Advancing AI ethics beyond compliance

From principles to practice

IBM **Institute for Business Value**

IBM.

How IBM can help

Clients can realize the full potential of artificial intelligence (AI) and analytics with IBM's deep industry expertise, technology solutions, and capabilities and start to infuse intelligence into virtually every business decision and process. IBM's AI & Analytics Services organization is helping enterprises get their data ready for AI and ultimately achieve stronger data-driven decisions; access deeper insights to provide improved customer care; and develop trust and confidence with AI-powered technologies focused on security, risk, and compliance. For more information about IBM's AI solutions, visit ibm.com/services/ai. For more information about IBM's AI platform, visit ibm.com/watson.

By Brian Goehring,
Francesca Rossi, and
Dave Zaharchuk

Talking points

Disparate views on AI's ethical challenges
Executives and consumers view AI risks with divergent emphasis: Businesses are focused on organizational implications, while consumers are concerned about societal topics like shared prosperity, inclusion, and AI's impact on jobs.

The CEO disconnect on AI ethics
While board members see AI ethics as a significant issue and area of corporate-oversight responsibility, CEOs view these issues as less important than either their C-level team or board members do.

Resolving the ethics dilemma
Most board members consider themselves unprepared to tackle AI ethics issues. Individual organizations are developing guidelines and implementation plans, but no organization can resolve this problem alone. The complexity and novelty of the issues require partnerships and alliances that act collaboratively.

—

Envisioning AI ethics

If you learned that a customer's credit application had been rejected with no discernible justification, would you condone the decision? If a doctor with no access to the latest medical literature recommended that a loved one undergo an invasive procedure, would you authorize the surgery? Surely not.

Critical areas of judgment – especially decisions that directly impact others' lives and well-being – are governed by standards of appropriate action. Humans live by communally governed ethical norms, enforced by laws, rules, societal pressures, and public discourse. While ethics may vary over time and across cultures, they have played a crucial role in guiding decisions since early human civilization. The Hippocratic Oath, for instance, finds its roots in ancient Greece and has been a mainstay of medical practice since the medieval era – from "first, do no harm" to patient confidentiality.[1]

In business, the topic of ethics is also not new. Traditionally, ethics are one of the much-talked-about but often-ignored elements of modern business culture. In the quest for financial success, sometimes corners are cut, advantages are taken, and long-term priorities – and even values – are sacrificed for short-term gains. In response, a compliance apparatus has been created inside companies. Organizations have worked to create "guardrails" and other reinforcement mechanisms to combat lapses, whether inadvertent or willful.

Arguably, ethical considerations have never been more critical than they are today. People around the world, from business executives and front-line employees to government ministers and individual citizens, have found themselves confronting critical decisions they wouldn't have imagined making in the past – decisions that could profoundly impact the lives of their fellow colleagues and citizens. And many have been forced to weigh seemingly impossible trade-offs between economic and health imperatives, guided only by their ethics, morals, and values.

81% of consumers say they became more concerned over the prior year with how companies use their data, and 75% percent are now less likely to trust organizations with their personal information.[2]

Four out of five directors say AI ethics issues are a board-level responsibility at least to a moderate extent, but barely half of CEOs view them as a CEO-level responsibility.

Well over half of all executives point to the CTO and CIO as primarily accountable for AI ethics.

Executives expect technology firms will greatly influence AI ethics, followed by governments and customers – with other companies last on the list.

Yet the existing system is ill-equipped for the challenges to come. In the last few years, the business environment has been rapidly changing, with a growing number of decisions facilitated by artificial intelligence (AI). Companies now use AI to help conduct talent screening, insurance claims processing, customer service, and a host of other important workflows. But the ethical parameters around AI remain vague and intangible, in some instances pushed aside as impediments to progress – without regard for possible short- and long-term ramifications.

AI ethics: The corporate landscape

Meanwhile, AI adoption is expected to continue growing rapidly. In fact, average spending on AI will likely more than double in the next three years, according to the results of a prior IBM Institute for Business Value survey of global executives.[3] With heightened AI use will come heightened risk, in areas ranging from data responsibility to inclusion and algorithmic accountability.

To gain a deeper understanding of executive views on the consequences and ethical considerations associated with AI's growing role in the enterprise, we sought input from 1,250 executives from a variety of industries and geographies. (For more information on the research, please see the "Insight: Study research and methodology" sidebar.) Our research suggests AI's importance to organizational strategy is likely to double in the next three years, underscoring the need to address the topic of ethics.

While almost all the executives surveyed say their own organizations are ethical, there is widespread concern that the application of AI could have serious repercussions if improperly unleashed. The unavoidable conclusion: Ethical considerations must be elevated in the dialogue about AI systems across the business landscape.

The level of cognitive understanding between humans and machines is inherently lower than it is between humans and other humans, yet the latter arena has been structured for centuries around ethics. Since AI relies on huge computing power, it can derive insight from massive amounts of data that would challenge human cognition. Relying only on traditional ethical approaches to decision making may be insufficient in addressing AI-powered decisions.

Indeed, more than half of the executives we surveyed tell us AI actually could improve their companies' ethical decisions. (Less than 10 percent were concerned about a negative impact.) A majority also say AI could be harnessed as a force for societal good, not just for good business. When one chief human resource officer was asked what came to mind when thinking about AI, the answer was "advancement in technology for betterment of human life," a sentiment echoed by others.

