Technology Innovation and AI Ethics

Sylvester A. Johnson, Assistant Vice Provost for the Humanities and Director of the Center for Humanities, Virginia Tech

Introduction

In 2017, Saudi Arabia granted citizenship to a machine, a humanoid robot powered by artificial intelligence (AI) and named Sophia. This woman-gendered robot is manufactured by the Hong Kong–based Hanson Robotics Corporation and it is machine-learning technology that enables her to deliver scripted speech and to participate in spontaneous conversation with humans, complete with facial gestures, intonation, and other forms of body language. Sophia had just delivered a speech at the nation's Future Investment Initiative summit, to which Saudi Arabia had invited hundreds of global investors to consider leveraging the financial growth opportunities the nation is charting for its future. Following Sophia's speech, it was announced that the government had granted her citizenship. Sophia responded with delight, even pondering the possibility of voting and attending college one day.

It seems undeniable that technology innovation is broaching fundamental questions about humanity and ethics. In the wake of Sophia's citizenship announcement, a mix of fascination and dissent emerged. Many people were quite amused that this humanoid AI robot could be so charming. Others lamented the fact that Saudi Arabia had granted citizenship rights to a machine while denying the same to millions of human immigrants. Still others noted that Sophia addressed an audience of elite men while unveiled, whereas human women in Saudi Arabia are traditionally required to veil in public. Amidst the various responses, one thing was certain: granting citizenship to an intelligent machine was a sure sign that AI technology is as much a social issue as it is a technical one.

In a prescient commencement address at Northwestern University back in 2015, IBM's CEO Ginny Rometty identified an emerging paradigm shift, declaring that "the dawn of a new era" is upon us, one in which "every important decision" of humankind will be made not by humans alone, but by human-machine alliances powered by "cognitive computing" systems to enable outcomes beyond anything humans might accomplish on their own.[1] Rometty was right to recognize that such themes as creativity, research, and culture have been traditionally conceived as uniquely human accomplishments in the past and are increasingly being performed by machines and humans working in concert. For several years now, IBM's Watson AI system has been working with human oncologists at the Memorial Sloan Kettering Cancer Center to learn how to develop treatments for cancer. Watson is also being used to assist decision-making in other domains, such as finance, marketing, and concierge services.

> "Granting citizenship to an intelligent machine was a sure sign that AI technology is as much a social issue as it is a technical one."

Although Hanson Robotics's Sophia and IBM's Watson can be competent at very specific tasks such as having a friendly conversation (Sophia) or reading and understanding thousands of articles on a given subject (Watson), humans still reign at so-called general intelligence. We think nothing of the fact that a single human might be equally adept at cooking, composing music, reading a data chart, and building furniture. This is just common sense. This range of ability simply does not exist with artificial intelligence. At least not yet. But in July of 2019, the Microsoft Corporation formally partnered with the formerly nonprofit OpenAI to inaugurate a new collaboration that aims to build the world's first machine intelligence capable of human-level general intelligent comprehension and skill.[2] If successful, such an AI system would be as adept at playing chess and creating recipes as it would be at curing cancer, analyzing foreign policy, planning urban development, and deriving practical solutions to address climate change. This would be a true know-it-all, capable of learning anything on its own without

having to be programmed. The scale of this machine intelligence would far outpace the capacity of any biological human, just as a simple calculator can outperform any human at solving math problems.

Technology innovation is creating immense opportunities to improve the lives of people throughout the world. As is especially evident through advances in artificial intelligence, this innovation is also producing startling quandaries that at one time seemed far-fetched and fictitious, but that now raise ethical challenges for the present and future of humanity.

Why AI Ethics?

As daunting as the technical questions are for fulfilling the vision of an AI-driven world, it appears that the ethics of governing innovation will be even harder. How should we manage technology—how do we shape outcomes, processes, and consequences—to ensure that human society is not only sustainable but also thriving? More bluntly, how will we create a future we actually want to inhabit, rather than one defined by destitution, technological cataclysm, and inhumane conditions? The answers to such questions are not simply technical; rather, they are profoundly humanistic and comprehensive. The judgments and decisions that will shape our human future are ultimately ethical in nature. They mandate consideration of social benefits and costs, of material advantage and disadvantage, and of security, wealth, and well-being.

