AI Ethics: What Leaders Must Know to Foster Trust and Gain a Competitive Edge
AI Ethics: What Leaders Must Know to Foster Trust and Gain a Competitive Edge ........ 1

1. Educate developers on how to avoid introducing bias.
2. Emphasize transparency and explainability.
3. Govern AI along its entire life cycle.
4. Seek external support for robust frameworks and rigorous tools.
5. Exercise an abundance of skepticism by asking tough questions and monitoring AI systems.
6. Embed AI ethics in your corporate culture.
7. Conclusion

AI Ethics Checklist .................................................................................................................. 4

Sponsor’s Viewpoint .............................................................................................................. 6
AI Ethics: What Leaders Must Know to Foster Trust and Gain a Competitive Edge

For years, organizations have been using artificial intelligence (AI) to automate manual tasks and improve products and services. But as real-world use cases for AI multiply, so too do the ethical implications of simulating human intelligence in machines.

Today, AI-powered chatbots help to screen job candidates, facial recognition systems keep workplaces safe from intruders, and sophisticated AI algorithms predict market trends and emerging customer demands. These examples deliver benefits ranging from increased productivity and employee well-being to competitive gains. But they also raise important questions about the ways in which organizations operate in society and how AI systems can impact the privacy, fundamental rights, and safety of the people they’re intended to serve.

“Until a few years ago, AI was all about optimizing the accuracy on some training data,” says Lise Getoor, professor of computer science at the University of California, Santa Cruz. But that’s changed, she says, as businesses place increasing emphasis on “the societal impacts” of what can happen when AI systems malfunction, are corrupted, or adopt human biases. These missteps can result in costly litigation, regulatory fines, lost revenue, reputational damage, and widespread mistrust in AI systems.

For years, organizations have been using artificial intelligence (AI) to automate manual tasks and improve products and services. But as real-world use cases for AI multiply, so too do the ethical implications of simulating human intelligence in machines.

“Over and over again, people who are incredibly well-meaning have found that AI can actually aggravate and introduce more bias.”

Lise Getoor, Professor of Computer Science, University of California, Santa Cruz

Fortunately, there are ways for organizations to capitalize on the competitive advantages of AI while fostering trust, eliminating bias, and embedding AI principles in the cultural fabric of an organization. This guide aims to share advice from industry experts on how AI, with the right best practices in place, can not only provide a strategic advantage but have a net positive impact on society. Following are six best practices for achieving that goal (and for more advice, see the “AI Ethics Checklist”).

1. Educate developers on how to avoid introducing bias.

The great promise of AI is its potential to remove human subjectivity and bias from critical business processes, such as hiring employees or granting customers loans or mortgages. But without the necessary precautionary measures, the opposite can, and does, occur.

Case in point: In 2014, Amazon discovered that an AI recruiting tool it had developed discriminated against women. Created to identify top job candidates, the model’s training data consisted of 10 years’ worth of resumes — mostly submitted by men. As a result, it automatically penalized resumes that included the word “women’s” and downgraded graduates of women’s colleges.

“Many times, when one applies AI in situations like candidate screening or recidivism-risk prediction, one is attempting to eliminate human bias,” says Getoor. “However, over and over again, people who are incredibly well-meaning have found that AI can actually aggravate and introduce more bias.”

Part of the problem, she says, is that “technologists just aren’t trained to think about bias” and its ethical implications when designing and developing AI systems.

Francesca Rossi agrees. An IBM Fellow and the IBM AI ethics global leader, Rossi says: “It’s important that designers and developers are aware that they could inject, even in an unconscious way, their own biases into an AI system when they choose the training data or when they make any other decision during the AI pipeline development process.”
Transparency is critical for the responsible development and use of AI. Yet all too often, developers don’t even know why an AI model has arrived at a particular outcome or conclusion. For AI to be ethical, it must be explainable, especially when making business-critical decisions.

For that reason, Rossi says that IBM has launched awareness-raising initiatives designed to educate and train developers, making them more aware of unconscious biases and helping them better understand this key AI ethics issue.