While xenophobia, misogyny, and other biases – even if unintentional – can be obscured behind human rationales, our study respondents suggest that AI can be designed with fairness, transparency, and even empathy. Detecting and correcting bias in AI – teaching technology to be more effective in relating to humans – may advance organizations' abilities to work together and achieve greater outcomes.

AI offers the opportunity to diagnose root causes of unintended results, essentially debugging biases. This could enable better understanding of past shortcomings and improvement in achieving social goals. But first, the right ethical frameworks have to be in place.

What matters most: Factors in addressing AI ethics

What is most important in ethically harnessing the power of AI? Who is responsible for helping ensure that ethics are integrated into AI, within corporations and outside? And how can society best use AI for good?

Those are the three overarching questions we set out to answer with our research and will address in this report. The dialogue about AI ethics so far has largely occurred among media, technology firms, consultancies, and academia (as well as some government bodies). There have been far fewer insights from companies that use AI. This study's aim is to give these overlooked constituencies – from banks to healthcare providers to retailers and beyond – an equivalent voice.

Insight: Study approach and methodology

In cooperation with Oxford Economics, the IBM Institute for Business Value surveyed 1,250 global executives in late 2018. Representing 20 industries and over 26 countries on 6 continents, survey participants included members of boards of directors, chief executive officers (CEOs), chief information officers (CIOs), chief technology officers (CTOs), chief data officers (CDOs), chief human resource officers (CHROs), chief risk officers (CROs), general counsels, and government policy officials.

According to more than half of the executives surveyed, AI actually could improve a company's ethical decisions.
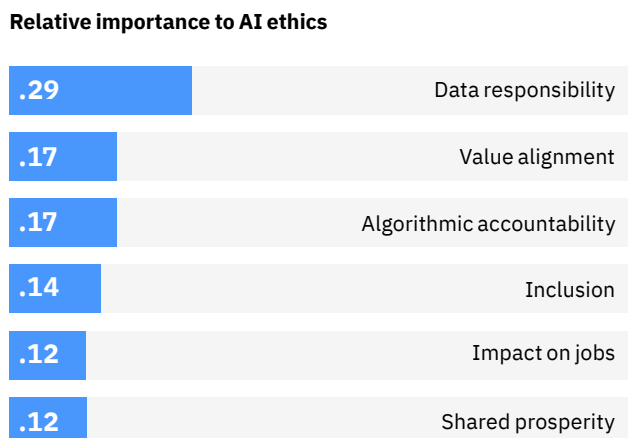
We confirmed that, unsurprisingly, AI has become a central and much-discussed technology. And nearly two-thirds of the executives surveyed also view AI *ethics* as an important business topic at least to a moderate extent. This even includes those in organizations that are not considering adopting AI at the moment. Among executives whose organizations are currently engaging with AI, the percentage is almost 90. And nearly all say they are formally considering ethics as part of their AI initiatives at least to a moderate extent.

As a part of our research, we also asked executives to rate the relative importance of a range of factors for developing ethical AI, including value alignment, algorithmic account-ability, and inclusion. Respondents were asked to make trade-off decisions regarding the importance of these factors, one to another, in a pairwise fashion.

We discovered that three main areas of ethical risk dominate the attention of organizations and their executives: data responsibility, value alignment, and algorithmic accountability (see Figure 1). The values in the figure below indicate the ratio of how important the individual factors are compared to one another.

—

Figure 1
Executives rank data responsibility among the most important factors in developing ethical AI

Relative importance to AI ethics

.29	Data responsibility
.17	Value alignment
.17	Algorithmic accountability
.14	Inclusion
.12	Impact on jobs
.12	Shared prosperity

Source: 2018 IBM Institute for Business Value Global AI Ethics Study. Q: Thinking of AI ethics, which of the following are relatively more important? N=1,250.

Data responsibility

The first risk area, by a wide margin, is data responsibility, including data ownership, storage, use, and sharing. Data responsibility is rated as twice as important as any other factor in AI ethics. This is undoubtedly influenced by the stream of examples of data breaches and misuse that have punctuated the news.

The customer implications are significant: According to a global consumer survey conducted by the IBM Institute for Business Value during a similar timeframe, 81 percent of consumers have become more concerned in the past year about how companies are using their data, with 75 percent saying they have become less likely to trust organizations with their personal information.[4] Along with this consumer concern, organizations face rising regulatory pressures, as evidenced by the Global Data Protection Regulation (GDPR) now in effect in the European Union (EU), as well as several rounds of U.S. congressional testimony on data privacy.[5]

The impact of data-related risk is already being felt in the AI realm: Forty percent of the executives we surveyed view various concerns about data trust, privacy, and transparency as a barrier to AI adoption. It may not be possible to realize the promises of AI without addressing the trust issue, as the power of AI depends almost entirely on the underlying data. Only half (54 percent) of the executives say they have a high degree of confidence in their business data. Increasing that confidence is critical to successfully adopting AI.

