4. AI developers, associations, and the academic community

Chapter 3 discussed how private technology companies, influenced largely by market imperatives, play an important role in AI governance as actors who are subject to that governance but are also participating in its creation. However, other actors also impact international governance, albeit in less direct ways. This chapter explores how programmers, professional associations and academic communities also participate to some extent in that governance.

ARTIFICIAL INTELLIGENCE DEVELOPERS

Estimates and occupational categories vary, but according to one survey, in 2021 there were about 26.9 million software developers in the world.[1] 4.3 million developers were in the United States and 6 million were in Europe. Excluding China, there were 6.5 million developers in Asia, with 5.8 million in India.[2] India is expected to overtake the United States in the number of developers in the next few years.[3] One article reports that large Chinese technology firms employ about 7 million programmers, software developers, data scientists and AI trainers.[4] Worldwide, developers are overwhelmingly male, in one survey making up 91.67 percent of developers, while women made up only 5.31 percent, and 1.42 percent of programmers were non-binary,

---

1 *Software Development Statistics*, TRUELIST, Dec. 20, 2021, https://truelist.co/blog/software-development-statistics/.

2 *How Many Programmers Are There in the World and in the US?* QUBIT LABS, Oct. 19, 2021, https://qubit-labs.com/how-many-programmers-in-the-world/.

3 *Worldwide Professional Developer Population of 24 Million Projected to Grow amid Shifting Geographical Concentrations*, Evans Data Corporation, Press Release, May 21, 2019, https://evansdata.com/press/viewRelease.php?pressID=278.

4 Shen Lu, *Your Favorite Startup Might Have its Engineers in China: But Few Founders Want to Talk about It, Fearful of Becoming "The Next TikTok"*, PROTOCOL, Apr. 2, 2021, https://www.protocol.com/china/many-startups-chinese-engineers. Another report refers to Lu's report, but gives a smaller number of 500,000 "tech professionals" in China, presumably referring to computer programmers. Andrew Burak, *Is Software Development Outsourcing to China on the Decline?*, Relevant, https://relevant.software/blog/outsourcing-to-china/#5_China.

genderqueer, or gender non-conforming.[5] That women make up such a small percentage of programmers is ironic, since the first programmers were women.[6] This percentage is slowly improving, at least with regard to artificial intelligence. Programmers are relatively young, with the average age between 25 and 35 years.[7]

Of course, not all software developers work in artificial intelligence. Ian Hogarth estimates that "there are perhaps 700 people in the world who can contribute to the leading edge of AI research, perhaps 70,000 who can understand their work and participate actively in commercialising it."[8] Other estimates range from 22,000 to 300,000 people who work in AI.[9] A significant barrier is the knowledge required to enter the field. An introductory textbook argues that only a high school mathematics education and some experience with programming is needed to start developing AI models.[10] Most would argue, however, that work in artificial intelligence requires extensive training. An industry report states that "AI demands advanced competencies in mathematics, statistics and programming. AI developers are seven times more likely to have a Doctoral degree than other developers."[11] AI developers are increasingly required to have knowledge about specific domains of application, engineering experience, and commercial experience.[12]

The Chinese government has estimated that as of 2019 there was a shortage of 5 million workers trained in AI and that, without government intervention, that number would exceed 10 million by 2025.[13] The 2021 AI Index Report cites results from an annual survey taken by the Computing Research Association

---

5 Lionel Sujay Vailshery, *Software Developer Gender Distribution Worldwide as of 2021*, STATISTICA, Feb. 23, 2022, https://www.statista.com/statistics/1126823/worldwide-developer-gender/.

6 NATHAN L. ENSMENGER, THE COMPUTER BOYS TAKE OVER: COMPUTERS, PROGRAMMERS, AND THE POLITICS OF TECHNICAL EXPERTISE 6 (2010) (discussing the 'ENIAC girls' and other women in the early days of computing).

7 Truelist, *supra* note 1.

8 Ian Hogarth, *AI Nationalism*, June 13, 2018, https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism.

9 MMC Ventures, The State of AI 2019: Divergence, https://www.stateofai2019.com/.

10 JEREMY HOWARD AND SYLVAIN GUGGER, DEEP LEARNING FOR CODERS WITH FASTAI AND PYTORCH: AI APPLICATIONS WITHOUT A PHD 13 (2020).

11 The State of AI 2019, *supra* note 9.

12 *Id.*

13 Institute of China Science, Technology and Education Policy at Zhejiang University and Baidu, China Artificial Intelligence Talent Training Report 21 (Jan. 2022) (Etcetera Language Group, Inc. trans., 2022), https://cset.georgetown.edu/publication/china-artificial-intelligence-talent-training-report/ [hereafter China AI Talent Training Report].

of trends in the United States and Canada that in 2019, 28,000 undergraduates completed computer science degrees, a three-fold increase from 2010.[14] It was unclear what percentage of these graduates took positions in the AI field. During the same period, most PhD recipients specialized in AI-relevant research. The increase in their numbers, however, was not as dramatic.[15] In 2010–11, computer science programs who responded to the survey reported that they had conferred 1,782 doctoral degrees.[16] In 2020–21, 140 computer science departments reported conferring 1,893 PhDs.[17] Compensation reflects the demand for these workers. According to a website where applicants for such positions share details about the offers they have received, the median annual compensation proposed was US$322,500, including base salary, equity grants, target bonuses, and signing bonuses. The offers came from companies such as Facebook, Google, Pinterest, Sales Force, Amazon, DeepMind, and Microsoft.[18]

Any generalizations must be taken with a grain of salt, but a few tentative observations can be made about these developers. One is that irrespective of the numbers, the role of programmer can be crucial for designing beneficial AI systems. This explains why programmers are the targets for ethical training in the academy and on the job site. Another is that an ethos has come to be associated with at least a significant number of programmers that has resulted in norms for artificial intelligence development and programmers as individuals that contrast at times with those of the technology companies that often employ these programmers. The relatively small pool of AI developers explains the close relationship between private firms and universities in artificial intelligence development, a relationship that builds from a larger trend in the academy to enter the market. This also impacts the kinds of norms that

---

[14] Daniel Zhang et al., Artificial Intelligence Report 2021, Stanford University Human-Centered Artificial Intelligence, at 114 [hereinafter 2021 AI Index Report].

[15] Stuart Zweben and Betsy Bizot, 2011 Taulbee Survey: Continued Increase in Undergraduate CS Degree Production; Slight Rise in Doctoral Production 4 (Apr. 4, 2012), https://cra.org/wp-content/uploads/2015/01/CRA_Taulbee_2010-2011_Results.pdf [herinafter 2011 Taulbee Survey], at 23 (63 PhD recipients went to academia; 64 went to industry); Stuart Zweben and Betsy Bizot, 2021 Taulbee Survey: CS Enrollment Grows at all Degree Levels, With Increased Gender Diversity, Computing Research Association 3 (May 2022), https://cra.org/wp-content/uploads/2022/05/2021-Taulbee-Survey.pdf [hereinafter 2022 Taulbee Survey], at 6 (84 PhD recipients went to academia; 201 went to industry).

[16] 2011 Taulbee Survey, *supra* note 15.

[17] 2022 Taulbee Survey, *supra* note 15.

[18] AI Payrades: Statistics of industry job offers in Artificial Intelligence (AI), https://aipaygrad.es/. The highest reported offer was from Facebook for a Research Scientist, with overall annual compensation of US $1,175,000. *Id.*

might arise from this relationship. Third, the limited supply of AI developers compared with the pool of general programmers means that at least for now these individuals might have some influence over norm development. Fourth, even though they might not be truly representative of the developer community, professional AI organizations for programmers do have some influence over policy. Finally, the small pool of AI developers is leading to the development of general-purpose AI software and packaged AI services so that people without specialized knowledge can have access to the technology.

Programmers and Developers as Ethical Actors

As discussed in Chapters 2 and 3, ethics has emerged as the principal framework for AI governance. Commentators have thus argued that it is crucial for AI researchers and developers to be educated in ethics and that a culture of responsibility be inculcated among them,[19] in particular because the technical choices that programmers and designers make in designing models and in training them can have significant impacts. Jeremy Howard and Sylvian Gugger make the case to programmers by listing some of the design choices they must make.