These guidelines, which can apply to any industry, include:

• Scheduling regular team reviews to avoid unintentional biases, including stereotyping, confirmation bias, and sunk-cost bias.

• Creating diverse teams comprised of individuals of different ages, ethnicities, genders, educational disciplines, and cultural perspectives to better detect bias.

• Encouraging greater accountability among developers for the success of AI systems by keeping detailed records of design processes and decision-making.

• Monitoring training data and outcomes continuously in order to quickly respond to issues.

• Investigating when bias is detected to understand where it originated and how it can be mitigated.

Organizations must be responsible for providing a complete methodology to developers on how to circumvent bias, Rossi says, such as what to do during the design and development process, including bias checking, mitigation, and testing, while making sure these processes are as easy as possible to adopt and integrate into current practices.

2. Emphasize transparency and explainability.

Transparency is critical for the responsible development and use of AI. Yet all too often, developers don’t even know why an AI model has arrived at a particular outcome or conclusion. For AI to be ethical, it must be explainable, especially when making business-critical decisions.

"AI can work well, but it can be a fluke," warns Abby Everett Jaques, a research fellow at Stanford University’s Center for Ethics in Society. Without greater transparency and openness into how an AI system is built, significant risks can emerge, Jaques says. These can range from ethical violations to simply not “knowing when an AI system might fail, especially when it’s out in the world and not in a controlled environment,” she says.

One step toward more transparent and trustworthy AI is “explainability” — a series of processes intended to help enable organizations to better understand the contributing factors that lead to decisions. To that end, the technical community has developed a number of open-source toolkits, such as IBM’s AI Explainability 360, that provide techniques for explaining and interpreting AI model decision-making. A range of explainability algorithms serves various audiences based on the types of questions they want answered.

For example, a bank customer whose mortgage application was denied may want to understand the primary reasons for the rejection and what she can do to produce a more favorable outcome. A data privacy officer, on the other hand, will want to examine multiple data points and decisions to better understand the behavior of an AI system as a whole and to ensure that it complies with stringent regulations.

Another transparency-focused initiative is the work of the FATE (Fairness, Accountability, Transparency, and Ethics in AI) research group at Microsoft that explores the complex social implications of AI, machine learning, data science, and automation. Projects include data sheet templates used for documenting the data sets used for training and evaluating machine learning models. The purpose of these templates is to encourage data set creators to carefully consider possible sources of bias in their data while, at the same time, providing data sheet consumers with greater transparency. FATE has also developed a fairness checklist, codesigned with AI practitioners, to support responsible AI products and services throughout their design, development, and deployment, in the same way a structural engineer relies on a checklist to ensure a building is sound.

3. Govern AI along its entire life cycle.

Even the most carefully collected and trained data can produce questionable outcomes if organizations fail to govern AI along its entire life cycle from ideation to final result. According to Getoor, this begins...
with business and technology leaders working together to determine, in the conceptual phase, “how an AI system will play out in terms of organizational and societal impact.”

Consider, for example, installing facial recognition systems in the workplace. Although intended to protect employees from ill-intended intruders, there are growing fears around how data sets of images of people’s faces are collected and can be redistributed without consent for more nefarious purposes.

To avoid crossing ethical boundaries in the early stages of development, an organization must consider AI as more than just a technology challenge, instead viewing it as a reflection of its corporate values.

“Almost any technology can be used in ways that are good and bad,” says Thomas Malone, professor of management at the MIT Sloan School of Management and the founding director of the MIT Center for Collective Intelligence. “Many of the issues people think of as ethical AI issues are really just issues about what we consider good and bad in general. We think of them as AI issues, but they’re really not.”

In fact, Malone says, AI can act as a magnifying glass, amplifying biases and questionable data sets that already exist in corporate databases. To return to the earlier example of AI’s use in employee recruiting, he offers the example of building a model trained using existing data sets on previously successful hires to screen new job candidates.

“On its face, I don't think there's anything problematic or unethical about the practice,” he says. “But it raises the issue of whether people who were successful in the company in the past were successful not because of their individual merit, but because of demographic characteristics, such as race or gender. The AI system could perpetuate such discrimination.”