Value alignment and algorithmic accountability

The next two highest-ranked AI ethics risks cited by executives are almost tied at 17 percent, perhaps because they are related. Value alignment refers to an AI algorithm's ability to operate as expected, generating decisions that reflect the appropriate values, constraints, and procedures. Algorithmic accountability refers to identifying who is responsible for the output of an AI algorithm – which is also related to explaining how decisions were reached, both in innocuous situations and when debate emerges. Algorithmic accountability is viewed as an important priority by two-thirds of C-level executives with technical roles. This mirrors their expecta-tions that consumer demands for transparency and explainability will continue to grow.

To understand these two linked risks, consider several examples of how actions or decisions made by AI-enabled systems can produce imperfect results or unintended consequences. In one study assessing AI's accuracy in identifying cancerous lesions, researchers noted that the AI system tended to flag photographs of lesions with rulers or other visible indications of measurements. Dermatologists often use such tools if they suspect malignancy after an initial assessment. The AI agent "learned" that lesions with these tools in the picture were more likely to be cancerous – but without any understanding of why.[6]

Similar issues arose in assessing AI's ability to predict pneumonia from radiographs: The AI agent associated a higher probability of illness with specialist hospital locations that admitted sicker patients; it anchored on the correlation, without identifying or even looking for an underlying root cause.[7]

Another real-world and well-known example: Amazon shelved an AI-powered recruiting engine that appeared to be inclined against selecting women. Because industry hiring practices from the last decade had resulted in an employee pool dominated by men, the AI algorithms "learned" from historical data that conforming to successful hiring practices from the (male-dominated) past meant screening out resumes with activities such as "Women's Field Hockey Team Captain."[8]

Amazon is not alone in grappling with these types of AI challenges; analogous issues have arisen related to law enforcement, customer interactions, translation, and image interpretation.[9] The value misalignment problem – when AI misunderstands what it is supposed to do – is often closely followed by questions of who is responsible and accountable for the algorithm. Ensuring that AI does what it is meant to do, and can explain why it did so, is critical.

Insight: AI ethics definition

AI ethics is a multidisciplinary field of study in which the main goal is to understand how to optimize AI's beneficial impact while reducing risks and adverse outcomes for all stakeholders in a way that prioritizes human agency and well-being, as well as environmental flourishing. To this aim, AI ethics research focuses on how to design and build AI systems that are aware of the values and principles to be followed in the deployment scenarios. It also involves identifying, studying, and proposing technical and nontechnical solutions for ethics issues arising from the pervasive use of AI in life and society. Examples of such issues are data responsibility and privacy, fairness, inclusion, moral agency, value alignment, accountability, transparency, trust, and technology misuse.

What matters most: Who cares about AI ethics – and why

One of the surprising results of our collective research is how the degree of what matters varies among different geographic populations. Even more surprising is how the substance of what matters diverges between executives and consumers.

Regional perspectives

The fact that variance in opinions about AI ethics exists across regions is not, on the surface, unexpected. Indeed, given differing cultural norms, full congruity is unlikely and would be rather surprising. But the specific geographic results are thought provoking (see Figure 2). Confirmed by our study: In the mature, developed markets of North America, Japan, and Western Europe, roughly half of the executives say AI ethics are significantly important to their organizations.

Also confirmed: In developing markets like Latin America, Africa, and Southeast Asia – and rapidly advancing countries such as India and China – less than half of executives indicate that same importance. Broadly speaking, organizations in less-developed economies are less concerned about the ethical implications of the AI technologies that could assist their growth. That gap may be closing, though, especially when it comes to trust, transparency, and fairness among IT decision makers in those firms adopting AI – as evidenced by the results from a study IBM commissioned in late 2019.[10]

—

Figure 2

The importance organizations place on AI ethics varies across regions
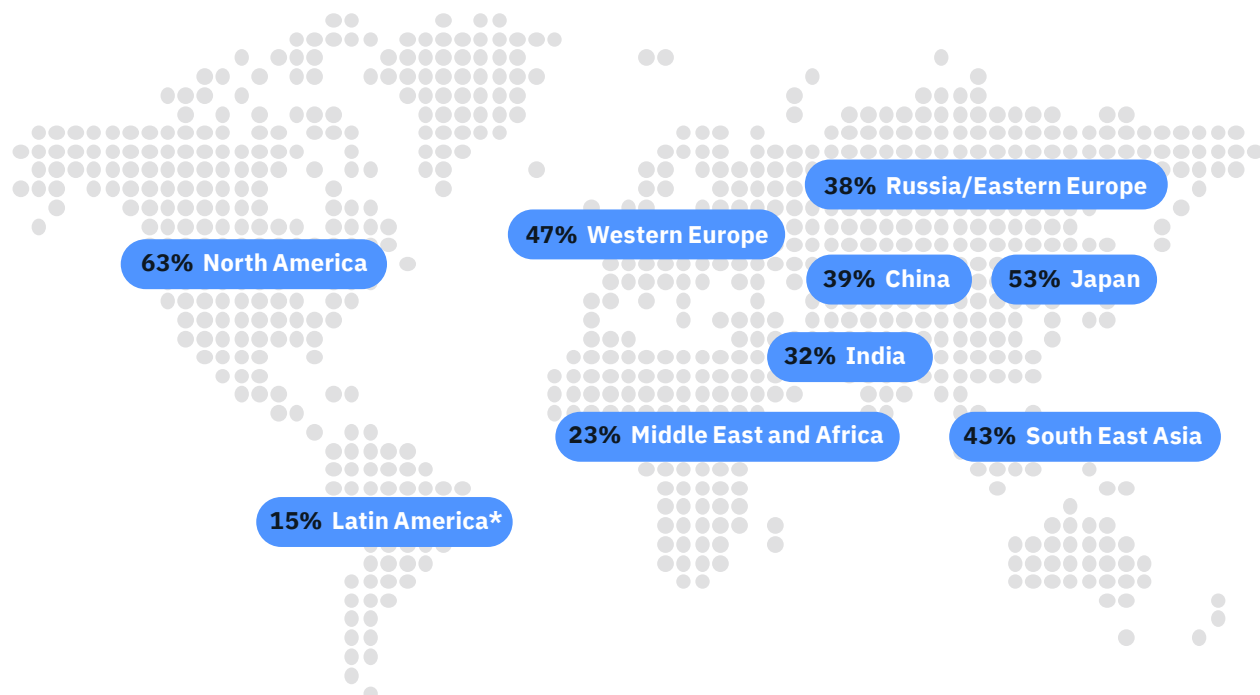
Count is less than 20.
Source: 2018 IBM Institute for Business Value Global AI Ethics Study. Q: Importance of AI ethics in your organization, N=1,247.