> Now, as you are collecting your data and developing your model, you are making lots of decisions. What level of aggregation will you store your data at? What loss function should you use? What validation and training sets should you use? Should you focus on simplicity of implementation, speed of inference, or accuracy of the model? How will your model handle out-of-domain data items? Can it be fine-tuned, or must it be retrained from scratch over time?[20]

Howard and Gugger continue by arguing that programmers play a unique role among stakeholders in artificial intelligence applications because of their knowledge of artificial intelligence techniques.

> These [design and training questions] are not just algorithm questions. They are data product design questions. But the product managers, executives, judges, journalists, doctors—whoever ends up developing and using the system of which your model is a part—will not be well-placed to understand the decisions you made, let alone change them.[21]

---

[19] Miles Brundage et al., The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Feb. 2018, at 56, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.
[20] Howard and Gugger, *supra* note 10, at 179.
[21] *Id.*

The authors concede that programmers alone cannot prevent harms caused by AI applications because other influences are at play. They write, "[i]ndividual behavior change can't address misaligned profit incentives, externalities…, or systematic failures."[22] But Howard and Sylvian urge that programmers still play an important role in preventing harms, because external norms such as law cannot cover all the issues that arise in AI design. "[The law will never cover all edge cases, and it is important that individual software developers and data scientists are equipped to make ethical decisions in practice."[23]

Recall from Chapter 2 that education is supposed to serve as a lever for compliance with soft law programs. In a survey of technology employees, however, only 28 percent of respondents were aware of their employer's vision, mission, and values, 15 percent less than the responses other kinds of employees.[24] If this result is representative of programmers, it becomes even more important that programmers as individuals be trained to think about the wider implications of their work. However, the strengths and weaknesses of only an ethical approach to governance surface in other surveys. In a 2018 poll of 100,000 programmers, nearly 80 percent of respondents agreed that developers have an obligation to consider the ethical implications of their code.[25] Most respondents (58.5 percent) said that they would not write code for an unethical purpose, while a significant number (36.6 percent) said that it would depend on what the unethical purpose was.[26] The latter result suggests that even those who would code for unethical purposes would at least consider the ethical implications of their work as they decided whether to participate in the project. However, programmers have a different attitude about who has ultimate responsibility for their programs: 57.5 percent felt that upper management at a company or organization was ultimately responsible for code that accomplishes something unethical, 22.8 percent believed the person who came up with the idea was ultimately responsible, and only 19.7 percent placed ultimate responsibility with the programmer.[27]

Regarding attitudes towards artificial intelligence itself, another survey indicated that most developers are positive about the future of artificial intel-

22 *Id.*

23 *Id.*

24 TINYPulse, The State of Employee Engagement in Tech: 4 Big Bad Trends That Are Hitting This Workforce Hard, https://www.tinypulse.com/resources/the-state-of-employee-engagement-in-tech.

25 Stack Overflow, Developer Survey Results 2018, Ethics, https://insights.stackoverflow.com/survey/2018#ethics.

26 *Id.*

27 *Id.*

ligence.[28] They believe that developers (47.8 percent) and governments or other regulatory bodies (27.9 percent) should be responsible for addressing issues raised by AI. Between 20 and 29 percent of respondents were concerned about algorithms making important decisions, AI surpassing human intelligence, fairness in algorithmic versus human decision-making, and increasing automation of jobs.[29] These results suggest that these issues, while salient, are not of overriding concern to the programming community. Interestingly, 41 percent felt that job automation was one of the exciting prospects for artificial intelligence.[30]

If these surveys accurately reflect the attitudes of most programmers, they suggest several things about ethical governance at this level. One is that regardless of the source of their values, most programmers do consider the ethical implications of their work. This is helpful with respect to self-governance, particularly when this is a group that has the technical expertise that other groups do not. Such an attitude might suggest that programmers would be open to further education in ethics and other normative training once they leave school. At the same time, the fact that programmers do not feel that they are ultimately responsible for code that accomplishes something unethical and that they believe ultimate responsibility lies with their superiors is equally telling. In part, this attitude probably reflects the reality of being members of hierarchical organizations and institutions in which individual developers ultimately have little power. This view is also consistent with the aversion most people have towards being held responsible for an outcome, particularly if responsibility leads to negative consequences.

Programmer Influence Within Organizations

Nathan Ensmenger traces the history of computer programming and programmers and argues that programmers have at times challenged other power centers within an organization.[31] According to Ensmenger, since the early days of computing, programmers gained power as intermediaries between the computer and the domains in which computer programs were applied. In the 1950s, as computer programming entered the business world, and as computer applications began to be used in business management and control, programmers

---

[28] Stack Overflow Talent, Global Developer Hiring Landscape 2018, Artificial Intelligence, https://oliver-dev.s3.amazonaws.com/2018/07/11/13/49/09/592/Global%20Developer%20Hiring%20Landscape%202018.pdf.

[29] *Id.*

[30] *Id.*

[31] Ensmenger, *supra* note 6.

came to be seen as threats to established corporate hierarchies.[32] Ensmenger argues that many developments in computer programming, including efforts to professionalize developers, were aimed at bringing computer programmers under the control of these corporate structures.[33]

The issue here is the degree to which programmers influence the organizations in which they work, particularly when it comes to establishing AI norms. There is of course the more general phenomenon of employee participation in organizational management.[34] Conceptions of such participation vary, and organizational purposes and cultures differ.[35] So far, evidence of employee participation in technology companies has been largely anecdotal. Sometimes employee programmers have influenced decisions about AI research and development. One of the most well-known examples is the protests by Google employees and other groups after Google agreed to participate in Project Maven, a program spearheaded by the US Department of Defense. The project involved using AI techniques to help process and analyze in real time the vast amount of visual data collected by the military. Some 3,100 employees signed a petition urging Google to end the agreement, which Google eventually did.[36] The company disavowed designing and deploying AI for autonomous weapons, among other applications.[37] Later, however, Google executives stated it was possible to contract with the Defense Department without violating those principles.[38] Similarly, a group of anonymous Microsoft workers called Microsoft Workers 4 Good circulated a petition to protest a contract to develop augmented-reality technology for the military, but this was rejected.[39]

---

32 *Id.*, at 16, 21–22.

33 *Id.*, at 13–14, 17–18, 22–24.

34 For an introduction to the subject, see The Oxford Handbook of Participation in Organizations (Adrian Wilkinson et al., eds., 2010).

35 Adrian Wilkinson et al., *Conceptualizing Employee Participation in Organizations*, *in* The Oxford Handbook of Participation in Organizations, *supra* note 34, at 3, 4–6.

36 Cade Metz, Genius Makers: The Mavericks Who Brought AI to Google, Facebook, and the World 240–50 (2021).

37 "[W]e will not design or deploy AI in the following application areas: … Weapons and other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people." Google AI, Artificial Intelligence at Google: Our Principles, https://ai.google/principles/.

38 Kate Conger and Daisuke Wakabayashi, *Google Executives Tell Employees It Can Compete For Pentagon Contracts without Violating its Principles*, N.Y. Times, Nov. 17, 2021, B.3.

39 Klint Finley, *Microsoft CEO Defends Army Contract for Augmented Reality*, Wired, Feb. 25, 2019, https://www.wired.com/story/microsoft-ceo-defends-army-contract-augmented-reality/.

Recall that large technology companies have established ethics committees and review boards to vet the ethical impacts of their products and services. Moshe Vardi, however, sees a conflict between ethics and business models that require "the commodification of personal data with the core purpose of profit-making." He cites in this regard the departure of Tinnet Gebru, former technical co-lead of Google's Ethical Artificial Intelligence Team, who was reported to have left the position after being refused permission to make public a risk assessment of large language models being developed by the company. This departure was followed by the firing of another member of the team.[40] Perhaps Vardi is correct that there is fundamental contradiction between profit-maximization and ethical principles, or that profit-maximization skews how companies interpret and apply the ethical norms for AI they have adopted. At a minimum, the departure of Gebru and the experience of Google and Microsoft employees and defense contracts illustrate the complexity of power dynamics within large organizations as they make ethical determinations.

PROFESSIONAL ASSOCIATIONS

Several professional associations are involved in creating artificial intelligence norms. Before discussing them more specifically, it is helpful to have a sense of the history of professionalization of computer programmers, as well as some understanding of the functions of professional associations.