In an astute article about the future of a digital society, Palmer Group CEO Shelly Palmer voiced a similar concern, explaining that the choice we face is not merely about opting in or opting out of "privacy," but "about our economic sovereignty and our national security."[3] There is abundant concern that resonates with such cautionary voices. Many readers will recall that in 2016, several entrepreneurs and scientists from Elon Musk to Stephen Hawking signed an open letter urging governments to ban weaponized forms of AI until experts

have developed a reliable way to control such technology. In 2018, Google responded to protests from their own employees by agreeing to refrain from developing AI for weapons systems and other forms of destruction.

Technology Innovation and Wealth Distribution

There can be little question that technology innovation is driving transformative changes nationally and globally. Joseph Stiglitz, former chief economist for the World Bank, observes that the economic growth resulting from technology will create unprecedented wealth in the years to come, albeit through drastically uneven distribution and with social implications that will require judicious foresight and humanistic guidance.[4] Consider that in just the next 10 years, digital technology is estimated to add around $100 trillion (net) in GDP to the global economy. Of that amount, AI alone will be responsible for about $15 trillion.[5] It is a staggering figure, but that's just the next decade. What about the next quarter-century? No less a leader than Kai-Fu Lee, the CEO of Sinovation Ventures and formerly president of Google China, has argued that AI will produce a scale of inequality that will create a gaping wealth divide between regions of the world as well as within individual nations. Without some drastic intervention, this inequality will escalate at a speed that previous analysts have scarcely imagined.[6] This is because the massive wealth that AI generates is concentrated into the hands of an increasingly smaller portion of humanity. This is happening at a time of increasing precarity for the middle-class—inflation-adjusted, real wages are stagnant or declining—and dissipating political support for the poor.

Were living-wage jobs to decrease rapidly due to AI automation (consider that AI is already replacing humans in finance and healthcare), millions more people of a previously middle- or upper-class would be plunged into underemployment, unemployment, and poverty. Stiglitz has urged that the current relationship between capital and politics, moreover, has already created an environment

that cultivates more scorn and contempt than compassion for the poor, among whom racial minorities are disproportionately represented. In the United States, particularly, federal and state policies have directed billions of dollars more toward prisons and militarized policing of the poor than toward education and healthcare for those same people. As a result, the country commands the number one spot as the world's top incarceration nation. Absent a drastic shift in American politics, it is difficult to imagine how a technologically driven, rapid increase in AI labor-automation would not end disastrously for most people. The rise of nationalist political parties on a global scale, moreover, that target the poor, immigrants, and racial minorities as an existential threat does not bode well for a future in which accelerating inequality will demand transnational synergies and collaboration to ensure a viable existence for humankind.

Data, Ethics, and Technocracy

Technology innovation, of which AI is an especially powerful example, has proceeded most vigorously through information science. This might be more familiar to contemporary readers as "data science." This latter term has become both a mantra and a chief paradigm for business, culture, entertainment, and security. It was only a few decades ago that most companies had never heard of a chief information officer (CIO). Today, executive administration of an organization's information is as standard as financial accounting. Information, in fact, is now the most valuable asset a company possesses.

Data science uses software (algorithms) to interpret massive data sets (the equivalent of millions of DVDs-worth of information, for instance) to produce insights into the real world. Such large sets of data are beyond what any human could possibly handle. But algorithms quickly churn through thousands of data points on a single individual to discern patterns of behavior so well that the software can reliably predict what people will do. Will you want to purchase a new hat

or lawn mower next month? What type of car will you buy in two years? Technology companies probably know before you do and can sell advertising to the highest bidder standing by to translate your purchasing power into their future revenue.

