Proactively Protecting Against the Singularity: Ethical Decision Making in AI

Dawn E. Schrader | Cornell University
Dipayan Ghosh | Harvard University

This article proposes an ethical framework for the development and implementation of artificial intelligence that is based on philosophical principles and perspectives upholding human rights and well-being above potential superior machine intelligence referred to as the *singularity*. Illustrative cases demonstrate the framework's application and suggest guidelines for future policy.

Artificial intelligence (AI) has subtly permeated our lives. We awaken to the machine-tuned aroma of our breakfast coffee and fall asleep reading our algorithmically designed social media feeds on platforms like Facebook and Twitter.[1] Increasingly, individualized applications of AI are being developed by businesses and used by consumers to power human decision-making processes, reduce the time needed to complete everyday tasks, and automatically inform us of events and incidents in the social and political world around us. Few would argue that this class of technologies has not enhanced our life experiences and made our lives more connected, efficient, and enjoyable.

Modern AI consists of complex algorithmic systems that train themselves to think and act as intelligently as humans. The concept of the *singularity* has been proposed and elaborated on by computing minds such as Alan Turing and J. Good, as well as by science fiction writer Vernor Vinge. Briefly stated, the singularity refers to that time in which artificially intelligent systems and their self-created descendent systems transcend the quality of human thought and operate beyond human control.[2] Although we may still be decades away from developing AI that can equal or best the quality of

human intelligence, algorithms are used daily in new ways that challenge us to consider what influence they may have on human psychology, self-determination, and autonomy.

This stated, it may be clear what AI's threat might be. Humans are psychologically influenced by interaction with others and their environment in the process of their learning, decision making, and behavior. For many, AI is increasingly a significant element of this environmental interaction; AI systems shape and define human behavior over time by revealing private information online, allowing continual voice recording in the home, influencing purchase preferences and even partner selection. While helpful in many cases, AI may be seen as a fundamental threat to the individual's privacy, security, and autonomy.[3] Thus, AI does not come without risk, particularly in dialogical consumer-facing contexts—that is, in those situations where an individual interfaces with a machine that has the capability to directly inform, shape, and alter behavior without human mediation.

In protecting against the singularity, where AI systems are indeed better than human intelligence, the fundamental threat is that AI systems would drive human

Copublished by the IEEE Computer and Reliability Societies

thought and behavior in ways that are distinctively calculated, logical, and utilitarian, while the more human aspects of empathy and intuition—which allow for the particularity of exceptions to rules which might better serve humanity—are relegated or ignored. The threat is that AI systems create and control humans, not vice versa. Essential for a safe and secure human future is the ethical mandate to bring heightened awareness to consumers about the manner and extent to which technology shapes our personal and work lives, and of the influential power that AI and advanced algorithmic technologies have over human thought processes in our political, social, and economic lives. This article presents a framework as a guide to ethical AI systems, particularly those that interact dialogically with humans.

Model of Ethical AI Systems

To address this potential intelligence reversal in which the machines instead of humans are "in charge," we propose a framework for ethical AI. Our premise is that AI systems interacting with humans via computer-mediated communication and human–computer interaction ought to follow the ethical principles derived from both traditional and communicative ethics principles. Further, AI systems should be developed with consideration for ethical practices prior to, rather than reactive to, ethical violations. We do not expect all developers and policymakers to become moral philosophers, but we provide a framework by which those involved in the development of AI consider the human and ethical impact of their work.

Even as AI technologies create new efficiencies and economic gains for consumers and the industry, deep ethical concerns about privacy and human autonomy arise. Incumbent on industry is the imperative to act ethically—that is, with integrity and credibility—particularly because the interactions between AI and humans shape human thought and action. The development and implementation of fair and transparent algorithmic operations and protections of privacy and security are foremost of the necessary ethical protections.

Industry has already taken note of some of these concerns. The "Partnership on Artificial Intelligence to Benefit People and Society" (www.partnershiponai.org) was created by industry players including Google, Facebook, Amazon, Microsoft, Apple, DeepMind Technologies, and IBM. Among their goals is to "address such areas as fairness and inclusivity, explanation and transparency, security and privacy, values and ethics, collaboration between people and AI systems, interoperability of systems, and of the trustworthiness, reliability, containment, safety, and robustness of the technology."

Guiding Framework: Three Dimensions for Ethical AI Development

The partnership mentioned above echoes classic ethical concerns that must be brought to light in the next technological revolution in which AI will be playing a dominant role. This article proposes a framework to help guide conversations of ethics in AI development and implementation. We contend that it is necessary for any and all conceptual developments of AI to be ethically grounded and be operative in the three dimensions that ensure ethical practice. Ethical AI should at once be:

- *proactive*—anticipative of potential ethical violations and problems before they arise;
- *engaged*—involving meaningful dialog and deliberation between people, systems, and industry; and
- *good*—targeting the enhancement of human life both individually and socially as well as meeting ethical standards of excellence.

The framework proposed here integrates established ethical principles from traditional philosophical thought together with newer philosophical ideas from communication ethics and flourishing ethics. These newer *ethical perspectives* are more relevant to dialogical contexts and reflect a "socially constructed" viewpoint.[4–6] We claim that technologies developed through participatory dialogical human engagement are more likely to be protective of human rights because equal participation and open transparency are inherent in both the process and the product of such engagement.