Computer Programmers as Professionals

In his history of computers and computer programmers, Ensmenger argues that the professionalization of computer programmers and related occupations has been fraught and never fully realized. In his account, the push for professionalization among programmers emerged from larger dynamics discussed above, as computers and their applications began to be adopted within organizations. "[C]onflicts within the computing community played out in the development of professional societies, programming languages, computer science curricula, and corporate training and recruitment programs."[41] Ensmenger continues, "[W]hat are dismissed as merely internal debates about the technical features of programming languages, the inclusion of a specific course in computer

---

[40] Moshe Y. Vardi, *ACM, Ethics, and Corporate Behavior*, 65 Comm. of the ACM, Mar. 2022, at 5 (including quotation); Dieuwertje Luitse and Wiebke Denkena, *The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI*, 8 Big Data & Soc'y https://doi.org/10.1177/205395172110477, at 2 (2021).

[41] Ensmenger, *supra* note 6, at 12.

science curriculum, or the imposition of software engineering methodologies for managing development projects are revealed rather as strategic moves in this negotiation over professional status and identity."[42]

Professionalization was also accompanied by two beliefs that persist to some extent today. The first, identified earlier, is a real or perceived shortage of qualified programmers. From early on, there has been a sense that the supply of talented programmers is low. For businesses, this lack of talent became obvious as software became increasingly integrated into business operations and the costs of software failures increased.[43] The second is a view of computer programming as a mysterious art, "whose success or failure was dependent on the idiosyncratic abilities of individual programmers."[44] This understanding has at least two implications. Computer programming as art adds to the sense of impenetrability of computer programs to the uninitiated. But it also contributes to the image of a lone programmer who operates without constraints in cyberspace, which in turn leads to efforts by organizations to control them.

As discussed above, Ensmenger believes that the professionalization of programmers was never fully realized. This failure can be attributed in his view to rivalries between professional organizations, a growing opposition to the movement from businesses, and the inability to justify how professionalization was relevant to business.[45] Ensmenger argues, however, that the root cause is that programmers are better understood as technicians. In this role, programmers "serve as mediators between the technological and social architectures of the organization."[46] For Ensmenger, the term

> captures the tension inherent in the practices of software development: the curious coexistence of high technology and artisanal sensibilities; the inability of programmers to conform to conventional professional, scientific, or engineering categories; the persistent attempts by corporate managers to restructure software development along the lines of traditional manufacturing; and the remarkable persistence of the forty-year-old software crisis.[47]

---

[42] *Id. See also*, *Google fires Margaret Mitchell, another top researcher on its AI ethics team*, Guardian, Feb. 19, 2021, https://www.theguardian.com/technology/2021/feb/19/google-fires-margaret-mitchell-ai-ethics-team.

[43] *Id.*, at 18–19, 24–25, 240.

[44] *Id.*, at 19.

[45] *Id.*, at 193. Brian Jesiek argues that the professionalization of computer design, as opposed to software, has undergone a similar fraught history. Brent K. Jesiek, Between Discipline and Profession: A History of Persistent Instability in the Field of Computer Engineering, circa 1951–2006 (Dec. 13, 2006) (unpublished Ph.D. dissertation, Virginia Polytechnic and State University), https://vtechworks.lib.vt.edu/bitstream/handle/10919/30212/bkj_diss_final.pdf?sequence=1.

[46] *Id.*

[47] *Id.*

This account suggests that as a group, programmers are at least to some extent thought of as independent of their organizations. It also supports the argument that programmers can have values that can differ from their supervisors, leading to the kinds of internal debates about organizational policies discussed above. This is accompanied by a sense of crisis that talented programmers are in short supply. This history also provides context for the professional programming associations that did in fact arise.

The Functions of Professional Associations

Regardless of whether the professionalization of programmers was completely successful, several organizations hold themselves out as professional associations dedicated towards advancing computer programming. Associations have long played an important role in the governance of professionals. Robert Merton argues that a professional association serves purposes for the individual practitioner, the profession, and the larger society.[48] For the individual, the association helps practitioners perform their role as such. An association often recognizes and rewards outstanding performance. It offers continuing education.[49] It gives the professional a sense of membership in a larger community. In this regard, Merton explains, "Each is expected to live up to or to exceed the acceptable standards of practice, and to see that others also do so. This means that the profession develops social and moral ties among its members who enter into a community of purpose."[50] These results have spillover effects that benefit professionals who are not members of the association.[51]

The profession itself gains from the association in several ways. An association can set standards for training and admission into practice, particularly when technical expertise is needed to determine those standards.[52] An association might dictate the requirements for continuing in the practice and set the agenda for future research in the field. It might also publish professional journals. These activities have the effect of raising standards, which tends to improve the profession's stature.[53] Royston Greenwood, Roy Suddaby, and C.R. Hinings note with Merton that professions often organize several associations. These organizations interact with each other "and collectively represent

48 Robert K. Merton, *The Functions of a Professional Association*, 58 Am. J. Nursing 50 (1958).
49 *Id.*, at 51–52.
50 *Id.*, at 52.
51 *Id.*, at 52.
52 *Id.*
53 *Id.*, at 52–53.

themselves to themselves."[54] As a result, "it is from those interactions that understandings of reasonable conduct and the behavioral dues of membership emerge."[55]

Finally, Merton suggests that the professional association performs functions for the wider society. It serves as an intermediary between the practitioner and profession and their "social environment, of which the most important parts are allied occupations and professions, the universities, the local community, and the government." With regard to the government, the association often suggests and monitors legislation.[56] Finally, the professional association monitors compliance with its norms.[57]

Michael Davis argues from a rational choice perspective that professional associations resolve asymmetries in information between professionals and consumers.[58] In his view, although professional associations can create inefficiencies, for example, by creating barriers to entry, associations can also create efficiencies in imperfect markets. In Davis's view, sometimes it is more efficient for the profession to self-regulate than to be regulated by law.[59] Moreover, a professional who joins a professional association gives signals to the market and information to consumers about their competency.[60] "Membership in a profession can thus tell a consumer that the professional in question is formally committed to a certain way of doing things, the profession's way."[61] Finally, standards set by associations allow professionals to protect themselves against market pressures to cut corners.[62] David Thacher explains that a practitioner can resist those pressures by identifying with ethical principles not just as an individual but with a group that has committed itself to them.[63] If the practitioner is "part of a profession whose members have *collectively* bound themselves to the relevant ethical principles, he can respond to this kind of pressure by pointing out that he has a duty to resist it and that

54 Royston Greenwood, Roy Suddaby, and C.R. Hinings, *Theorizing Change: The Role of Professional Associations in the Transformation of Institutionalized Fields*, 45 Academy of Mgm't J. 58, 61 (2002).

55 *Id*.

56 Merton, *supra* note 48, at 53.

57 Greenwood, Suddaby, and Hinings, *supra* note 54 at 62.

58 Michael Davis, *The Use of Professions*, 22 Bus. Econ. 5 (1987).

59 *Id.*, at 7.

60 *Id.*, at 7–8.

61 *Id.*, at 8.

62 *Id.* at 8–9.

63 David Thacher, *The Professional Association's Role*, Commentary, 32 Cities 169 (2013).

every other member of his profession who might replace or compete against him does too."[64]

Professional Associations for Computer Programming and AI Systems

This chapter cannot survey all the professional associations whose membership includes programmers and others involved in computer programming and artificial intelligence systems. This subsection will discuss the work of three organizations: the Association for the Advancement of Artificial Intelligence (AAAI), the Association for Computing Machinery (ACM), and the Institute of Electrical and Electronics Engineers (IEEE).[65] These associations illustrate some of the roles professional associations play in the development of norms related to AI applications.

Association for the Advancement of Artificial Intelligence

Consistent with Merton's description of professional associations in general, the AAAI aims to further research, development, and responsible use of artificial intelligence.[66] The organization publishes a journal, *AI Magazine*, which includes articles on research of interest to the general AI community. It holds academic symposia and conferences. These conferences include an annual conference on artificial intelligence, ethics, and society. One such conference sponsored presentations on topics such as machine ethics, gender bias in natural language processing, race in criminal charging positions, and formal models of fairness.[67]

Thacher argues that codes of professional ethics allow for shared commitments to standards that will apply vis-à-vis others outside the profession.[68] Consistent with that position, the AAAI has adopted a Code of Professional Ethics and Conduct.[69] The code was adopted in 2019 and is based largely on the ACM Code of Ethics and Professional Conduct. In the code, an AI professional, which includes "current and aspiring practitioners, instructors, students, influencers, and anyone who uses AI technology in an impactful way," agrees to a list of ethical principles, professional responsibilities, and

---

[64] *Id.* (citation omitted).

