Artificial Intelligence, Social Responsibility, and the Roles of the University

How universities can influence socially responsible use of AI technology development and use.

By Nigel Bosch, Anita Say Chan, Jenny L. Davis, Rochelle Gutiérrez, Jingrui He, Karrie Karahalios, Sanmi Koyejo, Michael C. Loui, Ruby Mendenhall, Madelyn Rose Sanfilippo, Hanghang Tong, Lav R. Varshney, and Yang Wang

Technologies that use artificial intelligence (AI) have become ubiquitous. AI technologies have produced numerous economic and social benefits, such as rapidly and reliably assisting radiologists with accurate diagnostic interpretations of medical images. Many harms of AI have also been documented, such as racial biases in predictive models used in the criminal justice system, and gender discrimination in automated screening of job applications. Some AI technologies have exacerbated biases that disproportionately affect historically marginalized communities, such as LGBTQ populations and members of racial, ethnic, and religious minorities.[4] Generative AI technologies are now widely available, and the potential harms are substantial: although anyone can use ChatGPT to draft messages and DALL-E to create artwork, others can use these tools to quickly produce deceptive news stories with specious images—misinformation that can spread quickly through social media.

AI technologies deployed in industry today are far more powerful than the early AI technologies created in university laboratories. We ask: What roles can the university now play in the socially responsible development and use of AI technologies? While many industrial organizations and governments have published statements of principles for social responsibility with AI technologies, we go beyond statements of principles to recommendations for actions by universities, particularly those in the U.S.

Since the first colleges were established in America in the 17th and 18th centuries, the purposes and missions of colleges and universities have evolved. The original mission of education has expanded beyond a fixed curriculum for upper-class youth to a multitude of subjects for all social classes. In the 19th century, universities added missions of research and public service. In the 20th century, many universities adopted missions of community engagement and economic development—the latter after the Bayh-Dole Act of 1980 accelerated the commercialization of

technologies developed at universities. With a great diversity of institutions in the U.S., different universities place different emphases on these missions. Here, we focus on four questions connected with the university missions of education, research, community engagement, and public service. For an extended discussion of these questions, with additional references, please refer to our white paper.[1]

Education

How can universities effectively educate students, technical professionals, and the public to consider social responsibilities in the design and use of AI systems?
In colleges and universities, issues in AI and social responsibility are currently covered in courses on computing ethics and in modules in technical courses.[5] These courses are sometimes taught by multidisciplinary teams, with members from computing, humanities, arts, data sciences, and social sciences. One example of multidisciplinary collaboration is the Social and Ethical Responsibilities of Computing (SERC) initiative at the Massachusetts Institute of Technology; case studies developed by SERC are freely available online. Since multidisciplinary instructional collaborations are not always valued by university reward structures, we recommend strategies that enable advocacy for the value of these collaborations, such as forming instructional teams that include a senior faculty member who can ensure junior colleagues receive credit toward promotion.

In disciplinary courses in computer science and engineering, students learn fundamental technical knowledge for developing AI technologies—the algorithms for machine learning and the mathematics of pattern recognition. To promote social responsibility, these courses should include instruction in techniques such as value-sensitive design that can reduce social biases, while recognizing the pitfalls of purely technical solutions.[8] Students should be encouraged to minimize the environmental impact of energy-intensive computations both in constructing AI models and in answering queries with these models.

Besides formal courses, universities should promote social responsibility in the use of AI technologies through public lecture series and existing outreach efforts common at many universities that include, for example, libraries, museums, lifelong learning programs, and other community spaces. Like the SERC initiative, instructional materials in these efforts should be inclusive (for example, to people with disabilities) and freely available online (for example, through the Online Ethics Center for Engineering and Science at the University of Virginia).

To date, there has been little empirical research on computing ethics education. In particular, there is currently no consensus about learning outcomes. We recommend education researchers undertake studies about education in AI and social responsibility: to define what learning outcomes could constitute AI literacy, and to determine what teaching methods are effective in achieving those outcomes.

Research

How can university and industry researchers collaborate on AI technologies in a socially responsible way? Many AI technologies are based on the application of machine learning algorithms to large datasets of data collected from individuals by e-commerce and social media firms. Even when the data are provided anonymously to researchers, individuals can sometimes be reidentified. When individuals are identifiable, university researchers have both an ethical obligation to protect their privacy and a legal obligation to comply with regulations on human subjects research. In the U.S., federally funded institutions must adhere to the Federal Policy for the Protection of Human Subjects, and research projects require oversight by an institutional review board (IRB). Other countries have equivalent provisions, with oversight by ethics committees. By contrast, industrial firms seldom have IRBs, with a notable exception of the Ethics Review Program at Microsoft Research. IRBs generally require the informed consent of the individuals whose data are used for research. In commercial datasets, however, the individuals are rarely aware of all research purposes to which their data could be applied. Even if they had technically given consent when registering on a commercial website, they were not fully informed about these purposes. University researchers should work with industry researchers to create datasets for clearly defined research purposes, following the ethical guidelines published by the Association of Internet Researchers. When appropriate, human subjects oversight should be provided.

