anxiety about the human and economic impact of new technologies is a constant historical theme. In 1492, a German Benedictine abbot named Johannes Trithemius published *De Laude Scriptorum (In Praise of Scribes),* in which he warned of the impact of Gutenberg's printing press on the quality of manuscripts and feared for the livelihood of the monks who painstakingly copied them. Roughly 350 years later, Samuel F. B. Morse sent the first telegraph message from Washington to Baltimore: "What hath God wrought?"

We're at a similar crossroads with AI. It's a powerful set of technologies with massive potential for business and society. But the ability of intelligent technology to enable organizations to act and learn autonomously and at scale is precisely why we need to be clear-eyed about its implications. The choices we make about AI today will affect our ability to innovate in the future. As importantly, those choices will determine the level of trust we are afforded by customers, consumers and the world at large. In an increasingly connected world, trustworthy digital experiences will become a critical differentiator.

Today the majority of news stories on responsible AI focus not on global business but on the large platform companies such as Facebook, Amazon, and Google. But the good news is that we have learned a lot from their very public successes and more public missteps. Today we've reached a teachable moment, an occasion to reflect on what happens when we innovate at the expense of trust or focus on trust at the expense of innovation. Trust and innovation aren't a zero-sum game; both must complement each other so we can build digital organizations and institutions that embody our values, treat people with dignity and respect, and, in so doing, drive growth and build brand preference and loyalty.

> **The choices we make about AI today will affect our ability to innovate in the future.**
>
> — Susan Etlinger, Altimeter

AI ETHICS DEMYSTIFIED

Given that our lives — as patients, customers, consumers, resident, and everyday people — are increasingly influenced and defined by algorithms, it stands to reason that we must be able to build, deploy, and use them confidently. To do this, we first need to clarify a few points:

1. **Definition**

   AI exacerbates a number of existing trends, such as automation of human labor, and introduces new ones, such as algorithmic bias. But to take into account the desirable and undesirable implications of AI, it's important to understand the unique challenges AI presents and how to think of "ethics" in this context.

2. **Scope**

   Given that AI research and development are predominantly located within technology organizations, it seems reasonable to assume that technologists are primarily responsible for how these technologies are deployed. But because AI effectively gives organizations the ability to act and learn at previously impossible scale, its implications are far-reaching, affecting teams from research to engineering to product development, marketing, human resources, and legal and compliance.

3. **Stakeholders**

   Given the potential impact of AI across the business, it's important to clarify both who should be responsible for identifying and addressing ethics issues and who are the stakeholders across and outside the organization.

4. **Risk**

   While the regulatory environment is still nascent, it's important to plan for the risks, both legal and reputational, of instituting (or not instituting) ethical AI practices, as well as norms for legal, research, and product teams to work together.

5. **Value**

   It's (relatively) easy to identify societal harms, such as discrimination, but businesses also have a responsibility to customers and shareholders to identify, prioritize, organize for, budget, and measure the impact of issues and the effectiveness of solutions.

Defining AI

While there are as many arguments for how to define AI as there are definitions, two key attributes distinguish it from other types of technology: It can act autonomously, and it can learn, sometimes without human intervention.[1] These characteristics mean that decisions made by AI can compound over time, for better or worse. Better outcomes can include efficiencies, insights, and unanticipated opportunities. Worse outcomes can include unwanted assumptions that, when encoded into algorithms, perpetuate and even amplify errors or inequalities, often without visibility, guardrails, or an understanding of the impact.[2]

Defining Ethics

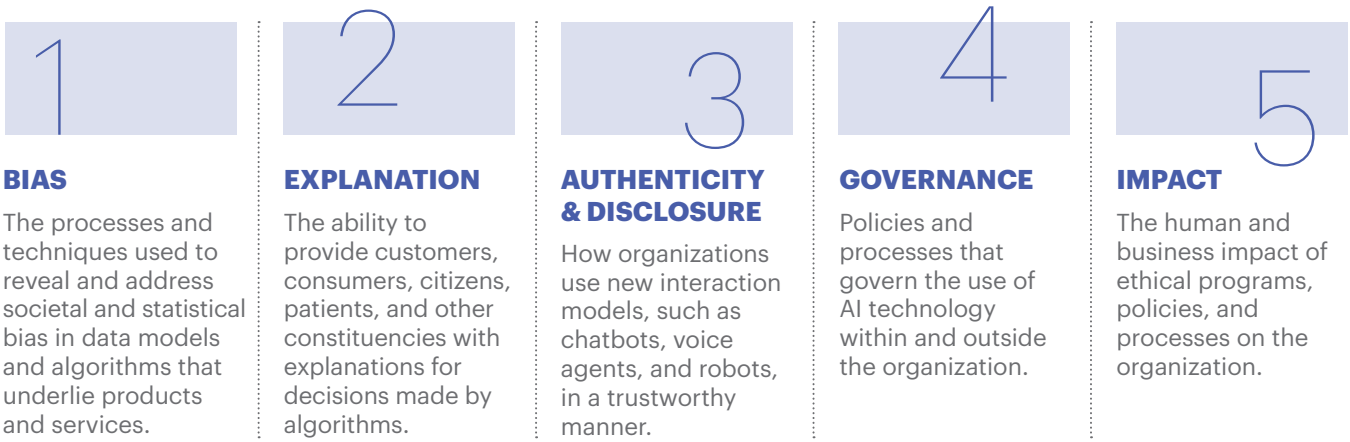
Ethics is, generally speaking, a system of moral principles that govern human behavior. We accept and understand these principles in the physical world, and they range from the trivial to the fundamental. Some are obvious and bound by law: Don't steal; don't hurt people. Some are subtle and unwritten: Face forward when standing in an elevator. Some are specific to certain cultures: Do present a business card with both hands. Don't show people the soles of your feet.