Only a little over one-third of CHROs say their organizations have an obligation to retrain or reskill workers impacted by AI technology.

Corporate versus citizen perspectives

A dichotomy also emerged when it came to parsing the AI ethics concerns of consumers and executives. Here, the differences hinge not on location, but the type of risk. In terms of the data responsibility risk, for instance, executives and consumers convey a similar high-priority concern. Yet their shared priorities do not persist across other areas.

Executive responses indicate they view AI's impact on societal well-being (i.e., inclusion, shared prosperity, and the effects on jobs) as substantially less important than factors directly impacting their organization (i.e., data responsibility, value alignment, and algorithmic accountability). In fact, shared prosperity and impact on jobs are identified by executives as the least important ethical considerations related to AI. These same areas, of course, are of central concern to the population in general.

The relative lack of importance executives place on societal well-being merits some consideration, particularly considering AI's impact on jobs and the workforce. There is no shortage of weighty studies concluding that AI will have a tremendous impact on workers and skills.[11] In 2019, we estimated that more than 120 million workers in the world's 12 largest economies may need to be retrained or reskilled in the next three years.[12] That number can only be expected to increase dramatically in the coming months and years.

According to our 2018 Country Survey, two-thirds of executives expect that advancements in AI and automation technology will require roles and skills that don't exist today.[13] A majority of global executives – 60 percent – estimate that up to 5 percent of their workforce will need to be reskilled or retrained in the next three years as a result of intelligent automation; more than a third – 38 percent – predict the percentage could be as high as 10.[14] Executives may be focused at the moment on the immediate organizational and stakeholder impacts of their AI deployments, but the effects on broader societal issues bear watching and vigilance as AI and attitudes toward it mature.

With so many jobs impacted by AI, the inward-focused emphasis on organizational impact expressed by the executives surveyed is short-sighted. Only a little over one-third of CHROs we surveyed say their organizations have an obligation to retrain or reskill workers impacted by AI technology. On a pragmatic level, investing in skills – including training employees to work with AI – will be critical to maintaining a quality workforce; on a societal level, deferring to other entities to resolve the dislocations that AI may produce could leave organizations exposed to backlash and distrust.

Influencing AI ethics: Who is responsible?

Given the risks related to AI ethics and the varied viewpoints geographically and between executives and consumers, the natural next question is how organizations can best position themselves to respond. To address this area, our survey explored who has primary responsibility for AI ethics within organizations and how much importance various leadership levels place on the topic. The results indicate significant misalignment.

Board of directors

When it comes to AI ethics in the private sector, a critical role can be played by the board of directors. According to the G20/OECD Principles of Corporate Governance: "The board has a key role in setting the ethical tone of a company, not only by its own actions, but also in appointing and overseeing key executives and consequently the management in general. High ethical standards are in the long-term interests of the company as a means to make it credible and trustworthy, not only in day-to-day operations but also with respect to longer-term commitments."[15]

The results of our study echo those principles: Four out of five directors tell us AI ethics should be a board-level issue, at least to a moderate extent. The downside: Only 45 percent of directors say they are fully prepared to tackle these issues, a gap with worrisome implications.

C-suite executives

What's more, board members will look to their executive teams to drive strategies that address AI ethics. However, our research indicates CEOs are somewhat out of step with their boards: Barely half of the CEOs we surveyed view AI ethics as a CEO-level issue at least to a moderate extent. (One CEO suggests, "These issues should be handled by a board committee with a focus on technology and data and the associated ethical principles.") What's more, CEOs are also out of step with their own executive teams, who share the responsibility for carrying out board directives: CEOs' average rating of the overall importance of AI ethics is well below that of their C-level reports.

This dual disconnect, between CEOs and boards and CEOs and their executives, is among the most disconcerting results we have seen in recent corporate studies. With top-down leadership critical to signaling the importance of any corporate-wide initiative, what will happen at companies with CEOs who are less motivated? Will boards have to force the issue, and how disruptive might that be?