Inherent biases in datasets can affect the quality of research that uses the datasets. For example, the ImageNet dataset contains more than 14 million images, which were labeled by 30,000 workers on Amazon's Mechanical Turk platform. After ImageNet was used in more than 300 research papers, researchers discovered social biases: images of individuals with lighter skin tones had more pleasant labels.[10] University and industry researchers should together develop auditing processes to identify biases in datasets and algorithms.

In the past, large collections of data were maintained primarily by government and academic organizations such as the U.S. Social Security Administration and the Inter-University Consortium for Political and Social Research based at the University of Michigan, with the purpose of serving the public interest. By contrast, today, data are collected and owned by business firms to serve private interests, though open source resources are emerging too. When industry and university researchers collaborate in AI research using proprietary datasets, the researchers need to negotiate, through their institutions' lawyers, who can access the data, what data can be accessed, what purposes would allow data access, and how the need for transparency in research publications can be reconciled with the need for confidentiality of proprietary information. University and industry researchers should collaborate to create equitable data access policies that balance public and private interests. Social Science One provides a model for these collaborations.

Community Collaborations

How can universities better collaborate with external organizations and local communities to address questions of bias and discrimination in AI technologies? In both industry and the university, AI technologies are often presented as one-size-fits-all solutions to problems in society.[3] These problems are defined and these solutions are

developed by entrepreneurs and technologists who are overwhelmingly white and male, from urban and middle-class backgrounds: the process of technology development systematically excludes marginalized populations such as women of color.[2]

Although popular "innovation frameworks" ignore marginalized communities, these communities can be sources of knowledge and wisdom in the design of AI technologies, centering care and reparation, to reduce bias and discrimination. Here, we describe three examples. The Our Data Bodies project comprises activists in marginalized communities in three cities in the U.S., who investigate how digital data about these communities are collected by corporations and local governments. The activists examine how these data systems inequitably affect decisions about housing access, public assistance, and community development. The Data for Black Lives movement brings together scientists, technologists, activists, and community organizers in meetings and conferences. They share research on how data are used as a tool of oppression of Black people, perpetuating inequality and injustice. They advocate for reducing discriminatory uses of data and for increasing civic engagement. The Global Indigenous Data Alliance aims to advance self-determination of Indigenous peoples around the world. The Alliance advocates against the expropriation and misuse of Indigenous data and works for uses of these data that benefit Indigenous peoples. The Alliance has developed a statement of data rights for Indigenous peoples.

Consistent with the mission of community engagement, universities can support and showcase the work of community organizations through ongoing partnerships. In particular, universities should recognize and value the scholarly work of faculty members who build relationships with community organizations and engage in the joint development of knowledge to reduce social biases in the design of AI technologies.

Governance

How can universities contribute to the governance of AI technologies? To limit the potential harms of technologies, social mechanisms are created, such as government regulations, technical standards, and institutional structures. At present, AI governance consists primarily of fragmentary regulations that respond to industry failures and that may reflect the industry-specific interests of the most powerful actors.[6] National governments and multilateral forums are, however, moving quickly on regulatory regimes. AI governance has been most effective when coordination occurs between stakeholders,[9] as with the Partnership on AI, and across systems or domains, as with contextually flexible frameworks like the NIST AI Risk Management Framework.[7] The coordination function can be performed by the university as part of its public service mission, because universities are networked across policymakers, governments, communities, media, and industry. Further, universities can be trustworthy partners because they are relatively independent from political influence and business interests. At the University of Chicago's Crown Family School of Social Work, Policy, and Practice, for example, the Office of Community Partnership and Impact brings together academic experts, government policymakers, and community organizers to address social issues such as reducing poverty in the city of Chicago. The Office is supported by the School's existing funds and by external grants for individual projects.

Individual academics frequently serve as external experts in the development of government policies and regulations. Besides advising on policies, academics play a key role in auditing processes, as consultants to regulatory agencies. Universities should recognize the importance of these scholarly forms of public service in promotion and tenure.

While individual academics can serve as independent experts in developing policies and in auditing technologies, universities can contribute to AI governance through institutional activities. As indicated in the "Community Collaborations" section, universities can collaborate institutionally with community organizations, who can identify the social impacts of AI technologies beyond the privileged viewpoints of industry and universities. To amplify these community voices, the university can provide a platform for responsive governance and participatory decision making, building on its role as a knowledge commons. Responsive governance is an alternative to technocratic governance, in which policies and standards reflect only the viewpoints of technical experts, not the perspectives of affected individuals. Responsive governance can ensure that in the governance of AI technologies, the status quo is not merely reproduced, but rather, those who have been historically overlooked or harmed have a say in what is appropriate. In short, universities should use the prestige of their institutional platforms to ensure those marginalized voices are heard.