In the digital world, however, we don't always have norms for appropriate behavior, nor do we have a way of addressing their short- and long-term impact. Some of the dynamics that tend to hinder ethics initiatives include:

- **Homogeneous teams**
  Lack of diversity in organizations often means that the very people who are most affected by technology do not have input into its design.
- **Confusion**
  While there are many issues that AI either accelerates or exacerbates (such as data collection and use, automation, and others), some are unique to autonomous, self-learning technologies (such as facial recognition, for example) and therefore require specific interventions.
- **Lack of visibility into the impacts**
  We know that hitting someone hurts them, but we don't always know how a flaw in a dataset or data model, or lack of transparency in decision-making, affects people and, by extension, institutions, populations, and brands.

Figure 1 lays out the primary ethics topics that are specific to AI:

Figure 1. Key AI Ethics Issues for Business



BIAS AND DISCRIMINATION

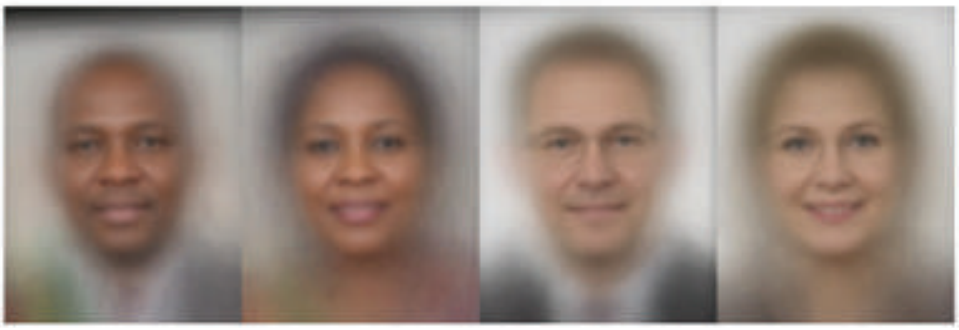
Bias in algorithms, and how it translates to human and business impact, is perhaps the most-discussed issue related to AI ethics. AI is fundamentally based on statistical reasoning: It uses patterns found in historical data (models) to predict the most-efficient path to the preferred outcome and executes a decision based on that prediction. Algorithms trained on data that naively represents a historical social or political bias will likely make decisions that replicate the patterns found in the historical data, often with harmful real-world consequences, especially for vulnerable

populations. This can take the form of who is deemed eligible for a loan, who gets parole, and who is served recruitment ads on social networks. From a business standpoint, bias can undermine audience segmentation, recommendation engines, ad targeting, personalization, or any system that relies on business, language, image, or video data to draw conclusions or make predictions.

For example, the "Gender Shades Project" led by Joy Buolamwini, a researcher at the MIT Media Lab and founder of the Algorithmic Justice League, and Timnit Gebru, a research scientist at Google AI, demonstrated how the facial recognition software of leading technology companies tends to return far higher error rates when analyzing the faces of women of color compared to white men. "In the worst case," Buolamwini wrote in a July 2018 *New York Times* OpEd, "the technology was 34% less accurate for [dark-skinned African] women than it was for white men" (see Figure 2).[3]

Figure 2. Gender Shades Company Performance Metrics

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



SOURCE: JOY BUOLAMWINI, GENDER SHADES PROJECT, CC-BY-NC-ND

This is one thing when tagging photos in social media; it's quite another when companies are using face-recognition technology to decide whom to hire, to identify a criminal suspect from a photograph, or simply using biometric data to unlock the door to a workplace.[4] Because platforms use the same core technology across multiple use cases, a bias of this type can spread widely throughout governments and the economy, replicating and exacerbating inequalities that those entities might otherwise wish to work against.

EXPLANATION

AI is an extremely powerful set of technologies with a wide range of use cases. Yet while we can see the inputs (data) and the outputs (decisions, recommendations, predictions) from machine-learning algorithms, they are not always easy or even possible to interpret. Sometimes, the lack of explainability is a problem that can be easily addressed; at other times, explainability may be unachievable due to the differences in type between human and computational cognition. There may also be trade-offs between explainability and accuracy, as a more interpretable model may unintentionally obscure predictive yet unknown factors.

Researchers around the world are working on building high-performance and explainable models to ensure that AI and machine-learning technology is able to provide customers, consumers, citizens, patients, and other constituencies the rationale for decisions that affect their lives.[5] This can be as simple as a "Why am I seeing this post?" option in social media or as complex as a rationale for granting or denying a loan, or parole.

The ability to interpret the rationale for recommendations or decisions made by algorithms is critical so that the organizations building and deploying AI can use these technologies in an informed, trustworthy, and reliable manner; in the European Union, according to GDPR, it is a right.[6] At the same time, unless a person has a similarly sophisticated algorithm advocating on their behalf, that right may be aspirational, or entirely unenforceable. Says Dr. Vivienne Ming, Executive Chair and Co-Founder, Socos Labs, "A right that you have no mechanism for exercising is not a right."

AUTHENTICITY AND DISCLOSURE

Currently, AI presents two primary issues related to trust in digital products, services, and systems:

- **Realistic robots**
  Increasingly "lifelike" chatbots and voice agents that can fool people into thinking they're human; and
- **Synthetic data and content (deepfakes)**
  Technology that can automatically generate synthetic content, whether it is text, audio, video, or a combination.