There are several other uncomfortable results, in terms of operational roles. Well over half of all executives in our study point to the CTO and CIO as primarily accountable

for AI ethics in the organization. In other words, AI ethics are seen fundamentally as a technical responsibility. When asked to choose a single accountable executive, only 15 percent selected a nontechnical role.

The need to take ownership

Our research also suggests a hesitancy among executives to unequivocally claim AI ethics as a corporate issue. In fact, executives expect multiple entities outside of their organizations to greatly influence AI ethics (see Figure 3). They point to technology firms first, followed by governments and customers – with other companies last.

This abdication of responsibility to external players is corroborated by the results of regional roundtables on governance and inclusion conducted in cities across the globe by Harvard University's Berkman Klein Center for Internet & Society.[16] As Ryan Budish, assistant director for research at the Center, observed: While the sessions were a constructive step toward more active engagement between the public and private sectors on AI ethics, "We observed a similar dynamic of finger pointing with regard to responsibility."[17]

—

Figure 3
Executives expect multiple factors outside their own organizations will greatly impact AI ethics

Perceived impact on AI ethics in three years

Technology firms	84%
Governments	73%
Customers	67%
Other companies	50%

Source: 2018 IBM Institute for Business Value Global AI Ethics Study. Q: To what extent do you expect the following to influence AI ethics [in three years]? N=1,250.

Our survey results
suggest a need for more
board-level education
about AI ethics issues.

Inside the organization: Taking action

What steps might be considered in response to the proliferation of AI, given the corporate ambiguities and disconnects regarding AI ethics?

Learning from the past

Lessons can be learned from the rise of biotechnology in the 1970s. Back then, the new discipline of bioethics emerged to address the implications of new scientific innovations. Through the years, various US Presidential Commissions on bioethics have been created, with the support of industry.[18] Some recent work on the principles of AI ethics explicitly references tenets of bioethics.[19] As the moral philosopher and bioethicist Peter Singer recommends, "Experts who are respected in their fields need to be involved."[20] The involvement and mandate of independent experts could help sharpen the dialogue, avoid pitfalls, and enhance trust.[21]

One difference Singer highlights between AI and biotech is the "alarmist responses exacerbated by present-day media dynamics."[22] Businesses should help ensure that serious issues worthy of careful deliberation are given their due – and not relegated to debates on social media. Substantive discussions of ethics typically don't translate well into 140 (or even 280) characters.

Another difference Singer notes is the prominence of nonprofit organizations in biotech (such as hospitals).[23] This observation underscores the need to consider the implications of profit-motivated business models, which fuel much of today's tangible innovation in AI.

Getting the board on board

Our survey findings suggest the need for more board-level education about and engagement with AI ethics issues. The World Economic Forum's AI Board Toolkit, developed through collaboration with various public and private partners including IBM, is a start in this direction.[24]

Other organizations have undertaken similar efforts, some focused exclusively on AI ethics.[25] For example, a team at Princeton's University Center for Human Values developed case studies that explore AI ethics in depth.[26] These cases have been used by educational institutions and global companies in AI ethics and governance activities.[27] Another example is the case study compendium "AI, Labor, and the Economy" from the Partnership on AI, a consortium of about 90 partner organizations, including IBM.[28]

Institutionalizing AI ethics

Beyond the board level, AI ethics need to be embedded into existing corporate mechanisms, from the CEO's office and the C-suite down to the operational level. This includes business conduct guidelines, values statements, employee training, and ethics advisory boards. A general counsel for a UK-based company indicates that the single most important action an organization can take is "forming a team comprised of ethicists, software developers, data engineers, and legal experts," a view echoed almost verbatim by a board director in Japan and a chief risk officer in Canada.

Executives need to be wary of only paying lip service to AI ethics, or what Harvard University's Budish highlights as a growing concern in the corporate governance community: ethics washing.[29] In an article on the ethical principles of AI, Luciano Floridi, a professor of philosophy at the University of Oxford, and his co-author Tim Clement-Jones, stress the need to prioritize substance over perception, demonstrating transparency about the impact that ethics advisory boards, educational programs, and other tools and techniques have on real business decisions.[30]

Intent matters, but so do outcomes.

There is some debate about whether commercial entities need to do more than simply comply with existing regulatory, legal, and industry standards – to get ahead of the status quo. Companies with strategic objectives that align with high-minded leadership might have an easier time justifying the effort and investment, while more traditional companies focused squarely on near-term profit seeking might resist. Increasingly, though, all enterprises will depend on sharing and using data from customers and other partners, which makes building trust even more essential to creating and protecting shareholder – and stakeholder – value.

Monday morning corporate playbook

1 **Boards**
Ensure CEO and C-level team are fully aware of and engaged in AI ethics issues; monitor progress.

2 **CEOs**
Establish internal AI ethics board to provide governance, oversight, and recommendations; ensure responsibilities are clear to C-level team.

3 **CHROs**
Assess AI impact on skills and workforce; take ownership for outcomes.

4 **Technical executives**
Embed ethics governance and training in all AI initiatives.

5 **Risk/legal**
Ensure AI ethics is incorporated in mechanisms for institutionalizing values.