Conclusion

From healthcare to policing, the rapid development and deployment of AI technologies have brought both social benefits and unintended harms, with disproportionate harms to marginalized communities. To promote the socially responsible development and use of AI technologies, universities should collaborate with industry, government, and community organizations in education, research, outreach, and public service activities. These activities should include teaching multidisciplinary courses on AI and social responsibility, both on campus and for the general public, and building networks with industry practitioners, government policymakers, and community partners to produce AI technologies and governance mechanisms that are responsive to community needs, rather than driven solely by business interests. Universities should ensure these activities are recognized as valuable forms of scholarship. By increasing engagement with external stakeholders, universities can contribute to social responsibility in the development and application of AI technologies.

References

1. Bosch, N. et al. *Artificial Intelligence and Social Responsibility: The Roles of the University*. University of Illinois, 2022; https://www.ideals.illinois.edu/items/125457

2. Brown, N. et al. Mechanized margin to digitized center: Black feminism's contributions to combatting erasure within the digital humanities. *Intern. J. of Humanities and Arts Computing 10*, 1 (Jan. 2016).

3. Chan, A.S. *Networking Peripheries: Technological Futures and the Myth of Digital Universalism*. MIT Press, Cambridge, MA, 2014.

4. Chun, W. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. MIT Press, Cambridge, MA, 2021.

5. Grosz, B.J. et al. Embedded EthiCS: Integrating ethics across CS education. *Commun. ACM 62*, 8 (Aug. 2019); 10.1145/3330794

6. Jung, M. and Sanfilippo, M.R. Mapping geographical biases of AI principles. *Poster Presentation, iConference 2022*; https://hdl.handle.net/2142/113756

7. National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework* (AI RMF 1.0) (Jan. 2023); 10.6028/NIST.AI.100-1

8. Selbst, A.D. et al. Fairness and abstraction in sociotechnical systems. *FAT* '19: Proceedings of the Conf. on Fairness, Accountability, and Transparency*. ACM, NY (Jan. 2019); 10.1145/3287560.3287598

9. Varshney, L.R., Keskar, N.S., and Socher, R. Pretrained AI models: performativity, mobility, and change. *arXiv:1909.03290* (2019).

10. Wiggers, W. Researchers show that computer vision algorithms pretrained on ImageNet exhibit multiple, distressing biases. *VentureBeat* (Nov. 3, 2020); https://bit.ly/4ey7lkl

About the Authors

Nigel Bosch (pnb@illinois.edu) is an assistant professor in the School of Information Sciences and Department of Educational Psychology at the University of Illinois Urbana-Champaign, IL, USA.

Anita Say Chan (achan@illinois.edu) is an associate professor in the School of Information Sciences and College of Media at the University of Illinois Urbana-Champaign, IL, USA.

Jenny L. Davis (loksi@illinois.edu) is an associate professor of American Indian Studies and Anthropology at the University of Illinois Urbana-Champaign, IL, USA.

Rochelle Gutiérrez (rg1@illinois.edu) is a professor in the Department of Curriculum and Instruction at the University of Illinois Urbana-Champaign, IL, USA.

Jingrui He (jingrui@illinois.edu) is a professor in the School of Information Sciences at the University of Illinois Urbana-Champaign, IL, USA.

Karrie Karahalios (kkarahal@illinois.edu) is a professor in the Department of Computer Science at the University of Illinois Urbana-Champaign, IL, USA.

Sanmi Koyejo (sanmi@illinois.edu) is an adjunct professor in the Department of Computer Science at the University of Illinois Urbana-Champaign, IL, and a professor in the Department of Computer Science at Stanford University, Stanford, CA, USA.

Michael C. Loui (loui@illinois.edu) is an emeritus professor of electrical and computer engineering and a University Distinguished Teacher-Scholar at the University of Illinois at Urbana-Champaign, IL, USA. He is the contact author for this Opinion column.

Ruby Mendenhall (rubymen@illinois.edu) is a professor of Sociology, African American Studies, and Carle Illinois College of Medicine at the University of Illinois Urbana-Champaign, IL, USA.

Madelyn Rose Sanfilippo (madelyns@illinois.edu) is an assistant professor in the School of Information Sciences at University of Illinois Urbana-Champaign, IL, USA.

Hanghang Tong (htong@illinois.edu) is an associate professor in the Department of Computer Science at the University of Illinois Urbana-Champaign, IL, USA.

Lav R. Varshney (varshney@illinois.edu) is an associate professor in the Department of Electrical and Computer Engineering at the University of Illinois Urbana-Champaign, IL, USA.

Yang Wang (yvw@illinois.edu) is an associate professor in the School of Information Sciences at the University of Illinois Urbana-Champaign, IL, USA.

All authors contributed equally to the development of this Opinion column.

Submit an Article to CACM

CACM welcomes unsolicited submissions on topics of relevance and value to the computing community.

Join the Discussion (0)