[65] Other organizations include the Association for the Advancement of Artificial Intelligence, the Association for Women in Computing, the Computing Research Association, and the Institute of Engineering and Technology.

[66] AAAI, About Us, https://www.aaai.org/.

[67] Proc. AAAI/ACM Conf. on AI, Ethics, & Soc'y (2021).

[68] Thacher, *supra* note 63 at 169.

[69] AAAI Code of Professional Ethics and Conduct, https://www.aaai.org/Conferences/code-of-ethics-and-conduct.php.

professional leadership principles. Among these principles, AI professionals agree that they should "[c]ontribute to society and to human well-being acknowledging that all people are stakeholders in computing."[70] The code at the same time reflects a preference for the less advantaged: it provides that "[w]hen the interests of multiple groups conflict, the needs of those less advantaged should be given increased attention and priority."[71] The principles also support non-discrimination[72] and acknowledge the value of intellectual property[73] and privacy.[74] Among professional responsibilities, AI professionals agree that they should "[g]ive comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks,"[75] and that "[e]xtraordinary care should be taken to identify and mitigate potential risks in machine learning systems."[76]

Compared with other codes, such as rules of professional responsibility that apply to lawyers, compliance mechanisms are relatively weak. Section 4.1 states that AI professionals "should adhere to the principles of the Code and contribute to improving them." When AI professionals recognize a breach of the Code, they "should take actions to resolve ethical issues they recognize."[77] An AI professional can respond to violations of the code by others, but such action is to be taken "when reasonable" and consists of "expressing their concern to the person or persons thought to be violating the Code."[78] Violations, however, can lead to expulsion from the organization.[79]

---

70 *Id.*, § 1.1.
71 *Id.*
72 *Id.*, § 1.4
73 *Id.*, § 1.5.
74 *Id.*, § 1.6. This article appears to be informed by the GDPR, discussed in Chapter 5. This includes requirements such as collecting data for legitimate ends without violating individual and group rights; preventing re-identification of anonymized data; ensuring accuracy and the provenance of data; protection from unauthorized access or unintentional disclosure; setting policies and procedures as to what and how data is collected; requiring informed consent; enabling individuals to review, obtain, correct, and delete personal data; collecting the minimum amount of personal information needed for a given purpose; establishing policies on retention and disposal; no repurposing data if collected for a specific purpose; and exercising care when merging data sets. Section 1.6 also requires that AI professionals should be informed about privacy. They "should become conversant in the various definitions and forms of privacy and should understand the rights and responsibilities associated with the collection and use of personal information." *Id.*
75 *Id.*, § 2.5.
76 *Id.*
77 *Id.*, § 4.1.
78 *Id.*
79 *Id.*, § 4.2.

Association for Computing Machinery

The AAAI is closely related to the ACM, and it is fair to say that as the older institution, founded in 1947, the ACM is the far more influential of the two organizations. The ACM has about 100,000 members.[80] From its inception, ACM has had close ties with the academic community.[81] It publishes more than 50 scholarly journals[82] and eight magazines, including the *Communications of the ACM*,[83] a periodical that is oriented towards practitioners. Twenty-eight special interest groups within the ACM allow members to focus on specific areas of research and practice. These groups include, among others, groups on artificial intelligence, algorithms and computation theory, computers and society, and data management.[84] As discussed, it has adopted a Code of Ethics and Professional Conduct[85] that served as the basis for the AAAI's code.

The ACM has a Technology Policy Council that is intended to provide "independent, nonpartisan, and technology-neutral research and resources to policy leaders, stakeholders, and the public about public policy issues, drawn from the deep technical expertise of the computing community."[86] The Technology Policy Council has given comments and statements on a number of topics involving AI applications. These include input concerning the European Commission's white paper on artificial intelligence, the UK's national data strategy, computer fraud and abuse, and the use of facial recognition technologies.[87]

Institute of Electrical and Electronics Engineers

The IEEE has over 409,000 members in more than 160 countries.[88] The association states that it publishes a third of the world's technical literature in electrical engineering, computer science, and electronics. Through the IEEE Standards Association, of which companies can be voting members, it

---

80 ACM, About the ACM organization, https://www.acm.org/about-acm/about-the-acm-organization.

81 Ensmenger, *supra* note 6, at 171–72.

82 ACM, About ACM Journals, https://dl.acm.org/journals.

83 ACM, About ACM Magazines, https://dl.acm.org/magazines.

84 ACM, Special Interest Groups (SIGs), https://dl.acm.org/sigs.

85 ACM, ACM Code of Ethics and Professional Conduct, https://www.acm.org/code-of-ethics.

86 ACM, Public Policy, Global Policy and Public Affairs, https://www.acm.org/public-policy. The Technology Policy Council also coordinates the work of a United States-focused policy group and a European policy group. ACM, About ACM's Public Policy Work, https://www.acm.org/public-policy/about.

87 ACM, Public Policy, Global Policy and Public Affairs, *supra* note 86.

88 IEEE at a Glance, https://www.ieee.org/about/at-a-glance.html?utm_source=linkslist_text&utm_medium=lp-about&utm_campaign=at-a-glance.

develops technical standards that become part of industry practice, often with international application. Like the AAAI and the ACM, the IEEE has adopted a code of ethics. Among other things, members of the association agree

> to hold paramount, the safety, health, and welfare of the public, to strive to comply with ethical design and sustainable development practices, to protect the privacy of others, and to disclose promptly factors that might endanger the public or the environment;[89]

The code prioritizes public safety, health and welfare. Moreover, it promotes ethical design and sustainability, privacy, and disclosure of possible dangers to the public and to the environment.

The IEEE has facilitated wide-ranging explorations on the ethical implications of AI systems. This mission is expressed in the second section of Article 1 of the code, where members commit "to improve the understanding by individuals and society of the capabilities and societal implications of conventional and emerging technologies, including intelligent systems."[90] In particular, the IEEE has facilitated discussions of the ethical implications of artificial intelligence through the IEEE Global Initiative, an initiative whose mission is:

> to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity.[91]

The Global Initiative has helped produce an influential document, Ethically Aligned Design (EAD): A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems, now in its third version. The initiative has also created 12 working groups whose goal is to develop standards that focus on specific matters related to the design and application of autonomous and intelligent systems.[92] The IEEE reports that its document on Ethically

---

[89] IEEE, Code of Ethics, art. I.1, https://www.ieee.org/about/corporate/governance/p7-8.html.

[90] *Id.*, art. I.2.

[91] IEEE, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Key Information, Milestones, and FAQs (undated), at 1 [hereinafter IEEE Global Initiative FAQs], https://standards.ieee.org/wp-content/uploads/import/documents/faqs/gieais-faq-11.22.2020.pdf.

[92] IEEE, IEEE Ethics in Actions in Autonomous Systems: Explore our approved IEEE 7000™ Standards & Projects (2022), https://ethicsinaction.ieee.org/p7000/. Such standards, some already issued and others in development, seek to provide guidance for topics such as ethical design; transparency, explainability, and accountability; data privacy; algorithmic bias; and the assessment of the impact of autonomous and intelligent systems on human well-being.

Aligned Design has been cited in work by the OECD in its development of the OECD Principles on Artificial Intelligence (discussed in Chapter 7), the Future of Life Institute, the EU High Level Experts Group, and various companies.[93] The organization has also joined with other standards-setting organizations for cooperation in such standards-setting.[94]

THE ROLE OF THE ACADEMY

The primary mission of universities has been three-fold: teaching, research to add to the general body of knowledge, and to prepare students to become contributing members of their societies. The university has also been seen as an institution that addresses broader societal problems, such as poverty and inequality. They have increasingly been seen as important contributors to economic growth and innovation.[95] More recently, this mission has included social innovation, to work with those who have been marginalized as they work towards freer participation in society.[96]

---

93 IEEE Global Initiative FAQs, *supra* note 91, at 1.

94 *Id.* This section has discussed how professionalism among developers and the associations that represent them play a role in principled AI development and use. Denise Carter suggests that information professionals also play a role in implementing AI ethics principles in their organizations. Denise Carter, *Regulation and Ethics in Artificial Intelligence and Machine Learning Technologies: Where Are We Now? Who is Responsible? Can the Information Professional Play a Role?* 37 Bus. Info. Rev. 60, 65–66 (2020).