The world learned from the 2016 US presidential election and from England's Brexit campaign that data science in the form of psychographics might also be used to modify people's behavior. This was the basis for Cambridge Analytica's business model. Using 5,000 data points on a given individual, the company had developed over many years effective means of directing the decisions that targeted voters would make at the polls.[7] Entities ranging from governments to schools to private corporations to law enforcement enjoy unprecedented access to previously unimaginable volumes of data about people throughout the world today. This fact alone poses an immense ethical issue: how much data is too much for any entity to possess? Should data sets be classified as public to prevent them from being monetized? If monetizing data represents a viable path for ethical outcomes in humanity's future, should individuals benefit financially from the use of their data? These difficult questions defy easy answers; but they must be met with deliberation at both practical and regulatory levels if we are to avoid the most undesirable consequences.

As it happens, information is also the central concern of library science and of the educational domain more broadly. This poses both a special opportunity and a perplexing challenge for academic libraries specifically and for the world's educational institutions more generally. The opportunity rests largely with the fact that libraries occupy the center of gravity in a technocratic society because they manage the most valuable asset category the world has known—data, "the oil of the digital era."[8] This also implies that libraries are uniquely positioned to leverage innovation for enhancing the delivery of information to a broad population of learners. The challenge, by contrast, is perhaps best demonstrated by the monetization of digital information.

This is precisely the nut to be cracked in the quandary over the Elsevier corporation, which is self-designated as an "information analytics" (more commonly termed data analytics) company. At a practical level, Elsevier controls access to most of the world's published research. Like other data analytics companies, Elsevier is able to leverage and monetize the insights gained from mining massive amounts of data about users. By an accident of history, academic libraries find themselves obligated to expend billions of dollars for contracts with Elsevier to ensure that the consumers of information (students, faculty, researchers, etc.) can access knowledge in the form of academic publishing. As libraries run up against the limits of their financial resources, they will have to consider what role they will play in the information economy.

Librarians have begun grappling with the ethical nature of this situation and with the imperative of structuring a viable and sustainable future for delivering information.

A major aspect of this ethical challenge that technology innovation is raising for libraries can be put more sharply: what relationship should exist between information (in the form of scholarly research, for instance) and markets? And never mind the adage that "information should be free." In the real world of employees, book purchases, journal subscriptions, capital assets, and institutional finance, such a refrain merely dodges the question. Will academic libraries remain conduits for the behavioral data their users are generating? Should libraries also participate in deriving insights from user data and monetize those insights for ethical ends? Do there exist inherent tensions in this enterprise?

As daunting as these challenges seem, the trajectory of technology innovation appears set to deliver even more complicated quandaries. As machine intelligence achieves greater capacity to read and understand expert material, at what point will AIs be recruited to write scholarly papers on subjects ranging from history to psychology to

economics to oncology and computer science? IBM's Project Debater is already capable of ingesting thousands of articles about a given subject, understanding the content, and debating a human by discerning that person's argument to generate a counter-argument rooted in information-rich analysis.[9] It is already the case that AIs can read in a few hours more than any single human could possibly read in their entire lifetime. So, there is certainly a compelling argument to be made that scholarship produced by AIs has a role to play in advancing expert knowledge to promote understanding, analysis, and innovation.

> "Now is the time for a broad array of experts to anticipate new directions in technology innovation in order to begin shaping an ethical and sustainable future rooted in equitable outcomes."

What would authorship mean in such a scenario? Should an AI be legally recognized as an author, particularly when no human could possibly generate the robust analysis and writing such a system might create? If AIs create publications and disseminate knowledge that relies on the research produced by academics (this reliance on expertise disseminated through academic publishing is the current model), then who should benefit from any monetization of such authorship? Machine-learning systems are already leveraging existing research (generated by humans) to derive insights for treating or curing disease. Who will own the capital (servers, algorithmic design, cloud-based services, data sets, and so forth) that is the basis for the digital domains of technology innovation as AIs join the rank of academic publishers?