Outside the organization: Preparing for tomorrow

Rules relating to AI ethics are emerging inside organizations: More than half of the companies from our ethics study have adopted business conduct guidelines, values statements, employee training, or ethics advisory boards around AI. But relying on enterprises alone is unlikely to provide a complete solution.

Educating from the ground up

The pipeline into organizations, to begin, can certainly be strengthened. As one board member notes, "Ethics in AI can be encouraged by educating students at the university level followed by imparting proper knowledge, as well as guidance, to the current professionals working within different industries."

Business and law schools, computer science programs, and technical organizations like the American Association for Artificial Intelligence (AAAI) and the Institute of Electrical and Electronics Engineers (IEEE) have started to pilot ethics-related curricula, establish certifications, set standards, and create guides and toolkits.[31] A crowd-sourced list of AI and tech ethics university courses, started by Casey Fiesler from the University of Colorado Boulder, has more than 250 entries as of February 2020 and continues to grow.[32] Companies that employ graduates and members of these institutions can play a role in helping ensure ethics training is effective.

Setting guidelines and standards

Then there is government action. The overall high level of importance ascribed to AI ethics by executives and the emerging legislative interest suggest that regulatory standards will play a material role in the evolving AI future. When asked where they think these standards will be set, executives from our AI ethics study say they are anticipating formal guidelines at the national, supra-national, and even global levels – rather than at the local/regional levels or from professional organizations.

> "Ethics in AI can be encouraged by educating students at the university level followed by imparting proper knowledge... to the current professionals working within different industries."

Board member survey respondent

Almost a year after GDPR went into effect, the Ethics Guidelines for Trustworthy AI were published in April 2019. They are the culmination of work conducted by the independent High-Level Expert Group on AI appointed by the European Commission.[33] The group also published the Policy and Investment Recommendations for Trustworthy AI in June 2019, and further versions of such recommendations, specific to various sectors, are planned through 2020.[34]

While most major technology firms have issued their own guidelines, some have explicitly endorsed those from the European High-Level Expert Group. These guidelines define a human-centric "trustworthy" AI approach built around seven requirements: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability.[35]

Creating a unified approach

The AI regulatory environment will continue developing. The disruptive nature of AI combined with the accelerating pace of adoption will challenge the agility of many governing bodies. Yet tackling ethical issues is a function of society. By recognizing this and working with shared responsibility, businesses can better address the needs of affected citizens.

Even if guidelines vary across regions and professions, the design principles from the European Commission's approach can serve as a best practice: a) an independent group with multidisciplinary, multi-stakeholder representation, b) an agreed declaration of human rights relevant to the mandate, and c) a stated direction toward concrete executional recommendations.

Some efforts look at the current state of AI technology and focus on mitigating the risks while enhancing the benefits. Another approach is epitomized by the United Nations' AI for Good platform, which focuses on determining where society wants to go in terms of AI.

The platform's future vision is guided by the UN's 17 Sustainable Development Goals and the methods, tools, and technological innovations required to achieve them.[36]

We encourage adopting a hybrid approach to ensure a robust, complete, and holistic assessment of present and future implications.

Corporate education, professional standards, and even effective regulation are not enough. No isolated government, technology firm, corporation, professional organization, concerned citizen group, academic institution, or other entity can unilaterally achieve the needed aims. Varied stakeholders must act collaboratively.

These and other observations were apparent in the vigorous yet civil discussions at various world business and AI forums – including the latest AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society in early February 2020.[37] This dialogue needs to continue moving outside conference and academic environs and into the boardrooms, executive suites, IT labs, and even front-line operations of those companies implementing AI today. The common ground is our shared humanity: the societal rights we foster and moral responsibility we bear.

Fostering a sustainable future

Issues of ethics are rarely black and white. The seriousness of AI's tangible implications demands an equivalent level of seriousness in addressing AI ethics. There are substantive questions about the tradeoffs between individual privacy and business value, regulation and innovation, and transparency and competitive advantage. Those tradeoffs deserve to be debated in a thoughtful, civilized, and engaged manner without inflammatory rhetoric.

What's at stake may be no less crucial than a wholesale rethinking of the social contract.

Ethics issues in the context of AI are not just the domain of scholars and pundits. They matter to companies, customers, and citizens. Organizations that proactively address these issues and take meaningful action have an opportunity to shape their competitive future – and make AI more trustworthy and, hopefully, more trusted.

About the authors

Brian C. Goehring
goehring@us.ibm.com
linkedin.com/in/
brian-c-goehring-9b5a453/

Brian Goehring is an Associate Partner and the AI Lead for the IBM Institute for Business Value, where he brings over 20 years' experience in strategy consulting with senior-level clients across most industries and business functions. He received an A.B. in Philosophy from Princeton University with certificates in Cognitive Studies and German.

David Zaharchuk
david.zaharchuk@us.ibm.com
bit.ly/DaveZaharchuk
@DaveZaharchuk

Dave Zaharchuk is Research Director and Government Lead for the IBM Institute for Business Value. Dave is responsible for directing thought leadership research on a variety of issues related to both the public and private sectors.

Francesca Rossi, Ph.D.
Francesca.Rossi2@ibm.com
linkedin.com/in/
francesca-rossi-34b8b95/
@frossi_t

Francesca Rossi is an IBM Fellow at the T.J. Watson Research Center and the IBM AI Ethics Global Leader. Prior to joining IBM, she was a professor of computer science at the University of Padova, Italy.