95 Aris Kaoudis et al., How Universities Contribute to Innovation: A Literature Review-based Analysis, Norwegian University of Science and Technology, 2019.

96 Blaise Booponoyeng Bayuo, Cristina Chaminade, and Bo Göransson, *Unpacking the Role of Universities in the Emergence, Development and Impact of Social Innovations—A Systematic Review Of The Literature*, 155 Tech. Forecasting & Soc. Change art. no. 120030 (2020); Christina Puente et al., *Role of Universities as Drivers of Social Innovation*, 13 Sustainability art. no. 13727 (2021); Citizenship and Higher Education: The Role of Universities in Communities and Society (James Arthur and Karen Bohlin eds., 2005) (a series of essays on the role of universities in preparing students for citizenship and civic responsibility). James J. Duderstadt, *The Future of Higher Education in the Knowledge-Driven, Global Economy of the Twenty-first Century*, *in* Creating Knowledge, Strengthening Nations: The Changing Role of Higher Education 81 (Glen A. Jones, Patricia L. McCartney, and Michael L. Skolnik eds., 2005).

For a critical assessment of modern universities, see Neil J. Smelser, Dynamics of the Contemporary University: Growth, Accretion, and Conflict (2013). Smelser lists and questions what functions a university truly performs because it is unclear what is meant by the term 'university.' Functions could simply describe what universities accomplish, form a set of goals, justify the university's existence, or be used as a form of advocacy for the university (Part 1, "functions").

Universities and AI Governance

Universities and research institutions have always played an important role in technological development, and artificial intelligence is no exception. Their most significant contributions have obviously come from original research, but universities are also contributing to AI governance. Urs Gasser identifies at least five roles that universities can play in this area. First, universities can provide open resources for AI research, "particularly AI applications and technologies in the public interest and for the social good."[97] Second, universities can devise ways to promote access and accountability. This includes means to assess the accuracy and fairness of AI systems. Schools can help design corrective systems and mechanisms to respond to algorithmic decisions that have undesirable impacts. Third, Gasser argues that universities can help develop methods for assessing the social impact of AI. Fourth, universities can serve as conveners, gathering different stakeholders to discuss issues raised by AI applications. Finally, universities can help translate technical aspects of artificial intelligence to the public.[98] That there may be such a role for universities is supported by a survey conducted by Baobao Zhang and Allan Dafoe. While Americans did not place much trust in any group to develop or manage artificial intelligence, research scientists and the US military were the groups that were most trustworthy. About half of respondents said that they put a "great deal" or "a fair amount" of trust in these two groups.[99]

Visits to the websites of 50 universities that offer well-regarded programs in artificial intelligence indicate that virtually all of them have faculty members whose research areas include some aspect of AI governance or policy. At least 27 of them house institutes, centers, or initiatives that focus on the social and policy impacts of artificial intelligence, along with other transformative technologies. Perhaps it is no surprise that the work of these centers parallels that of the professional associations discussed above, and as will be seen in later chapters, the work of groups within international organizations as well. Activities include engaging in original research; holding conferences and symposia; publishing work, sometimes in scholarly journals, sometimes on blog posts; and making policy proposals. The Montréal Declaration for a Responsible Development of Artificial Intelligence, setting out 10 principles

---

[97] Urs Gasser, *The Ethics and Governance of AI: On the Role of Universities*, Berkman Klein Center Collection, Jan. 21, 2017, https://medium.com/berkman-klein-center/the-ethics-and-governance-of-ai-on-the-role-of-universities-6c31393fe602.

[98] *Id.*

[99] Baobao Zhang and Allan Dafoe, Artificial Intelligence: American Attitudes and Trends, Jan. 2019, https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/general-attitudes-toward-ai.html.

for the development of artificial intelligence,[100] is an example of work sponsored largely by the academy that has been influential in AI governance.

Academic and Commercial Research and Development

As discussed in Chapter 3, substantial resources are needed for artificial intelligence research and development. Governments are an important source of such funding. For example, the US National Science Foundation and other US government agencies have established and funded national artificial intelligence institutes, housed mostly in US universities.[101] Nevertheless, substantial funding for university research also comes from industry. For example, some of the same national institutes are partially funded by companies such as Amazon, Google, Intel, and Accenture.[102] Benaich and Hogarth report that 88 percent of "top AI faculty" have received funding from large tech firms. This includes 84 percent of all computer science faculty, 88 percent of computer science faculty in artificial intelligence, and 97 percent of computer science faculty in ethics.[103]

As Elizabeth Berman recounts, universities have long had relationships with business. This trend became noticeable in the late 1970s. This was due in part to universities' search for additional resources and for industries' need for basic research as companies began to cut their own research and development budgets.[104] Berman argues, however, that the primary reasons for this move were, first, that government began urging universities to view academic science as a valuable product and, second, that it became accepted by government and universities alike that technology and innovation were keys to economic growth.[105]

The linkage between innovation and economic growth had earlier origins, but in the United States grew stronger by the late 1970s and early 1980s.[106] Berman relates that the emergence of the biotech industry broke down what-

---

[100] Text available in 2018 Report: Montréal Declaration for a Responsible Development of Artificial Intelligence, https://monoskop.org/images/b/b2/Report_Montreal_Declaration_for_a_Responsible_Development_of_Artificial_Intelligence_2018.pdf.

[101] National Science Foundation, *NSF partnerships expand National AI Research Institutes to 40 states*, NSF News, July 29, 2021, https://beta.nsf.gov/news/nsf-partnerships-expand-national-ai-research-institutes-40-states.

[102] *Id.*

[103] The State of AI 2019, *supra* note 9, at slide 84.

[104] Elizabeth Popp Berman, Creating the Market University: How Academic Science Became an Economic Engine 2 (2012).

[105] *Id.*

[106] *Id.*, at 45–55.

ever walls existed between academics and firms. "From the earliest years of the biotech industry, university faculty acted not only as consultants but as entrepreneurs. They started firms, served on boards, and announced their advances not only in scientific journals but also at press conferences."[107] The then-nascent industry required both scientists and venture capital, which led to new arrangements between academics and industry.

> [V]ery few people had the knowledge needed to conduct biotech research, and almost all of them were at universities. Most of these scientists had no interest in quitting their academic jobs to become entrepreneurs in an unproven industry. Venture capitalists who wanted to start firms had to find ways to entice academics to participate without asking them to leave their universities.[108]

More generally, university scientists are understood to make valuable contributions to private business.[109] Ulrich Kaiser and his co-authors point out that such scientists are seen as bringing three benefits to firms: general scientific and research skills that can be more effective in defining and solving problems than technical skills; specific problem-solving skills, including the ability to translate information from scientific journals to the firm; and access to university social networks that might be useful to firms.[110] They find that hiring a scientist with academic research experience leads to more innovation output than hiring a scientist without that experience.[111]

Close relations between industry and the university persist in the development of artificial intelligence. The 2021 AI Index Report cited research by Michael Gofman and Zhao Jin, who observed that between 2004 and 2018, there were a substantial number of departures of AI faculty to work in industry. These departures were associated with certain negative impacts on graduates of the programs from which the faculty departed.[112] Benaich and Hogarth report a trend of professors from North American universities moving into large technology companies that has been ongoing since 2004. Thirty-three faculty members associated themselves with tech firms in 2019. Eighty-five percent of such professors were tenured, thus with permanent employment at

---

107 *Id.*, at 59.

108 *Id.*

109 Ulrich Kaiser et al., *Experience matters: The Role of Academic Science Mobility for Industrial Innovation*, 39 Strat. Mgmt J. 1935, 1936 (2018).

110 *Id.*, at 1939.

111 *Id.*

112 2021 AI Index Report, *supra* note 14, at 123–24, *citing* Michael Gofman and Zhao Jin, *Artificial Intelligence, Education, and Entrepreneurship*, posted Sept. 17, 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3449440.

their home universities.[113] This could mean that some faculty members retain their academic appointments while working for technology companies.