The conversation about authenticity, which began in earnest in the aftermath of the 2016 elections, peaked again during the 2018 Google I/O conference, when CEO Sundar Pichai demonstrated Google Duplex, a technology that enables a voice agent to replicate naturalistic human speech.[7] The demonstration was so effective that it alarmed critics concerned about transparency and worried about how realistic robotic technology could be misused. Google quickly clarified that the product would include disclosure so that people interacting with it would be aware that they were interacting with a voice agent rather than a human being.[8]

Deepfakes

The ability for AI to generate synthetic data and content (also known as "deepfakes") is also extremely powerful. It can be used legitimately or illegitimately for entertainment purposes, business purposes, or political ones. In May 2019, the Dali Museum in St. Petersburg, Florida, introduced an exhibition entitled "Dalí Lives" in which visitors could watch a video of Dalí constructed entirely from images and audio of Salvador Dalí during his lifetime and even take a selfie with the artist, who died in 1989.[9]

IMAGE: THE DALÍ MUSEUM

Recently, the nonprofit OpenAI announced GPT-2, a new model that can generate language, useful for applications such as virtual assistants, prediction, and content generation. OpenAI claimed its model was so effective that it was concerned about the potential for abusing it. They chose to release a partial version, prompting both accolades and criticism.[10]

Synthetic content has legitimate applications in entertainment, media, and elsewhere, but it can also be used as a weapon of disinformation. In April 2018, comedian and film director Jordan Peele teamed with BuzzFeed to educate viewers about this issue; they produced a fake public service announcement in which former President Obama seemingly warned viewers about the dangers of fake news.[11] In another example, hackers manipulated a video of Parkland shooting survivor Emma Gonzales tearing up a gun-range target, replacing it with a copy of the Constitution.

Says John Villaseñor, a Senior Fellow at the Brookings Institution, "Because they are so realistic, deepfakes can scramble our understanding of truth in multiple ways. By exploiting our inclination to trust the reliability of evidence that we see with our own eyes, they can turn fiction into apparent fact. And, as we become more attuned to the existence of deepfakes, there is also a subsequent, corollary effect: They undermine our trust in all videos, including those that are genuine. Truth itself becomes elusive, because we can no longer be sure of what is real and what is not."[12] This is clearly a danger to civic life, but it is also potentially threatening to brand marketers and content strategists, as it is becoming relatively easy to hijack a brand or a person's image for nefarious purposes.

REGULATION AND GOVERNANCE

Given the speed at which AI technology is moving, regulations, policies, and processes that govern its use lag far behind. As typically occurs with major technology shifts, such as the Internet and cloud computing, legislation will eventually address some issues, but legal frameworks conceived in one century inevitably fail to account for the technologies of the next.

Currently, the General Data Protection Regulation in the European Union has some provisions for algorithmic transparency and explanation, in addition to detailed provisions for data collection and use, but it is enforceable only for "data subjects" (e.g., people who can be identified by their data) in the European Union. In the United States and elsewhere, situations vary given regulatory and cultural differences.[13]

It will inevitably fall to governments and organizations to determine how AI technologies fit within their norms and how they will address the circumstances that fall outside applicable regulations and governance frameworks.[14] These issues include questions such as internal controls and documentation, allowable use cases, and employee training and development.

MEASUREMENT

There is a growing conversation within organizations about how to measure the business value of deploying (or not deploying) ethics programs, policies, and processes and the impact to brand reputation, trust, customer loyalty, and so forth. Alphabet Inc. recently warned investors that flawed AI could affect revenues and operating results, while Microsoft warned of potential "brand or reputational harm."[15]

What does all of this mean for business leaders? Simply that if we intend to unlock the potential of AI both to optimize and even transform organizations, we need to be clear-eyed about the values of our organizations, the issues we need to manage, the risks we will likely encounter, the opportunities that lie on the other side, and how we intend to measure progress.

A ROADMAP FOR RESPONSIBLE AI

In many organizations, the impetus for responsible AI comes from people within and outside the organization: data scientists, developers, marketers, product managers, and, most importantly, customers concerned about the implications and effects of autonomous technologies. Yet even in organizations that were early advocates for responsible use of technology, not every leader or every employee agrees on precisely what it means and what form(s) it should take.

It's tempting to assume that because AI projects originate within data science and engineering organization that these groups are exclusively responsible for designing and building ethical applications, products, and services. Some organizations have instituted a sort of Hippocratic Oath for data scientists and engineers in which they are asked to swear to "do no harm" with data, while others have built checklists for data scientists to help them identify areas where data sets or data models may yield problematic results.

While these tactics help to promote awareness and a sense of accountability, they can also set up a false sense of security and can shift responsibility and blame to technical staff. Technical teams cannot operate in an ethics vacuum, nor can they solve every issue of bias, interpretability, or transparency using technology alone. Says Jake Metcalf, Founding Partner at Ethical Resolve, a company that helps companies operationalize their values, "A lot of what we hear from data scientists and machine-learning engineers is, 'We're talking about this at work, but we don't know what to do or how to operationalize it.'"

> **This is not a technology problem alone.**
>
> — Priya Vijayarajendran, Vice President Data and AI at Microsoft

"This is not a technology problem alone," argues Priya Vijayarajendran, Vice President Data and AI at Microsoft. "It's a multidisciplinary problem that many people have to come together to solve." To build trustworthy products and services requires a set of values, principles, processes, policies, and shared accountability — from research to product development to operations.

To be clear, no organization has yet built out a fully functioning operating model for ethical AI or ethical technology; we are still at the very earliest stages of AI, not to mention understanding what ethics means in the digital world. But practices are emerging that, taken together, can form the basis for a roadmap.

Figure 3 lays out the steps needed to build and socialize a foundation for responsible AI.

Figure 3. Planning for Responsible AI

1 IDENTIFY ISSUES TO ADDRESS

Ethics is by its nature contextual, so it's critical to build ethics capacity in a way that supports organizational values and culture. The first step is to identify and clearly define key issues that matter to the organization as a whole: how it learns about customers, how it treats them, and situations in which machine-learning technologies can introduce risk.