However, by carefully examining the business objectives to be addressed by an AI model, scrutinizing the data it ingests, and continuously monitoring outcomes, organizations can work within the ethical limits of AI without compromising innovation.

4. Seek external support for robust frameworks and rigorous tools. Internal ethics committees, comprised of members from disparate teams, can help organizations address complex ethical questions and stay accountable for their AI initiatives. Yet many organizations are recognizing the added value in relying on external entities to provide guidance on the effective use of AI technology, along with robust tools.

Chief among these benefits is the ability to tap into a vast reservoir of experience. “Because AI can touch — and threaten — so many aspects of our lives, the only way to build and deploy it properly is for technologists to collaborate with experts in the social impacts of innovation: ethicists, social scientists, and so on. And those experts must be empowered,” Jaques says. In fact, because most third parties work with a wide array of clients from varying industries, they are typically adept at recognizing common patterns that could indicate red flags. This experience also lends itself to a better understanding of the AI software solutions in the marketplace, many of which organizations choose to purchase rather than build proprietary systems.

Another advantage of seeking outside assistance for AI ethics: a broader and more objective perspective on AI ethics issues. Many technologies allow for standardized testing and monitoring — best practices that companies can apply repeatedly across departments and industries. However, “there's something unique about ethics,” Malone notes. “If you're measuring cybersecurity risks, these are fairly objective questions. Most people are knowledgeable enough to have a similar definition of the right answer. But in the case of ethical questions, that's not always true. Different people have different opinions about what's ethical and what's not.”

Whereas employees’ opinions can be shaped by colleagues and influenced by incentives, third parties are more likely to provide independent guidance on ethical issues and consider multiple perspectives. “It’s easier to receive objective and more trustworthy answers from
people who aren't conflicted by the incentives offered inside an organization, such as pleasing the CEO,” Malone says.

At the same time, many organizations are looking to external sources for valuable self-policing tools. These include AI ethical and design audits — independent assessments that rigorously test and validate AI systems using a wide array of assessment frameworks and testing methods. In fact, an IDC survey reveals that only 50% of organizations reported having a formalized framework to encourage considerations of ethics, bias, and trust, leaving plenty of room for improvement.

Even government agencies are establishing ethical guidelines for AI systems. For instance, in April 2019, the European Commission (EC) released its ethical AI guidelines, developed by an independent group of experts. The guidelines define trustworthy AI as lawful (“respecting all applicable laws and regulations”), ethical (“respecting ethical principles and values”), and robust, both technically and socially. To qualify as trustworthy, AI initiatives must address the following seven requirements: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, nondiscrimination, and fairness; social and environmental well-being; and accountability. In April 2021, the EC followed up by releasing its recommendation for a formal AI regulatory framework that addresses the potential risks AI could pose to people's security and fundamental rights.

5. Exercise an abundance of skepticism by asking tough questions and monitoring AI systems.

Impressed by AI’s potential to mimic human intelligence, “many people aren’t critical enough when analyzing AI,” Getoor says. “It’s just an optimization algorithm. If you have bad data coming in or a poor objective, you’ll produce bad outcomes.”

One way for technology leaders to put their AI systems to the test is by retaining a healthy degree of skepticism and asking hard questions, says R. David Edelman, director of public policy at MIT’s Internet Policy Research Initiative and a former technology adviser to President Obama. Among the questions to consider, he says: “Who designed this system, and what are the incentives? Who’s accountable if they get it wrong — who’s at risk, whose brand, and whose reputation? What is the system designed to do, and is that really what it’s doing? What data was used to train it? Is the model designed for that specific purpose, or is it being repurposed? And have we done enough to think about how that system can be abused?”

AI ETHICS CHECKLIST

Keep the following practices in mind to successfully establish the fundamentals of trust in AI while reaping its competitive advantages.