Manuela Fernández Pinto identifies three concerns with industry funding of university research. First is a concern with scientific fraud, with prominent scandals associated with the tobacco, pharmaceutical, and petroleum industries. Second is the concern that private funding could impact the type of knowledge produced by universities. Banaich and Hogarth argue in this regard that private funding of AI programs allows the firms to "indirectly craft a desirable public image and influence events, decisions, and research agendas of the universities they fund (particularly top tier institutions)."[114] Third, intellectual property rights given to private industries as a result of research would prevent increases to the fund of knowledge needed for further development.[115] Bennett Holman is equally concerned about these impacts. He points out that the answer to the question of whether industry funding of university research is helpful or harmful might depend on the discipline from which one asks the question. The literature from science policy studies tends to be more favorable to industry funding, but the literature from the philosophy of science is the opposite.[116] Regarding norms, "industry involvement with science almost necessarily infuses commercial values into scientific inquiry."[117] "The primary concern is that the very nature of the university is changing in a self-reinforcing cycle with the nature of science from a mission of public service to a mission for private enrichment."[118]

---

[113] The State of AI 2019, *supra* note 9, at slide 83.

[114] *Id.* As disclosed in the Preface and acknowledgements, my home institution, Seattle University, has received funding from Microsoft Corporation for work in technology and ethics.

It should be pointed out that government funding might also influence research outcomes. One study indicates that federal funding of academic research tends to lead to less patenting by the academic researcher and the number of patents produced (except for general patents). Tania Babina et al., The Color of Money: Federal vs. Industry Funding of University Research, National Bureau of Economic Research Working Paper Series, Dec. 2020, https://www.nber.org/system/files/working_papers/w28160/w28160.pdf.

[115] Manuela Fernández Pinto, *Open Science for Private Interests? How the Logic of Open Science Contributes to the Commercialization of Research*, 5 Frontiers in Res. Metrics & Analytics art. no. 588331, at 2 (2020). *See also* Bennett Holman, *What, Me Worry? Research Policy and the Open Embrace of Industry-Academic Relations*, 6 Frontiers in Res. Metrics & Analytics art. no. 600706, at 2 (2021) (describing potential harms from industry funding).

[116] Holman, *supra* note 115.

[117] *Id.*, at 7.

[118] *Id.*, at 8.

A study of the impact of industry funding on scientific research showed that research funded by the private sector had several effects. As might be expected, corporate-funded research focused on research topics that could be commercialized.[119] The study also examined internal corporate documents that showed a common thread among the tobacco, alcohol, sugar, and mining industries of "establishing research agendas within the industry that are favorable to its positions, strategically funding research along these lines in a way that appeared scientifically credible, and disseminating these research agendas by creating collaborations with prominent institutions and researchers."[120]

Chapter 2 discussed arguments that the various actors in norm formation for artificial intelligence are informed by different logics. This argument appears when contrasting the respective foci of academia and business research. Academic logic is characterized by "the search for fundamental knowledge, research freedom, rewards from peer recognition, and the open disclosure of research results."[121] In contrast, business logic focuses on "applied research in a setting shaped by bureaucratic control, limited disclosure, and the private appropriation of financial returns from research."[122] Henry Sauermann and Paula Stephan argue, however, that distinguishing between logics can be taken too far. Based on their survey of research scientists in academia and industry, they find that there are indeed significant differences in the kind of research the two groups perform, salary levels, and levels of patenting, but there are smaller differences between academic and industry scientists in the freedom they feel in pursuing research and in their likelihood to publish academic work.[123] The authors also find that the two sectors themselves are not uniform; for example, within academia, degrees of freedom, publication, etc., vary across types of universities, types of research positions, and scientific fields.[124] Sauermann and Stephan speculate that different logics could be at play within each sector.[125] At the same time, the authors conclude that to the extent that some differences between sectors do exist, such as pay or patenting, these are not explained by the nature of the work (basic research versus applied science), but rather "reflect deeper differences in missions and value systems."[126]

---

[119] Alice Fabbri et al., *The Influence of Industry Sponsorship on the Research Agenda: A Scoping Review*, 108 Am. J. Pub. Health e1, e11–e12 (2018).

[120] *Id.*, at e14 (citations omitted).

[121] Henry Sauermann and Paula Stephan, *Conflicting Logics? A Multidimensional View of Industrial and Academic Science*, 24 Organization Sci. 889 (2013).

[122] *Id.*

[123] *Id.*, at 904.

[124] *Id*

[125] *Id.*, at 905.

[126] *Id.*

Richard Lester and Michael Piore, writing from the US perspective, argue, however, that higher education is indeed marked by a different ethos that would be weakened if industry were to exert too much influence on the academy, to the detriment of innovation from which industry benefits.[127] The authors argue that certain structural aspects of the university do not exactly match those of businesses: the central functions of the university—education and research—are not directly tied to sectors of the economy.[128] Faculty are organized by disciplines that span all academic institutions. Within a university, even those faculty members who are in disciplines that have potential ties to business are expected to spend the bulk of their time on academic activities.[129]

More importantly, Lester and Piore suggest that universities contribute to innovation by engaging in two processes that industry also performs, but in a different way that would be lost if industry were to have too much influence on university research. All effective organizations, in their view, engage in "analysis" and "interpretation." Analysis refers to basic problem-solving. In analysis, "business consists of a series of discrete problems and an associated series of decisions and choices about which of those problems to solve and how best to solve them."[130] However, analysis might not be aware that there is a problem to be solved.[131] The authors conclude that innovation also requires a second process, interpretation. Interpretation is an ongoing process that has no defined endpoint. This is the activity "out of which something innovative emerges—a new insight about the customer, a new idea for a product, a new approach to producing or delivering it."[132]

Lester and Piore argue that successful universities also engage in analytic and interpretive processes, yet those processes differ from businesses in ways that diversify and thus strengthen innovation in general. With regard to analytic problem-solving, university researchers can choose problems whose solution need not necessarily have economic impacts.[133] Such problem identification can be influenced by the academic disciplines in which university researchers are organized.[134] To the extent practical purposes motivate problem selection, those purposes tend to be public purposes as opposed to pecuniary

---

[127] Richard K. Lester and Michael J. Piore, Innovation: The Missing Dimension (2006).
[128] *Id.*, at 150–51.
[129] *Id.*, at 151.
[130] *Id.*, at 6.
[131] *Id.*, at 7.
[132] *Id.*, at 8.
[133] *Id.*, at 155.
[134] *Id.*, at 155–56.

ones.[135] Similarly, with regard to the interpretive process that encourages conversation as opposed to problem-solving, "the conversation is directed toward advancing the frontier of discovery or toward a particular public purpose."[136] Further, university interpretive processes tend to welcome diversity and more participants through interdisciplinary programs and the participation of faculty and students, potentially from all parts of the world.[137] This is of value to industry, "What mainly attracts product developers to university campuses is the opportunity to participate in interpretive conversations that are different in direction and broader in scope and participation than those within firms."[138]

As discussed, Lester and Piore argue that innovation will be threatened if industry exerts too much influence on the university because it will curtail the distinctive ways in which universities analyze and interpret. Their argument is relevant to this book because it suggests that universities in fact are affected by a different mission, form of organization, and logic. As a result, it should not be surprising that universities have sponsored or housed AI research centers and employ faculty who, despite close ties to industry, nevertheless are influenced by the norms that inform their home institutions. Members of university communities appear to be aware of these issues. Based on information provided by the Center for Security and Emerging Technology, the 2022 AI Index Report indicates that since 2010, there has always been about twice as many research collaborations between educational institutions and nonprofit organizations than between educational institutions and industry.[139]

THE NORM OF OPENNESS

In this chapter, I have argued that it is fair to say that programmers, professional organizations that comprise them, and the universities and research institutions that train or employ them are guided by norms that at times differ from those of business firms. These actors impact to some extent the norms of governance that are emerging at the international level. A good example of such norms is openness in sharing research and tools for the development of technology. This is evident in a general norm in the academy towards open sharing of research (counteracted in part by pecuniary interests and professional competition) and more particularly the open-source software movement.