The right partner for a changing world

At IBM, we collaborate with our clients, bringing together business insight, advanced research, and technology to give them a distinct advantage in today's rapidly changing environment.

IBM Institute for Business Value

The IBM Institute for Business Value, part of IBM Services, develops fact-based, strategic insights for senior business executives on critical public and private sector issues.

For more information

To learn more about this study or the IBM Institute for Business Value, please contact us at iibv@us.ibm.com. Follow @IBMIBV on Twitter, and, for a full catalog of our research or to subscribe to our monthly newsletter, visit: ibm.com/ibv.

Related reports

Brenna, Francesco, Giorgio Danesi, Glenn Finch, Brian Goehring, and Manish Goyal. "Shifting toward Enterprise-grade AI: Confronting skills and data challenges to realize value." IBM Institute for Business Value. September 2018. ibm.biz/enterprisegradeai

LaPrade, Annette, Janet Mertens, Tanya Moore, and Amy Wright. "The enterprise guide to closing the skills gap: Strategies for building and maintaining a skilled workforce." IBM Institute for Business Value. September 2019. ibm.co/closing-skills-gap

Mantas, Jesus. "Intelligent Approaches to AI." NACD Directorship magazine. November/December 2019. ibm.co/intelligent-approaches-ai

Notes and sources

1 "Greek Medicine." U.S. National Library of Medicine website, accessed December 3, 2019. https://www.nlm.nih.gov/hmd/greek/greek_oath.html

2 Unpublished data from the 2018 IBM Institute for Business Value Global Consumer Study. IBM Institute for Business Value.

3 Unpublished data from the IBM Institute for Business Value survey on AI/cognitive computing in collaboration with Oxford Economics. IBM Institute for Business Value. 2018.

4 Unpublished data from the 2018 IBM Institute for Business Value Global Consumer Study. IBM Institute for Business Value.

5 Jaffe, Justin, and Laura Hautala. "What the GDPR means for Facebook, the EU and you." CNET. May 25, 2018. https://www.cnet.com/how-to/what-gdpr-means-for-facebook-google-the-eu-us-and-you/

6 Patel, Neel V. "Why Doctors Aren't Afraid of Better, More Efficient AI Diagnosing Cancer." *The Daily Beast.* December 11, 2017. https://www.thedailybeast.com/why-doctors-arent-afraid-of-better-more-efficient-ai-diagnosing-cancer

7 Zech, John R, Marcus A. Badgeley, Manway Liu, Anthony B. Costa, Joseph J. Titano, and Eric Karl Oermann. "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study." PLOS Medicine. November 6, 2018. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002683

8 Dastin, Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters. October 9, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

9 Dodds, Laurence. "Chinese businesswoman accused of jaywalking after AI camera spots her face on an advert." *The Telegraph*. November 25, 2018. https://www.telegraph.co.uk/technology/2018/11/25/chinese-businesswoman-accused-jaywalking-ai-camera-spots-face/; Blier, Noah. "Stories of AI Failure and How to Avoid Similar AI Fails in 2019." Lexalytics blog. October 30, 2019. https://www.lexalytics.com/lexablog/stories-ai-failure-avoid-ai-fails-2019; Sonnad, Nikhil. "Google Translate's gender bias pairs 'he' with 'hardworking' and 'she' with lazy, and other examples." Quartz. November 29, 2017. https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/; Doctorow, Cory. "Two years later, Google solves 'racist algorithm' problem by purging 'gorilla' label from image classifier." Boing Boing. January 11, 2018. https://boingboing.net/2018/01/11/gorilla-chimp-monkey-unpersone.html

10 Survey commissioned by IBM in partnership with Morning Consult: "From Roadblock to Scale: The Global Sprint Toward AI." January 2020.

11 LaPrade, Annette, Janet Mertens, Tanya Moore, and Amy Wright. "The enterprise guide to closing the skills gap: Strategies for building and maintaining a skilled workforce." IBM Institute for Business Value. September 2019. ibm.co/closing-skills-gap

12 2018 IBM Institute for Business Value Global Country Survey. IBM Institute for Business Value; "Labor force, total by country." The World Bank. 2017; IBM Institute for Business Value analysis and calculations. 2019.

13 Unpublished data from the 2018 IBM Institute for Business Value Global Country Survey. IBM Institute for Business Value.

14 Ibid.

15 Principle VI:C. "G20/OECD Principles of Corporate Governance." Organisation for Economic Co-operation and Development. http://www.oecd.org/corporate/principles-corporate-governance.htm

16 Berkman Klein Center for Internet & Society at Harvard University website, accessed November 20, 2019. https://cyber.harvard.edu

17 Interview with Ryan Budish, Levin Kim, and Jenna Sherman. May 2019.

18 Gray, Bradford H. "Bioethics Commissions: What Can We Learn from Past Successes and Failures?" National Center for Biotechnology Information website, accessed November 20, 2019. https://www.ncbi.nlm.nih.gov/books/NBK231978/

19 Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. "AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." Minds and Machines: Volume 28, Issue 4. December 2018. https://link.springer.com/article/10.1007/s11023-018-9482-5