- Educate and train developers to increase their understanding of the current issues around AI ethics and heighten their awareness of any unconscious biases.
- Create diverse data science teams comprised of individuals of varying ages, genders, ethnicities, and educational disciplines. Strive to include new culture perspectives and diverse ways of thinking by, for instance, including neurodiverse members on the team.
- Leverage a wide array of AI explainability tools to better understand and interpret the contributing factors to AI model decision-making.
- Strike a balance between AI innovation and continuously questioning an AI model’s alignment with corporate values, data integrity, and ethical boundaries.
- Turn to external partners for AI ethical and design audits — independent assessments that rigorously test and validate AI systems using a wide array of assessment frameworks and testing methods.
- Conduct an inventory of all AI algorithms to identify potential risks and evaluate their impact on various stakeholders.
- Establish in-house standards and principles for ensuring ethical use of AI. Seek external expert advice on using or adapting existing codes. Ensure that they’re routinely applied and updated as needed.
- Encourage leadership to play a critical role in embedding AI ethics into an organization’s DNA by modeling ethical behavior and communicating its value to employees.
Another application of critical thinking involves conducting an inventory of all AI systems. Doing so not only helps to identify potential risks and evaluate their impact on various stakeholders, but it can also expose AI technologies being developed without proper oversight and governance. In addition to this preemptive strike, organizations must continue to monitor AI systems for potential deficiencies on an ongoing basis to account for changes in data, algorithms, and even societal norms.

An abundance of skepticism should also be exercised before fully deploying an AI system. Luciano Floridi, professor of philosophy and ethics of information at the University of Oxford, encourages organizations to explore “what-if” scenarios by testing the AI system prior to deployment. He points to the example of a company that quickly launches a new same-day delivery service, only to discover that it unintentionally bypasses areas with predominantly minority populations. “If the organization had tested its software in a lab, it would have immediately discovered that its rules discriminate between communities,” Floridi says. Recognizing margins for error prior to launch can eliminate the risk of reputational damage, legal liabilities, and unwelcome media exposure.

6. Embed AI ethics in your corporate culture.
From increasing transparency to exercising caution, the skills and best practices around AI must become embedded in an organization’s culture to bring about sustainable change. “If developers are taught to use ethical skills in the same way they’re taught to look for bugs in code, then it just becomes part of their daily practice,” Jaques says. To prepare for this cultural shift, organizations must rely on innovative strategies to drive adoption of ethical AI practices. For instance, experts recommend rewarding developers who take steps to actively avoid bias, increase transparency in decision-making, and design AI models that align with corporate values.

Leadership is fundamental to incorporating AI ethics into an organization’s DNA. “It’s important to have ethics embedded in your company ethos. It should be part of who you are as a company,” Getoor says. “That’s something that comes from leadership.”

Malone agrees. “If the people in power in an organization let it be known, often implicitly by their behavior, that ethical violations aren’t tolerated, then it’s easier to prevent people who look up to those leaders from making those same ethical violations.”

For those same reasons, organizations should also develop, enforce, and update their own ethical frameworks for ethical AI. No set of standards will cover every eventuality, but an approach that touches on the FATE approach of fairness, accountability, transparency, and ethics is certainly a good place to start.

Conclusion
AI technologies are permeating every aspect of the business world, from manufacturing goods to meeting customer needs. The strategic advantages are clear to companies from all industries, provided they adopt an ethical approach to AI. Education, explainability, robust frameworks — all can minimize the risk of bias, foster trust, and lead to a positive societal impact.

“Ethics is an aspirational endeavor,” Getoor says. “Companies should always be looking to create a better organization — and a better society.”

Thomas Malone, Professor of Management and Founding Director of the Center for Collective Intelligence, MIT
A Q&A With Nigel Duffy

In this Q&A, Nigel Duffy, EY global AI leader, discusses why AI ethics should align with corporate values, who should oversee ethical standards, and how to view the issue in terms of risk management.

This conversation has been edited for clarity, length, and editorial style.

Q: What role should corporate values play in shaping AI ethics, and how can an organization position AI ethics in the broader context of its responsibility to a fair and just society?