---

135 *Id.*, at 156.
136 *Id.*
137 *Id.*, at 159–60.
138 *Id.*, at 161.
139 Daniel Zhang, Artificial Intelligence Index Report 2022, Stanford University Human-Centered Artificial Intelligence, at 23 [hereinafter 2022 AI Index Report].

As we have seen, the proprietary model of artificial research emphasizes practical applications and the ability to monetize gains, often through intellectual property rights. This contrasts with the norm of openness within AI development and in broader computer science research. As Brundage and his co-authors note:

> Today, the prevailing norms in the machine learning research community strongly point towards openness. A large fraction of novel research is published online in papers that share anything from rough architectural outlines to algorithmic details to source code. This level of openness has clear benefits in terms of enabling researchers to build on each others' work, promoting collaboration, and allowing theoretical progress to be incorporated into a broad array of applications.[140]

Such collaboration continues to predominate within the artificial intelligence community. The 2022 AI Index Report states that there has been an increase in cross-border and cross-sector artificial intelligence publications over the past decade. The largest number of collaborations have been between researchers from the United States and from China, with more than 2,000 publications in 2010, increasing to over 10,000 in 2020, and dropping off slightly to 9,660 in 2021.[141] Cross-country collaborations not involving US and Chinese researchers also rose during the same period.[142] Cross-sector collaborations are growing between educational institutions on the one hand and nonprofit organizations, industry, and educational institutions on the other,[143] and as mentioned earlier, more collaborations occur between educational institutions and nonprofit organizations.

Openness as a value is reflected in surprising ways. As will be discussed in Chapter 5, several countries restrict exports of technology for national security and foreign policy reasons, but exceptions are made for the results of academic research. In the United States, for example, as a general matter, technology or software that results from fundamental research is not subject to export controls.[144] The US regulation defines fundamental research as "research in science, engineering, or mathematics, the results of which ordinarily are published and shared broadly within the research community."[145] However, the definition is not met if researchers have accepted restrictions on their research

[140] Miles Brundage et al., The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Feb. 2018, at 54. The co-authors discuss how this norm might need to be altered to account for the dual use of AI. *Id.*, at 54–55.

[141] 2022 AI Index Report, *supra* note 139, at 22.

[142] *Id.*, at 23.

[143] *Id.*

[144] 15 C.F.R. § 734.8(a) (2022).

[145] *Id.*, § 734.8(c).

for proprietary or national security reasons.[146] The definition suggests that the primary purpose of the exemption is to promote the open sharing of academic research; it will not protect research collaborations between academic institutions and private companies or the government if those companies or the government imposes restrictions on disseminating the results of those collaborations. Other countries also make allowances for certain kinds of academic research, with varying degrees of restrictiveness.[147]

Another way in which the norm of openness manifests itself is in open-source programming. Saumo Pal defines open-source software as "a type of computer software in which source code is released under a license in which the copyright holder grants users the right to use, study, change, and distribute the software to anyone and for any purpose."[148] Pal argues that the primary motivation for open software and free software "is to propagate the idea of liberty, not price, so that every user is legally free to do what they want with their copies of [open-source software], while encouraging them to voluntarily participate in the design of the software."[149] Open-source software has become an ethos and the basis for communities separate and apart from formal organizations of which members may be a part. Consider this definition of an open-source community:

> An open source software community is a group of people united by the shared purpose of developing, maintaining, extending, and promoting a specific body of open source software. These communities are often globally distributed—their members occupy different geographic regions and work across numerous industries. What unites them is their common vision for the open source software project—as well as the spirit of camaraderie and collective identity that participating in the community affords them.[150]

---

146 *Id.*

147 *See*, *e.g.*, U.K. Export Control Unit and Dept. of Int'l Trade, Guidance: Export controls applying to academic research, Mar. 31, 2021, https://www.gov.uk/guidance/export-controls-applying-to-academic-research (discussing the application of UK export controls to academic research); Commission Recommendation on internal compliance programmes for controls of research involving dual-use items under Regulation (EU) 2021/821 of the European Parliament and of the Council setting up a Union regime for the control of exports, brokering, technical assistance, transit and transfer of dual-use items, 2021 O.J. (L 338) 18–19 (discussing that general technology that results from "basic scientific research" is not subject to EU export controls).

148 Saumo Pal, *History of Open Source Software*, btw Blog, https://www.btw.so/blog/history-of-open-source-software/.

149 *Id.*

150 Contributors, The Open Source Way 2.0, Dec.16, 2020, at 7.

This description shows something of the ethos of the movement. To be sure, a primary motivation for joining such a community is pragmatic: to develop effective software. Open source benefits from crowdsourcing: open-source projects are not limited to only a certain number of employees; in theory, anyone can contribute something to improve the software. But there is more to an open-source community. The vision for the community is cosmopolitan in nature, with individuals from all parts of the world, and transcends individual business firms.

This culture has impacted proprietary technology companies. Open-source software originated in part in reaction to the growing privatization and commercialization of software.[151] Microsoft once criticized the movement as unpatriotic.[152] Now, large technology companies incorporate open-source software as part of their business strategy.[153] GitHub, a repository for open-source software, reports that 84 percent of Fortune 100 companies use GitHub Enterprise.[154] Jim Spohrer argues that this interest is driven by programmers,[155] and Pal suggests that large private companies are trying to promote open-source software and culture at least in part "in a bid to get more developers to join their firms."[156] In a 2018 article, Matt Asay found that programmers from Microsoft, Google, IBM, Amazon and Alibaba were among the top contributors to GitHub.[157] It is not surprising that AI developers also use open-source software. (Richard Stallman, one of the founders of the free

---

151 For a discussion of that reaction from the perspective of a proponent of the open-source movement, see Eric S. Raymond, The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary (1999).

152 Tom Warren, *Microsoft: we were wrong about open source: Microsoft has embraced open source and even Linux in recent years*, The Verge, May 18, 2020, https://www.theverge.com/2020/5/18/21262103/microsoft-open-source-linux-history-wrong-statement.

153 *See*, *e.g.*, Microsoft, Open Source: Our program, https://opensource.microsoft.com/program/; Google Open Source: Projects, https://opensource.google/projects; Amazon.com, AWS: Open source at AWS, https://aws.amazon.com/opensource/?blog-posts-content-open-source.sort-by=item.additionalFields.createdDate&blog-posts-content-open-source.sort-order=desc; Meta Open Source, https://opensource.fb.com/projects/.

154 GitHub, Inc., The 2021 State of the Octoverse, https://octoverse.github.com/#developer-feedback-helps-steer-git-hub-public-policy-commitments.

155 Jim Spohrer, *The Role of Open-Source Software in Artificial Intelligence*, AI Mag., Spring 2021, at 93, 94.

156 Pal, *supra* note 148.

157 Matt Asay, *Who really contributes to open source: New data debunks several myths around which companies lean in open source contributions*, InfoWorld, Feb. 7, 2018, https://www.infoworld.com/article/3253948/who-really-contributes-to-open-source.html.

software movement, closely related to the open-source movement, was a programmer at MIT's Artificial Intelligence Lab in the 1970s and early 1980s.[158]) In 2021, there were about 300 AI-related open-source community projects.[159] Spohrer reports that all the major technology vendors maintain developer websites using open source for the development of AI projects.[160]

Open-source software and the values it represents are international in scope, almost by definition. They have also made their way to formal international organizations. The United Nations is seeking to expand the use of open-source software.[161] The UN believes that open source can help fulfill the mandate to make emerging technologies available to all UN members, thus contributing to the UN's Sustainable Development Goals. Open-source software is seen as a cost-effective way to give this access, particularly to low- and middle-income countries.

TECHNOLOGICAL APPROACHES TO AI GOVERNANCE

It will take a combination of approaches to address concerns about possible adverse impacts of artificial intelligence applications.[162] Because this chapter has explored the technology community as a source of norms for governance, it is appropriate to discuss efforts to use technology itself in the governance process. A significant part of AI research is devoted to these questions, with a large literature being developed. This research area and the tools that are emerging from it are becoming part of the normative structure for AI development and use.

---

[158] David Bretthauer, *Open Source Software: A History*, 7 Published Works (2001), at 4, https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=1009&context=libr_pubs.

[159] Spohrer, *supra* note 155, at 94.

[160] *Id.*

[161] From Open Source to Open Culture—Opportunities and challenges of open source to support the United Nations Mandate, Office of the Secretary-General's Envoy on Technology and Office of Information and Communications Technology, Sept. 23, 2021, https://media.un.org/en/asset/k1o/k1ovtqdxd9.