For many organizations, that means starting by looking for and trying to remediate unwanted bias in data sets and data models. Says Ethical Resolve's Metcalf, "Bias is the most promising area for leveraging these topics into organizational change because it's computational. It feels to engineering departments that this is something for which you can build a solution that is technically tractable in ways other things might not be."

At the same time, however, Metcalf cautions teams not to assume that addressing bias in data automatically equals fairness. "Bias is a subset of the fairness problem, and fairness is not technically tractable. There is no answer to what is fair. When we talk about bias, we're often talking about fairness, but there is no universal agreement on fairness, and there never will be." As a result, bias and fairness issues must honor corporate values and existing governance frameworks to the extent possible.

One of the most common fears about AI ethics initiatives is that they will slow development and time to market. But some organizations have found the opposite to be the case. Rumman Chowdhury, Responsible AI Lead at Accenture, believes that establishing norms for responsible AI can actually be an accelerator for progress and is critical to the success of AI overall. "To get to better decision-making, you have to have responsible AI in place," she states. "People need to be able to make ethical decisions, so there's a balance between guidance and good judgment. We can't rely only on checklists, because there is a point at which you have to exercise good judgment. People are worried right now because the industry doesn't have a culture of norms. So the challenge is to cultivate that culture and community within organizations."

> **To get to better decision-making, you have to have responsible AI in place.**
>
> — Rumman Chowdhury, Responsible AI Lead, Accenture

Recommendations

Map AI ethics to corporate values	If inclusiveness or trust is a corporate value, are there places where AI-enabled products, services, or systems can enhance or undermine it?
Agree on priority issues	Find issues that are important and tractable, such as a lack of representation in a dataset, potentially discriminatory attributes in a data model, or an update of terms and conditions or master service agreements to start. Useful questions might include the following: 1. Does our data reflect our current and desired customer base? 2. Are there unwanted biases in our data that lead to stereotypical or otherwise undesirable outcomes? 3. Do we have policies in place that clarify how our values translate to decisions related to product development, sales, recruitment, marketing? 4. Do we have an opportunity to deliberate on these issues and determine an approach during product development? 5. Do we have processes in place to ensure that we continually assess our data and systems for unwanted bias? 6. Do we have provisions in our Master Service Agreements or Terms of Service to ensure that our partners and customers know, understand, and abide by our policies?
Align with legal teams	Work with counsel to understand legal implications to ensure that teams do not unwittingly put the company at risk of liability.
Foster a culture of deliberation	Clarify rules of engagement and decision-making processes for internal discussions. Part of the value of an ethics initiative lies in the deliberative discussions that it stimulates and the organizational capabilities it helps develop internally.
Communicate	Set expectations up front about what issues the team seeks to address, who the stakeholders are, criteria for decision-making, and when and how you will communicate progress.

2 EMPOWER A CORE TEAM

The first principle of AI ethics teams is that they should be small, agile, trusted, and resourceful. They should not become a bureaucracy, a roadblock to innovation, or, worse, "the ethics police." And because AI is still relatively new, AI ethics teams should function as a change-management resource that is useful until the organization develops enough knowledge and resilience internally to address AI ethics more broadly. As a result, even in the largest organizations, ethics teams are generally very small, are composed of existing staff, and focus on ethics in addition to their "regular" roles. Few are exclusively dedicated to AI or technology ethics more broadly.

Recommendations

Domain expertise and passion	Identify people with domain expertise on AI and/or ethics and social science and a passion for the topic.
Diversity	Diversity (on the core team and among stakeholders) is critical to reflect as many types of experience as possible. This goes for role and level as well.
Tolerance for ambiguity	This is a new and complex topic for many businesses, so it's important to be able to navigate through ambiguity and be willing to take considered risks.
Clear scope	Every organization is different, but following is a representative scope for corporate AI ethics teams: 1. **Identify issues.** Help researchers, data scientists, engineers, operations, sales, marketing, and other colleagues identify, understand, prioritize, and resolve issues; 2. **Principles.** Develop and/or refine and socialize a set of principles to guide decision-making and policy development; 3. **Governance.** Develop and/or refine policies and processes; 4. **Products and services.** Embed ethics into product and service development processes and practices and help colleagues in compliance, legal, HR, and elsewhere instantiate them into the business; 5. **Employee engagement.** Create safe channels for employees to raise issues and for teams to deliberate on them; 6. **Policy.** Collaborate with internal and external advisors to collect feedback and propose industry standards; and 7. **Communication.** Educate employees, customers, consumers, and external stakeholders on salient issues, and collect feedback from them.

3 DEFINE A STAKEHOLDER LIST

Larger organizations typically have a wide-ranging group of internal and external stakeholders, comprising research, engineering, product management, design, legal, human resources, communications, public policy, ecosystems, human rights organizations, and beyond. Beyond teams, however, one of the most fundamental drivers of ethical AI is a diverse and empowered workforce. Lily Jampol, Ph.D., Social Scientist at ReadySet, a diversity and inclusion consultancy, says, "AI is not really distinct from other technologies in that the more diversity of perspective and backgrounds you have, the more likely you are to spot problems you wouldn't have seen otherwise."

It's also critical to recognize that trustworthy technology is something that customers want and will ask for. "Our technology needs to be transparent and accessible," says Tatiana Mejia, Head of Production Marketing for Adobe Sensei. "We need to help people understand the levers and attributes that drove a particular decision. As our tools continue to enable scale, we need more transparency because they're moving so fast. Our customers are requesting ways of understanding how AI is driving the result that they're seeing."