Nigel Duffy: A company’s use of AI should be consistent with the values it espouses. Most corporations want to live up to these value statements, but to do so, they have to make sure that the way they use AI is consistent with them. The key is for organizations to see AI as a strategic tool, not a tactical one. This framing of AI promotes a more holistic discussion and ensures that AI is being implemented in ways that have significant and beneficial impacts both for the organization and for society as a whole.

For example, when organizations only discuss AI ethical questions or AI risks at an engineering level or at a technical level, they miss critical opportunities to mitigate those risks. In fact, many of the negative consequences of AI are largely driven by incorrect framing of the business problem, as opposed to a technical issue. By having conversations around what those risks are, and how they relate to the values and goals you have as an enterprise, your organization can put the right governance and controls in place.

Q: It seems that senior leaders are taking greater ownership of their organizations’ AI ethics today than they did in the past. Why is this so critical to AI success?

Duffy: I believe AI should be the responsibility of the most senior people in an organization for two reasons: One, most AI ethical problems arise around how the technology is being used, not at a technical level. And two, again, AI is a strategic tool, and it will define the future of many organizations. As a result, the opportunities and, of course, the risks associated with that strategic tool must be addressed by the most senior people in the organization.

Another important point is that attention to AI ethics is growing much faster than many organizational leaders understand. When AI risks materialize, they receive a lot of media attention. But regulators and legislators are also moving much faster than industry leaders recognize (as detailed in a 2020 global EY study conducted with The Future Society). For example, the new European Union (EU) draft document on AI regulation is likely to impact many organizations in the coming 18 to 24 months. It takes time to mobilize a response, so leaders really need to be thinking about these matters now.
“Many of the negative consequences of AI are largely driven by incorrect framing of the business problem, as opposed to a technical issue. By having conversations around what the risks are, and how they relate to the values and goals you have as an enterprise, your organization can put the right governance and controls in place.”

Just look at the impact AI has had on our democracies globally in the last few years, and how people are paying increasing attention to privacy and the way AI — and automated decision-making systems in general — have led to some undesirable outcomes. This is an issue the general public cares about, and regulators and legislators are responding accordingly.

AI systems already exert a huge amount of power and influence over our society, and the influence they have over our behaviors is pretty profound. For example, if you look at the social justice movement in the last year or two, AI systems are increasingly becoming a battleground for issues of social justice and the kind of society we want to build. Because of this, civic organizations, legislators, and regulators are rallying around these issues — and, again, responding to them faster than many industry leaders recognize.

Q. What steps can organizations take to better prioritize AI ethics and manage potential risks?

Duffy: One way to encourage organizations to pay closer attention to AI ethics is to view the issue through a risk management lens. It’s definitely important to think about AI in terms of societal consequences and social justice. But oftentimes, these terms are too squishy for enterprises to get their arms around or even to recognize who owns them. However, if you view AI ethics through a risk management lens, it becomes a little bit clearer who should own the effort and why it’s important to enterprises.

The European Commission appears to have taken exactly that approach to AI oversight. First, in 2019, it issued its “Ethics Guidelines for Trustworthy AI,” based on recommendations from its High-Level Expert Group. And earlier this year, the EC released a proposal for an AI regulatory framework that takes a risk management point of view heavily influenced by existing product safety regulations.

The reality is that ethical questions around AI are part of a larger AI risk management discussion — and a very important part of it. In the end, the AI ethics agenda will achieve a faster and more powerful impact if we can help bring the risk management organizations into the conversation early on. In fact, according to a global survey that EY teams conducted with the World Economic Forum and the University of Cambridge, leading adopters of AI have taken exactly that approach.

ABOUT EY

EY exists to build a better working world, helping to create long-term value for clients, people, and society and build trust in the capital markets.

Enabled by data and technology, diverse EY teams in over 150 countries provide trust through assurance and help clients grow, transform, and operate.

Working across assurance, consulting, law, strategy, tax, and transactions, EY teams ask better questions to find new answers for the complex issues facing our world today.

EY refers to the global organization and may refer to one or more of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a U.K. company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via ey.com/privacy. EY member firms do not practice law where prohibited by local laws. For more information about our organization, please visit ey.com.