[162] *See, e.g.*, Miles Brundage, et al., Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims (2020), http://www.towardtrustworthyai.com/ (a report by over 50 scholars, researchers, and others on combining regulatory and technological approaches to making AI trustworthy) [hereinafter Towards Trustworthy AI Development]; Davinder Kaur et al., *Trustworthy Artificial Intelligence: A Review*, 55 ACM Computing Surv. art. 39, Jan. 2022 (a review of the literature on the same topic).

As discussed earlier, Gasser identified the needs for corrective systems and mechanisms to prevent or counteract undesirable algorithmic decisions and for methods to assess the social impact of AI. In a similar vein, to promote trustworthy AI, Brundage and his co-authors recommend the development of software to facilitate audit trails of algorithmic decisions, interpretability, and privacy-preserving machine learning.[163] These systems and mechanisms need not be technological, but algorithms are being developed to perform these functions. This includes methods to mitigate bias in data sets to prevent biased outcomes in natural language processing and facial recognition applications.[164] Other research proposes using blockchain to keep audit trails of decisions,[165] and several methods are being used and developed to explain or interpret them.[166] The same is true with research on approaches to data analytics while preserving privacy.[167] Hardware is being developed to reduce the amount of energy required for computation.[168] This research has resulted in software tools to detect problems as models are being developed, trained, and implemented.[169]

It should be noted that these activities do not take place within a vacuum. Research into technical solutions to issues raised by AI is an area where the

---

[163] Towards Trustworthy AI Development, *supra* note 162, at 21. These recommendations are accompanied by requirements regarding reproducibility of decisions, *id.*, at 21–22, formal verification and validation methods, *id.*, at 22, "scientific protocols to characterize a model's data, assumptions, and performance," *id.*, at 22–23, and enhancing adversarial robustness, *id.*, at 23. The report also identifies several recommendations for assuring trustworthiness of hardware. *Id.*, at 31–37.

[164] *See*, *e.g.*, Timo Schick, Sahana Udupa, and Hinrich Schütze, *Self-Diagnosis and Self-Debiasing: A Proposal for Reducing Corpus-Based Bias in NLP*, 9 Trans. Assoc. Comp. Linguistics 1408 (2021); Yifan Yang, Yang Liu, and Parinaz Naghizadeh, *Adaptive Data Debiasing through Bounded Exploration and Fairness* (2021), https://arxiv.org/pdf/2110.13054.pdf.

[165] Suparna Dutt Dcunha, *Blockchain and AI Is a Good Match*, Datatechvibe, Feb. 16, 2022, https://datatechvibe.com/ai/blockchain-and-ai-is-a-good-match/.

[166] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis, *Explainable AI: A Review of Machine Learning Interpretability Methods*, 23 Entropy art. no. 18 (2020).

[167] *See*, *e.g.*, José Cabrero-Holgueras and Sergio Pastrana, *SoK: Privacy-Preserving Computation Techniques for Deep Learning*, Proc. Privacy Enhancing Tech. 139 (2021); Rongxing Lu et al., *Toward Efficient and Privacy-Preserving Computing in Big Data Era*, IEEE Network, Jul./Aug. 2014, at 46.

[168] Adam Zewe, *New Hardware offers faster computation for artificial intelligence, with much less energy*, MIT News, Jul. 28, 2022, https://news.mit.edu/2022/analog-deep-learning-ai-computing-0728.

[169] Technology companies are making these tools available to developers as part of the companies' software development services packages. *See, e.g.,* IBM, AI Fairness 360, http://aif360.mybluemix.net/ (describing AI Fairness 360, an open-source toolkit for developers to detect and mitigate machine learning bias).

ethical principles and guidelines for artificial intelligence applications have provided some direction, as members within the technology community try to put these principles into practice. Jessica Morley and her co-authors find that programmers see a disconnect between the two.[170] Mark Ryan and Bernd Stahl try in this regard to guide developers implementing high-level ethical norms or values into the design process. For example, many statements affirm the value of non-discrimination. For this value, Ryan and Stahl recommend that

> AI should be designed for universal usage and not discriminate against people, or groups of people, based on gender, race, culture, religion, age or ethnicity. There should be mechanisms in place to effectively prevent, remedy and reverse discriminatory outcomes resulting from AI use…. [O]rganizations should create 'discrimination impact assessments' to identify issues before their AI are used.[171]

Recommendations like these are also relatively broad, but it can be argued that they reflect what is taking place in technical approaches to AI development. The 2022 AI Index Report devoted a section to benchmarks and metrics being used to address problems of bias that arise in natural language processing. For example, Perspective API is a software tool that has been widely adopted by the natural language processing community to identify toxicity, defined as "a rude, disrespectful or unreasonable comment that is likely to make someone leave a conversation."[172] The tool is used to test how often a natural language processing model will respond to a prompt with a toxic comment.[173] More recent large language models with billions of parameters that have been trained on large data sets have been shown to produce more toxic comments.[174] Methods exist to detoxify models, but these have been found to reduce performance of the models.[175] Benchmarks and metrics are also used to detect or measure stereotype bias, gender bias, and stereotypical word associations.[176]

Most hard law on artificial intelligence systems has tried to remain technology-neutral, in part because it is difficult to predict future developments in the field. Such developments might include software tools and methods that can be useful in detecting and perhaps mitigating unwanted effects of AI

---

170 Jessica Morley et al., *Operationalising AI Ethics: Barriers, Enablers and Next Steps*, AI & Soc'y, Nov. 15, 2021, at 6–7.

171 Mark Ryan and Bernd Carsten Stahl, *Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications*, 19 J. Info. Comm. & Ethics in Soc'y 61, 68 (2021) (citations omitted).

172 2022 AI Index Report, *supra* note 139, at 109.

173 *Id.*, at 110.

174 *Id.*, at 111–12.

175 *Id.*, at 113.

176 *Id.*, at 114–22.

decision-making, which suggests that it is useful for a regulatory system to keep abreast of these tools and to encourage their use when appropriate. The use of artificial intelligence for content moderation discussed in Chapter 3 is one example. However, there are several reasons why most commentators view technical approaches as a part of a broader approach to governance. For example, methods to improve interpretability are being developed, but as Brundage and his co-authors remind us, there is a lack of consensus on the meaning of the term.[177] The same is true of methods for preserving privacy: there are few standards for evaluating such methods, and the 'typical' AI developer is unable to implement them.[178] Finally, the very definition of trustworthiness is multi-faceted.

> An AI developer might also be more or less trustworthy based on the particular values they espouse, the extent to which they engage affected communities in their decision-making, or the extent of recourse that they provide to external parties who are affected by their actions.[179]

These components of trustworthiness might be at odds with each other. For example, if a developer is subject to legal recourse, this might create a disincentive to verifiability, because a developer might not want to be held responsible when the measure is not met. This would be true even if, or perhaps even more so, there is a third-party auditor. This is illustrative of a wider problem: bias, fairness, explanations, and recourse can all be construed in different ways. For purposes of governance, persons from outside the developer community must participate in determining what these terms will mean for their communities.

CONCLUSIONS

This chapter suggests that artificial intelligence developers, the professional associations that represent them, and academic institutions are participants in the development and implementation of norms for artificial intelligence governance. Developers of artificial intelligence applications are in a unique position to make technical decisions that have broader impacts. Professional associations in which AI developers, academics and companies participate have recognized the need for the governance of artificial intelligence and have responded by developing ethical guidelines for AI and participating in the development of other forms of governance at the international level. The university, part of whose mission is to engage in basic research, has also

---

177 Towards Trustworthy AI Development, *supra* note 162, at 26–27.
178 *Id.*, at 28.
179 *Id.*, at 6.

responded by establishing centers for the study of AI governance and is itself responsible for AI norms that have had some influence. Individual academics engage in international collaborations and provide their expertise to international governing bodies. At the same time, the chapter has shown how these actors are often in close relationships with private firms, often as employees, consultants, or recipients of funding. Finally, the norm of openness and the practice of developing technical tools that detect and mitigate harms might not stem solely from these actors, but they have been championed by them. Openness seems to be well established even in the face of national security concerns and is represented by the adoption of open-source practices at the international level.