Over time, as ethics programs mature within organizations, there will be few people explicitly tasked with ethics and more shared accountability across the organization. "There shouldn't be just one person in charge," says Michelle Carney, Lecturer at the Stanford d.school. "If we want to design ethical products and services for everyone, the best way to do it is to empower the people actually working on them to have accountability and make sure incentive structures are in place to enable employees to call things out as needed."

Recommendations

As with the core team, identify stakeholders that will represent the widest possible spectrum of background, experience, culture, domain expertise, and role within the organization.

INTERNAL STAKEHOLDERS	
Research and Data Science	Consider impacts of ethical and societal implications of AI models, and disclose them in research papers. This is standard practice in academia but emerging in business.
Product Development (Product Management, Engineers, User Experience Researchers and Designers)	Consider impact of using a protected data class in a data model, as that may lead to a biased decision. Look for collinearity (e.g., highly correlated variables, such as zip code and race) and data leakage, both of which can result in biased outcomes. Consider places in digital products where it is possible to alert users to low confidence or potentially ambiguous results. Conduct an ethics pre-check as a way to discuss whether a proposed product, service, or feature could potentially result in biased or unintended outcomes. If using agile methodologies: • Include the ethics pre-check in the "definition of ready" and "definition of done." • Build in a retrospective to identify what went well and what should be done differently next time. If your organization uses agile methodologies, adding ethics as an attribute of readiness, completeness, and success will scale.

INTERNAL STAKEHOLDERS, CONTINUED

Customers	Work with customers to educate and empower them: • If you are a business-to-consumer company, this may consist of disclosures in digital products and services, provisions in terms of use, and other digital content to educate consumers about ethical AI. • If you are a business-to-business company, this may consist of the above, plus white papers, executive briefings, and consultations with customers if they have questions or concerns.
Legal	Partner with your legal department to: • Better understand data privacy laws, such as GDPR and others, to ensure that you are not unintentionally exposing the company to risk. • Craft acceptable use policies that cover how customers and consumers may use your products and services in situations where there is no legal precedent. Educate yourself on emerging guidelines and regulations on how AI may be used within governments or countries.
Human Resources	Engage and partner to understand and update employee processes (onboarding, performance reviews, incentive structures) and resources (intranet, managers) to account for issues and opportunities raised by AI ethics.
Diversity and Inclusion	Diverse teams make scenario planning and bias identification easier and reduce the likelihood that a similar group of people will view a situation in the same way. Engage with D&I teams to identify places where ethics and diversity and inclusion programs can support each other.
Communications	Work with marketing and PR to understand how they engage with customers at the point of purchase and to develop clear, trustworthy messages about how you do (and don't) use AI.

OTHER STAKEHOLDERS

Non-Governmental Organizations	Engage with NGOs or other expert advocacy groups like World Economic Forum, Omidyar, Partnership on AI to ensure you are listening to and learning best practices, as well as sharing what you have learned.
Ethicists	Engage with applied ethicists at academic institutions to preview key decisions and collect insight on potential ethical issues and solutions.
Government	Engage and maintain dialogue with think tanks and policy advisors on issues related to using artificial intelligence inclusively, fairly and for the public good.
Shareholders	Begin to build, and continually test, a way to evaluate the impact of ethical technology use on key metrics such as brand preference, risk mitigation, customer and employee loyalty, and other key performance indicators (KPIs).

2 DEVELOP A STRATEGY AND AN ACTION PLAN

Once you've assembled an initial stakeholder list, the first thing to do is listen — not only to internal stakeholders, but to external stakeholders as well. Paula Goldman, Chief Ethical and Humane Use Officer at Salesforce, says, "These issues are really complex, and opinions differ in and across cultures, so we must be much more inclusive around finding solutions from civil society and beyond in a really meaningful way." This means actively engaging with outside and internal stakeholder groups, listening, collaborating on solutions, and building accountability.

A good first step is to collectively agree on a few key priorities to tackle first. This may mean clarifying responsible AI principles, evaluating the product development process, or updating terms of use. Whatever the starting point, plan both at a micro (where can we show impact first?) and a macro level (what issues are likely to emerge that we should plan for?).

The good news is that many organizations, such as the World Economic Forum, Omidyar Network, Data & Society, The AI Now Institute, and others, are excellent resources. "There is a community of practice around these issues", says Salesforce's Goldman. "It's really important not to reinvent the wheel. These problems can't be solved as islands — it's important to have as many civil society voices at the table as possible. If we're going to shift norms about technology, we have to work together."

Leverage Corporate Values and Principles

Companies like Microsoft, Workday, and Salesforce tie their ethics programs to established corporate values. For example, principles for trustworthy AI are prominently featured on the Microsoft website and permeate development at the firm.[16] They include:

- **Fairness**
  AI systems should treat all people fairly.
- **Inclusiveness**
  AI systems should empower everyone and engage people.
- **Reliability & Safety**
  AI systems should perform reliably and safely.
- **Transparency**
  AI systems should be understandable.
- **Privacy & Security**
  AI systems should be secure and respect privacy.
- **Accountability**
  AI systems should have algorithmic accountability.

Leverage Existing Processes

One way to foster norms is to leverage established processes and operations. "Companies like Intel and others have a commitment to human rights principles and an ethical code of conduct," says Chloe Autio, Policy Analyst, AI and Privacy Policy, at Intel Corporation. "So we see AI ethics as an extension of these principles and values rather than as something new." One place where companies can focus early on is the agile development process (or the product development process more generally); specifically:

- Instituting an ethics "pre-check" at the beginning of a sprint;
- Adding in ethics questions or checks to the definition of "ready" and the definition of "done" (DoR and DoD); and
- Instituting a retrospective to identify what went well and what should be done differently next time.

The benefit of this type of approach is that if all development teams use the same development methodology, these ethics "checks and balances" will scale and, as importantly, become consistent across the company. This also goes for companies using cloud-based AI platforms and services from companies such as Microsoft, Accenture, IBM, and Salesforce.