It might be tempting to simply hope such developments never occur. And yet, Microsoft and OpenAI have already forged an alliance to make this scenario look like child's play. As existing AIs can already write poetry, short stories, and newspaper articles, it is certain that even a minimally successful product that barely approximates the goal of artificial general intelligence (AGI) could mean the irreversible transformation of the expert knowledge economy. It should be clear

that hoping against the future is not a strategy. Instead, now is the time for a broad array of experts to anticipate new directions in technology innovation in order to begin shaping an ethical and sustainable future rooted in equitable outcomes.

Future Humans

Perhaps the horizon of technology innovation will increasingly be shaped by developments in human enhancement or human engineering. Advances in human-machine combining (cybernetics) and genetic engineering promise to create radical changes to human society and unprecedented questions of ethics, equity, and accountability that will easily match or exceed those being generated by AI. Today, every major military industrial state is racing to develop capacities in military soldiers that surpass those of unmodified humans. These efforts include exoskeletons, drug enhancements, brain implants, and "smart" (AI-driven) prostheses; such efforts would permit soldiers to carry heavier loads greater distances, control tools or weapons by thinking, process information more quickly than a normal human by interfacing with an AI system, or function on alert for days without sleeping. Given the high stakes of military dominance for which the world's most powerful military nation-states compete, there is every reason, as well, to expect genetic engineering to emerge in military applications on a global scale.

> As our global society increasingly recognizes that technology is not merely technical but also societal and human-oriented, new doors of opportunity are opening for humanists to take leadership of the most important efforts that might shape the future of society.

Medical therapies constitute arguably the most compelling motivation for aggressively pursuing human enhancement. It is one thing, after all, to rationalize modifying humans for warfare, which inherently involves killing and destruction. It is quite another to justify modifying humans

as a means of preserving life and restoring capacities; these efforts can enable patients to regain speech and motor function, for instance. Even this medical context for enhancement is generating well warranted concerns about ableism and eugenics, particularly as the meaning of a "normal body" or "normal capacity" is reshaped by this technology. It seems unlikely on ethical grounds, however, that such valid concerns will be used to deny everyone even the possibility of regaining the ability to walk again or to have impaired vision or hearing restored through technological enhancements as the development of these technologies continues to advance.

All of this means that technology innovation is on pace to reshape the future of humanity in deeply consequential ways, including at the foundational level of what it means to be (a) human.

What Should AI Ethics Look Like?

As our global society increasingly recognizes that technology is not merely technical but also societal and human-oriented, new doors of opportunity are opening for humanists to take leadership of the most important efforts that might shape the future of society. The University of Oxford announced with great fanfare in June of 2019 that Blackstone CEO Stephen Schwarzman had gifted more than $188 million to fund a humanities center housing a new institute for AI ethics. The billionaire-philanthropist had previously donated $350 million to MIT to create an institute-wide "College of AI" that will emphasize the role of the liberal arts and human sciences. In 2019, Stanford University launched a new institute harnessing university-wide efforts to support human-centered AI, placing at the helm a philosopher and a computer scientist.

We are witnessing growing efforts to ensure that technology serves human interests through regulatory efforts, ethical frameworks, and more comprehensive education. "Public interest technology" is among the key growth areas devoted to ensuring that social justice and

equity guide the development and implementation of technology. Since 2016, philanthropic foundations have devoted millions of dollars to support this new approach for American universities to prepare a new generation of "technologists" to work in civil service, education, and a full range of humanistic endeavors. The daunting challenges of AI have motivated major technology corporations such as Apple, Google, and Microsoft to emphasize fairness,

> "After decades of worries that the popularity of science and technology paradigms threaten humanistic learning and scholarship, it is now becoming evident that unique opportunities are emerging to demonstrate why humanistic expertise and informed considerations of the human condition are essential to the very future of humanity in a technological age."

ethics, and humanistic approaches to innovation. The London-based company DeepMind has made AI ethics central to the guidance of its technology. Since 2018, Google has published AI principles underscoring their commitment to fairness and avoiding developing AI weapons and other harmful forms of AI technology. In 2019, China's Beijing Academy of Artificial Intelligence published ethical guidelines emphasizing fairness and sustainability. The World Economic Forum has likewise articulated ethical guidelines for the use of technology.