20 Interview with Peter Singer. April 2019.

21 Zimmermann, Annette, and Bendert Zevenbergen. "AI Ethics: Seven Traps." Freedom to Tinker. Princeton Center for Information Technology Policy. March 25, 2019. https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-traps/

22 Interview with Peter Singer. April 2019.

23 Ibid.

24 "Empowering AI Leadership." World Economic Forum's Shaping the Future of Technology Governance: Artificial Intelligence and Machine Learning platform. https://www.weforum.org/projects/ai-board-leadership-toolkit

25 Butterfield, Kay Firth, and Ana Isabel Rollan Galindo. "AI Toolkit for Boards of Directors." Australian Institute of Company Directors. September 26, 2018. https://aicd.companydirectors.com.au/membership/membership-update/ai-toolkit-boards-directors; Else, Shani R, and Francis G.X. Pileggi. "Corporate Directors Must Consider Impact of Artificial Intelligence for Effective Corporate Governance." Business Law Today. February 12, 2019. https://businesslawtoday.org/2019/02/corporate-directors-must-consider-impact-artificial-intelligence-effective-corporate-governance/; Bethke, Anna, and Kate Schneiderman. "AI Ethics Toolkits." Intel AI blog. February 19, 2019. https://www.intel.ai/ai-ethics-toolkits/#gs.cbuub4

26 "Princeton Dialogues on AI and Ethics Case Studies." Princeton University Center for Human Values and the Center for Information Technology Policy. https://aiethics.princeton.edu/case-studies/

27 Zevenbergen, Bendert. "Princeton Dialogues of AI and Ethics: Launching case studies." Freedom to tinker website (accessed December 13, 2019). May 21, 2018. https://freedom-to-tinker.com/2018/05/21/princeton-dialogues-of-ai-and-ethics-launching-case-studies/

28 "AI, Labor, and the Economy Case Study Compendium." Partnership on AI. https://www.partnershiponai.org/compendium-synthesis/

29 Interview with Ryan Budish, Levin Kim, and Jenna Sherman. May 2019.

30 Florid, Luciano, and Lord Tim Clement-Jones. "The five principles key to any ethical framework for AI." NS Tech. March 20, 2019. https://tech.newstatesman.com/policy/ai-ethics-framework

31 Zittrain, Jonathan, and Joi Ito. "The Ethics and Governance of Artificial Intelligence." MIT Media Lab. November 16, 2017. https://www.media.mit.edu/courses/the-ethics-and-governance-of-artificial-intelligence/; Jagadish, H.V. "Data Science Ethics." Michigan Online, University of Michigan. https://online.umich.edu/courses/data-science-ethics/ and https://www.wsj.com/articles/university-is-rolling-out-certificate-focused-on-ai-ethics-11558517400?mod=djemAIPro; Sullins, John. "Responsible Innovation in the Age of AI." IEEE Xplore. April 2018. https://ieeexplore.ieee.org/courses/details/EDP496; "The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)." IEEE Standards Association. https://standards.ieee.org/industry-connections/ecpais.html; "Ethical Considerations in Artificial Intelligence Courses." *AI Magazine*. July 1, 2017. https://aaai.org/ojs/index.php/aimagazine/article/view/2731; Cutler, Adam, Milena Pribić, Lawrence Humphrey, Francesca Rossi, Anna Sekaran, Jim Spohrer, and Ryan Caruthers. "Everyday Ethics for Artificial Intelligence." IBM Design for AI. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf; "Ethics in Action." IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. https://ethicsinaction.ieee.org/; "AI Fairness 360 Open Source Toolkit." IBM Research. http://aif360.mybluemix.net/?cm_mc_uid=46348957889315487882623&cm_mc_sid_50200000=71493781559770063924; "AI for Good Global Summit." International Telecommunication Union. https://aiforgood.itu.int/

32 Fiesler, Casey. "Tech Ethics Curricula: A Collection of Syllabi." Post on Medium.com, accessed November 15, 2019.

33 "High-Level Expert Group on Artificial Intelligence." European Commission. May 2, 2019. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

34 Ibid.

35 "Artificial Intelligence at Google: Our Principles." Google AI. https://ai.google/principles/; "Microsoft AI principles." Microsoft AI. https://www.microsoft.com/en-us/ai/our-approach-to-ai; "DeepMind Ethics & Society Principles." DeepMind. https://deepmind.com/applied/deepmind-ethics-society/principles/; "Trusted AI." IBM Research. https://www.research.ibm.com/artificial-intelligence/trusted-ai/; Rende, Andrea. "Europe's Quest For Ethics In Artificial Intelligence." *Forbes*. April 11, 2019. https://www.forbes.com/sites/washingtonbytes/2019/04/11/europes-quest-for-ethics-in-artificial-intelligence/#5137a7a57bf9; "Ethics Guidelines on Trustworthy AI." European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

36 "AI for Good Global Summit." International Telecommunication Union website for AI for Good Global Summit, accessed December 12, 2019. https://aiforgood.itu.int

37 Third AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. February 7-8, 2020. New York, NY.

About Research Insights

Research insights are fact-based strategic insights for business executives on critical public and private sector issues. They are based on findings from analysis of our own primary research studies. For more information, contact the IBM Institute for Business Value at iibv@us.ibm.com.