Leverage Emerging Technology Tools and Resources

While no technology tool or service can fully automate ethics within an organization (or elsewhere), research and tools are emerging that will begin to add scale and consistency to responsible technology programs. Over time, we will begin to see existing platforms include ethics assurance capabilities, and systems of record will emerge. Following are a few examples:

1. **Algorithmic impact assessments**

   The AI Now Institute, an interdisciplinary research center focused on understanding the social implications of AI, has published an "Algorithmic Accountability Policy Toolkit" that, while it is intended for legal and policy advocates, is also a highly useful resource for industry.[17]

2. **Ethics checklists for data scientists**

   Deon, a tool developed by DrivenData Labs, is a tool that enables data scientists to add an ethics checklist to data science projects to better understand and address issues such as informed consent, exposure of Personally Identifiable Information (PII), various types of bias, and other issues.[18]

3. **Bias checks**

   Accenture offers an AI testing service to enable users to check for bias in data sets and data models.[19] IBM has deployed trust and transparency capabilities for AI on the IBM cloud to enable users to better understand, monitor, and manage both inputs and outputs of data models.[20]

4. **Explainability**

   Microsoft has open-sourced InterpretML, an AI toolkit to enable developers to experiment with ways to interpret models and systems.[21]

5. **Governance and transparency**

   IBM has released a research paper that proposes the Supplier Declaration of Conformity (SDoC) as a framework to build trust in AI-enabled products and services. An SDoC is essentially a fact sheet that would include the product's "purpose, performance, safety, security, and provenance information to be completed by AI service providers for examination by consumers."[22] To operationalize ethics across the organization will require a combination of "red teams"; e.g., teams that can use adversarial methods to audit algorithms and systems of record that can document drivers for machine-based decision-making, changes to data models over time, and decisions made by humans.

Leverage Existing Customer Experience Principles and Guidelines

Because ethical norms underpin the customer experience, it's critical to think of them as an enabler of quality and customer experience rather than a hurdle to address. To do this, it's important to build relationships from the ground up and raise awareness internally. "You have to cultivate an environment of psychological safety," says Kathy Baxter, Ethical AI Architect at Salesforce, "where the people who are closest to the work feel safe finding issues. They're the ones most likely to find issues early, so why wouldn't you listen to that?"

One way to build an ethics mindset into product development is to incorporate discussion of ethical issues into the development process as a matter of course, rather than viewing it as an outlier or unanticipated roadblock. Following is a set of simple "thought-starter" questions that can be used to inform a business case, create a Product Requirements Document (PRD), during weekly scrum, and at other key points during the product development process.

Recommendations

As you develop a scope of work for responsible AI or other technology, consider these three categories:

1. Organizational values or principles
2. Processes and practices already in place
3. How the technology, product or service will be used

Following is a set of starter questions to inform planning.

Ethical Principles	What values do we uphold as an organization? Have we communicated and socialized those values? How could AI technology threaten or help us uphold those values? How could this product or service be abused or attacked, and can it be prevented? What external resources or stakeholders, such as human rights organizations, would help us better understand the potential impact of AI on our business?
Ethical Practices	What data science, engineering, design, and governance practices can we use to mitigate biases that conflict with our values AND our ability to innovate? How complex/valuable are they? Are there any places we can focus for quick wins and learning? What methodologies could make our products and services more trustworthy? Does our product create or exacerbate unwanted outcomes for customers, customers' customers, or society based on race, religion, gender, or other sensitive categories? What is the worst possible article that could appear in the press as a consequence of our product or service? How will we educate, reward, and protect employees for identifying and mitigating ethical risk? How will we know whether these practices work? What can we share with our ecosystem: vendors, partners, academia, the industry at large?
Ethical Use	What are our primary use cases for AI technology? Are there use cases we will not allow? What are the human, brand, legal, and business implications? How do we track the human impact of our product or service? Are some groups disproportionately affected? How do we make decisons and learn from them? How do we manage the trade-offs? Who will make those decisions? Whose interests, desires, skills, experiences, and values have we simply assumed rather than consulted? How will we educate our employees, customers, and other constituencies? How will we measure the impact on our business? What is our responsibility to employees? Shareholders? Vendors and partners? Society?

5 MEASURE IMPACT, TEST, LEARN

One of the core issues in AI ethics is the question of how organizations evaluate the impact of ethical AI programs on the business. One way to approach measurement is to set goals for the ethics program itself while aligning it with established organizational metrics. This would entail two types of evaluations: one to gauge momentum of the ethics initiative itself, and another to connect it to the business at large.

Recommendations

Start small

Focus on one to three key metrics that you want to better understand: Is the team able to identify and socialize principles and issues? Are there roadblocks to alignment or implementation? What is accelerating or slowing momentum?

Measure to learn

Especially at the outset, use measurement as a learning tool rather than as an indicator of performance until you have a relatively stable set of policies, processes, and tools and enough history to start to understand patterns and trends.

Look both at deliverables and impact

Potential deliverables:

- Has the team developed, socialized, and helped to establish recommendations, such as principles, processes, and practices?
- Are they being implemented internally? What is the feedback?

Potential impact:

- Is it possible to identify places where the team has remediated risk within the company?
- Is trust a differentiator for your product, service, or company?
  - Can you begin to understand the impact of ethics programs on customer and employee trust?
  - Are customers mentioning ethics and/or responsible processes as an indicator of trust in your brand?
- Can you see, anecdotally or with quantitative data, an impact on any of the following?
  - Customer churn
  - Brand reputation
  - Customer experience
  - Employee engagement/loyalty
  - Growth
  - Legal risk
  - Innovation/time to market given time spent debating courses of action

When in doubt, align with values

"A learning mindset is fundamental," says Tim O'Brien, GM, AI Programs at Microsoft. "Your organization may or may not already have a process for applying judgment to potential issues. You may have to invent a new workflow, but it must be motivated by supporting the values of the company."