Notwithstanding this important beginning, bringing ethical governance to technology will require a thoroughgoing transformation of humanities leadership. Colleges and universities will need to invest greater resources in humanistic programs of study. Humanistic disciplines must focus more urgently on recruiting and producing far more racial minorities and women in technology. Because technology innovation will bring massive changes to our democratic institutions and social systems, future technologists will have to include people with expertise in the human condition, policy, and social services.

Equally important will be transdisciplinary communities of research and collaboration that must provide teams of diverse talent and expertise to guide the use of AI in higher education, law enforcement,

medicine, finance, and warfare. As things currently stand, there exists no regulatory framework for governing technology innovation. The good news is that the challenges posed by the ethical guidance of AI and other forms of technology innovation will require our social institutions to embrace new forms of leadership from humanities experts. After decades of worries that the popularity of science and technology paradigms threaten humanistic learning and scholarship, it is now becoming evident that unique opportunities are emerging to demonstrate why humanistic expertise and informed considerations of the human condition are essential to the very future of humanity in a technological age.

Endnotes

1. Virginia Rometty, "Northwestern University Commencement Address," C-SPAN, June 19, 2015, video, 14:33, https://www.c-span.org/video/?326217-1/virginia-rometty-commencement-address-northwestern-university.

2. Kelsey Piper, "Microsoft Wants to Build Artificial General Intelligence: An AI Better than Humans at Everything," *Vox*, July 22, 2019, https://www.vox.com/2019/7/22/20704184/microsoft-open-ai-billion-investment-artificial-intelligence.

3. Shelly Palmer, "Governing a Digital Democracy: Unanswered Questions," LinkedIn, February 10, 2019, https://www.linkedin.com/pulse/governing-digital-democracy-unanswered-questions-shelly-palmer/.

4. Joseph E. Stiglitz, *People, Power, and Profits: Progressive Capitalism for an Age of Discontent* (New York: W. W. Norton, 2019), 35–38, 46.

5. Barry B. Hughes et al., "ICT/Cyber Benefits and Costs: Reconciling Competing Perspectives on the Current and Future Balance," *Technological Forecasting and Social Change* 115 (February 2017): 117–130, https://doi.org/10.1016/j.techfore.2016.09.027; Frank Holmes, "AI Will Add $15 Trillion to the World Economy by 2030," *Forbes*, February 25, 2019, https://www.forbes.com/sites/greatspeculations/2019/02/25/ai-will-add-15-trillion-to-the-world-economy-by-2030/. PwC Global

puts the figure in the same range, at $15.7 trillion, exceeding the current combined GDP of China and India. See *Sizing the Prize: What's the Real Value of AI for Your Business and How Can You Capitalise?*, PwC Global, 2017, https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. The McKinsey Global Institute estimates a slightly lower but similar figure, $13 trillion. See Jacques Bughin et al., *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*, McKinsey Global Institute, September 2018, https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx.

6. Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (Boston: Houghton Mifflin Harcourt, 2018), 170–173.

7. Varoon Bashyakarla, "Psychometric Profiling: Persuasion by Personality in Elections," *Our Data Our Selves*, Tactical Tech, May 18, 2018, https://ourdataourselves.tacticaltech.org/posts/psychometric-profiling/; Terrell McSweeny, "Psychographics, Predictive Analytics, Artificial Intelligence, & Bots: Is the FTC Keeping Pace?," *Georgetown Law Technology Review* 2, no. 2 (July 2018): 514–530, https://georgetownlawtechreview.org/psychographics-predictive-analytics-artificial-intelligence-bots-is-the-ftc-keeping-pace/GLTR-07-2018/.

8. "The World's Most Valuable Resource Is No Longer Oil, but Data," *The Economist*, May 6, 2017, 9, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

9. "Project Debater," IBM, accessed September 10, 2019, https://www.research.ibm.com/artificial-intelligence/project-debater/.