CONCLUSION

AI is still an emerging discipline, and no one has all the answers. Listening, educating yourself and your organization, and agile experimentation are key. But while this is a complex topic, it is by no means a zero-sum game between trust and growth. O'Brien says, "The world is changing because of this technology, so we need to change too."

END NOTES

[1] Taddeo, Mariarosario and Floridi, Luciano. "How AI Can Be a Force for Good." Science magazine. August 24, 2018 (https://science.sciencemag.org/content/361/6404/751). This article says, "On the one hand, AI is fueled by data and therefore faces ethical challenges related to data governance, including consent, ownership, and privacy. These data-related challenges may be exacerbated by AI, but would occur even without AI. On the other hand, AI is a distinct form of autonomous and self-learning agency and thus raises unique ethical challenges."

[2] O'Neil, Cathy. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books. September 6, 2016 (https://www.amazon.com/dp/B019B6VCLO/ref=dp-kindle-redirect?_encoding=UTF8&btkr=1). This is a very useful guide to how algorithmic bias can creep into decision-making.

[3] Buolamwini, Joy. "When the Robot Doesn't See Dark Skin." The New York Times. June 21, 2018 (https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html). Please note also that some of the vendors mentioned in the Gender Shades study have used these findings to improve their results on face recognition.

[4] Levin, Sam. "Amazon Face Recognition Falsely Matches 28 Lawmakers With Mugshots, ACLU says." The Guardian. July 26, 2018 (https://www.theguardian.com/technology/2018/jul/26/amazon-facial-rekognition-congress-mugshots-aclu). In another example from the past year, Amazon's "Rekognition" software incorrectly matched the faces of John Lewis and six members of the Congressional Black Caucus with criminal mugshots. Of course, bias doesn't exist in a vacuum. In a report by the AI Now Institute entitled Discriminating Systems: Gender, Race, and Power in AI, the authors argue that: "... as the focus on AI bias and ethics grows, the scope of inquiry should expand to consider not only how AI tools can be biased technically, but how they are shaped by the environments in which they are built and the people that build them." (https://ainowinstitute.org/discriminatingsystems.pdf).

[5] One recent example is the Alpha release of InterpretML, developed by Harsha Nori and others at Microsoft Research (https://github.com/microsoft/interpret).

[6] Burt, Andrew. "Is there a 'Right to Explanation" for Machine Learning in the GDPR?" IAPP. June 1, 2017. (https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/).

[7] Pichai, Sundar. "Google Duplex Demo." 2018 I/O Developer Conference video via YouTube. May 13, 2018 (https://www.youtube.com/watch?v=zTtBRI9AhOU).

[8] Statt, Nick. "Google Now Says Controversial AI Voice Calling System Will Identify Itself to Humans." The Verge. May 10, 2018 (https://www.theverge.com/2018/5/10/17342414/google-duplex-ai-assistant-voice-calling-identify-itself-update).

[9] "Behind the Scenes: Dali Lives — Art Meets Artificial Intelligence." The Dali Museum video via YouTube (https://youtu.be/BIDaxl4xqJ4).

[10] Whittaker, Zack. "OpenAI Built a Text Generator so Good, It's Considered too Dangerous to Release." TechCrunch. February 2, 2019 (https://techcrunch.com/2019/02/17/openai-text-generator-dangerous/).

[11] Peele, Jordan. "Jordan Peele Uses AI, President Obama in Fake News PSA." Good Morning America video via YouTube. April 18, 2018 (https://www.youtube.com/watch?v=bE1KWpoX9Hk).

END NOTES

[12] Villaseñor, John. "Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth." The Brookings Institution. February 14, 2019 (https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/).

[13] In the United States, the AI Accountability Act, proposed in April 2019 by Sen. Cory Booker (D-NJ) and Rep. Yvette D. Clarke (D-NY) "requires companies to study and fix flawed computer algorithms that result in inaccurate, unfair, biased, or discriminatory decisions impacting Americans." The Commercial Facial Recognition Privacy Act proposed by Sen. Brian Schatz (D-HI) and Sen. Roy Blount (R-MO), "would strengthen consumer protections by prohibiting commercial users of facial recognition technology (FR) from collecting and re-sharing data for identifying or tracking consumers without their consent." On May 14, 2019, San Francisco became the first major city to ban facial recognition technology by law enforcement, a move that has been widely hailed by human rights advocates. For a sense of priorities in the United Kingdom, Singapore and elsewhere, see: Renstrom, Joelle. "The U.K. Wants to Become the World Leader in Ethical A.I." Slate. August 1, 2018 (https://slate.com/technology/2018/08/the-u-k-wants-to-be-the-world-leader-in-ethical-a-i.html); "The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems." AccessNow. May 16, 2018 (https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/); and Tan, Sumiko. "Singapore Releases Framework on How AI Can Be Ethically Used." The Straits Times. January 23, 2019 (https://www.straitstimes.com/singapore/singapore-releases-model-governance-framework-for-ai).

[14] Robertson, Adi. "A New Bill Would Force Companies to Check Their Algorithms to Bias." The Verge. April 10, 2019 (https://iapp.org/news/a/us-lawmakers-introduce-algorithmic-accountability-act/).

[15] Vincent, James. "Google and Microsoft warn investors that bad AI could harm their brand". The Verge. February 11, 2019. (https://www.theverge.com/2019/2/11/18220050/google-microsoft-ai-brand-damage-investors-10-k-filing).

[16] See https://www.microsoft.com/en-us/ai/our-approach-to-ai. Microsoft also includes specific guidelines for developers of conversational AI (chatbots and voice agents): See https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf.

[17] Algorithmic Accountability Policy Toolkit. AI Now Institute. October 2018 (https://ainowinstitute.org/aap-toolkit.pdf).

[18] See http://deon.drivendata.org/examples/.

[19] "Accenture Launches New Artificial Intelligence Testing Services." Accenture. February 20, 2018 (https://newsroom.accenture.com/news/accenture-launches-new-artificial-intelligence-testing-services.htm).

[20] Puri, Ruchir. "It's Time to Start Breaking Open the Black Box of AI." IBM. September 19, 2018 (https://www.ibm.com/blogs/watson/2018/09/trust-transparency-ai/).

[21] Caruana, Rich. "Creating AI Glass Boxes — Open Sourcing a Library to Enable Intelligibility in Machine Learning." Microsoft. May 10, 2019 (https://www.microsoft.com/en-us/research/blog/creating-ai-glass-boxes-open-sourcing-a-library-to-enable-intelligibility-in-machine-learning/).

[22] Arnold, Matthew et al. "FactSheets: Increasing Trust in AI Services Through Supplier's Declarations of Conformity." Cornell University. August 23, 2018 (https://arxiv.org/abs/1808.07261).

METHODOLOGY

This document was developed based upon online and in-person conversations with experts, market influencers, technology vendors, brands, think tanks, and others involved in AI and data science, as well as secondary research, including relevant and timely books, articles, and news stories. Our deepest gratitude to the following:

Accenture, Rumman Chowdhury, Responsible AI Lead

Adobe, Tatiana Mejia, Head of Product Marketing, Adobe Sensei

Code4Thought, Yiannis Kanellopoulos, Founder

Ethical Resolve, Jake Metcalf, Founding Partner and Ben Roome, Founding Partner

Intel, Chloe Autio, Policy Analyst, AI and Privacy Policy

Microsoft, Tim O'Brien, GM, AI Programs; Priya Vijayarajendran, Vice President Data and AI

ReadySet, Lily Jampol, Ph.D., Social Scientist and Diversity and Inclusion Consultant

Salesforce, Paula Goldman, Chief Ethical Use Officer; Kathy Baxter, Ethical AI Architect

Socos Labs, Vivienne Ming, Founder and Executive Chair, and Bulbul Gupta, Founding Advisor, Artificial Intelligence Think Tank for Human Potential

Stanford d.School Michelle Carney, Lecturer

Some of the material in this report previously appeared in Ethical AI: A Quick Start Guide for Executives, sponsored by Salesforce. All findings and analysis are independent and represent Altimeter's body of research.

ACKNOWLEDGEMENTS

I would also like to convey my gratitude to my many colleagues, friends, and contacts and industry experts who have made this report possible, either by contributing ideas, introductions, perspective, design, editing, research, scheduling, or moral support: Sharmeen Ahmed, Mitra Azizirad, Chris Benson, Jesse Berrett, Maureen Blandford, Amanda Casari, Danielle Cass, Alistair Croll, Ester de Nicolas, Alix Dunn, Jana Eggers, Brian Green, Michael Kleinman, Cheryl Knight, Mike Loukides, Arti Narayanan, Orlagh Neary, Lawson Parker, Carole Piovesan, Heather Roff, Alyssa Samuelson, Aubrey Littleton, Sanjna Parulekar, Irina Raicu, Yoav Schlesinger, Pete Skomoroch, Aaron Shull and Shannon Vallor. Your help and insights were invaluable and will enrich future research on this topic.

I would also like to acknowledge the many academics and industry experts whose work on this topic is foundational to our understanding of the human impact of technology and who deserve far wider recognition: danah boyd, Joy Buolamwini, Kate Crawford, Virginia Dignum, Timnit Gebru, Mary Gray, Anna Lauren Hoffman, Os Keyes, Safiya Umoji Noble, Cathy Smith, Molly Wright Steenson, Zeynep Tufecsi, Meredith Whittaker and so many more. As always, any omissions or errors are mine alone.

ABOUT US

SUSAN ETLINGER, INDUSTRY ANALYST

Susan Etlinger (@setlinger) is an industry analyst with Altimeter, a Prophet Company, where she publishes research and works with clients on issues related to data strategy, artificial intelligence, digital ethics and trust, and the impact of emerging technologies on business and society. Etlinger was named a 2016 LinkedIn top Voice for Technology, and her TED talk on Big Data has been viewed over 1.25 million times. She has been quoted in media outlets, including *The Wall Street Journal*, *The New York Times*, and *BBC*.

HOW TO WORK WITH US

Altimeter research is applied and brought to life in our client engagements. There are several ways Altimeter can help you with your business initiatives:

- **Strategy Consulting.** Altimeter creates strategies and roadmaps to help companies act on business and technology trends, including enterprise AI maturity. Our team of analysts and consultants work with global organizations on ideation and priority-setting, needs assessments, strategy roadmaps, and playbook development to address a range of strategic objectives.
- **Education and Workshops.** Engage an Altimeter speaker to help make the business case to executives or arm practitioners with new knowledge and skills.
- **Advisory.** Retain Altimeter for ongoing research-based advisory: Conduct an ad hoc session to address an immediate challenge or gain deeper access to research and strategy counsel.

To learn more about Altimeter's offerings, please contact sales@altimetergroup.com.

OPEN RESEARCH

This independent research report was 100% funded by Altimeter, a Prophet Company. This report is published under the principle of Open Research and is intended to advance the industry at no cost. It is intended for you to read, utilize, and share with others; if you do so, please provide attribution to Altimeter, a Prophet Company.

PERMISSIONS

DISCLAIMER