Ethical Perspectives

Six prevailing streams of philosophical ethical thought form the *ethical perspectives* of this proposed framework. We suggest these six perspectives be used along with *core functions* described below to guide ethical design of AI systems.

1. *Rights:* Deontological ethics, focusing on universal rights, moral duties, prescriptions, and obligations as rational moral agents. Assessments are based on intrinsic, generally applied acts and laws that meet the criteria of Kant's categorical imperative. As such, this view concerns fundamental human rights, upholding personhood, and ensuring all treat others equally with fairness and respect (Immanuel Kant, John Rawls).

2. *Goods/Harms:* Teleological ethics, focusing on the "goods" that ought to be pursued, often considering the harmful or beneficial consequences to individuals or society. Utility theory is incorporated in this viewpoint, including hedonic pleasure and

utilitarianism of many kinds. As such, this view concerns maximizing benefits, minimizing harms, and promoting goods (Jeremy Bentham, John Stuart Mill, Thomas Hobbes, Henry Sidgwick).

3. *Virtue:* Aretaic ethics, focusing on "virtue" of the actors—their character and integrity; beneficence. For example, Aristotle identified a list of 11 moral virtues in his Nicomachean Ethics, also articulating the "golden mean" between two vices. As such, this view concerns being "a good person" (Plato, Aristotle, David Hume; sometimes Alistair McIntyre).

4. *Community:* Communitarian ethics, focusing on the mutual constructions of self and community and that community as fundamental to, and imperative for, shaping common values that form the basis of all citizens' obligations and responsibilities as members of society, and that all have equal power and participation. As such, this view concerns citizen-constructed responsibilities for moral conduct (Amitai Etzioni, Michael Sandel, Charles Taylor).

5. *Dialog:* Communicative ethics, focusing on reason and rationality as people discuss things to understand each other's thought processes and perspectives and judgments. This view necessitates free and equal participation in dialog to create ethical norms for living together in a mutually respectful world. Both the process of dialog and the resultant agreements are moral; dialog continues until the "better" argument for what is right is achieved through consensus (Jürgen Habermas).[4]

6. *Flourishing:* Flourishing ethics, focusing on the human flourishing to reach one's fullest positive psychological and social functioning; goodness and well-being. This view concerns bringing in earlier ethics theories to the human, animal, and technological world to ensure maximizing existence (Douglas Rasmussen, Terrell Ward Bynam).[5,6]

Core Functions

Our goal of protecting against the potential harms of the singularity is to be met by leveraging the above *ethical perspectives* in the design of AI, and to do so in a way that is proactive, engaged, and focused on the good. To do so, we suggest some *core functions* to implement. Our proposed framework mirrors the structure of the National Institute of Standards and Technology (NIST) Cybersecurity Framework that defines a "Framework Core" as "a set of cybersecurity activities, desired outcomes, and applicable references that are common across critical infrastructure sectors."[8] Our proposed AI Ethics Framework similarly implements the idea of *core functions* to represent ethical activities and outcomes across philosophical ethical perspectives. The *core functions* for ethical development of AI

incorporated in this framework include the following five components.

1. *Identify ethical issues of AI*: This function encourages all parties to recognize the role of AI systems in human–technology interactions and to further acknowledge that ethical concerns around fairness, transparency, equity, goodness, beneficence, social utility, human flourishing and happiness, and protections for human agency exist. This function assists in identifying that privacy protections and security in systems are grounded in the ethical principles of rights to self-determination and happiness.

2. *Improve human awareness of AI*: This function addresses human understanding and cognizance of how AI systems work within the devices used, and how industry is creating algorithms from collected data and using, storing, protecting, and responding to threats and breaches or invasion. For example, are there aspects of informed consent? Are there "just in time" notifications?

3. *Engage in dialogical collaboration with AI*: This core function includes the idea of collaboration between humans and technology—comprising dialogical interaction, listening, and understanding. The function can ensure that the development of AI systems contains opportunities for ethical engagement between human and AI and computer technology as well as opportunities to co-construct with industry the nature and direction of the development of the systems that daily interact in human lives.

4. *Ensure the accountability of AI*: Borrowing from the ACM Code of Ethics and Professional Conduct,[8] this function addresses the adherence of ethical conduct of AI systems and those who design them. Specifically, the moral imperatives of the ACM code adapted for this framework include the initiatives to: contribute to society and human well-being, avoid harm to others, be honest and trustworthy, be fair and non-discriminative, honor human and property rights (including intellectual property), respect privacy, honor confidentiality, and evaluate and improve on an ongoing basis.

5. *Maintain the integrity of AI*: Drawing from and adapting the ACM code, this core function contends that AI systems should be of the highest quality, maintain professional competence, and be limited to the purpose for which the technology was developed. The use and development of AI systems should respect the law and honor agreements.

These *core functions* and *ethical perspectives* are to be applied and examined across the dimensions introduced earlier, namely proactivity, engagement, and

Table 1. AI Ethics Framework Think Sheet

Core Functions	Ethical Perspectives					
	Rights	**Goods/Harms**	**Virtue**	**Community**	**Communicative ethics**	**Flourishing**
Identify ethical issues in context	issues of justice, fairness, rights "is this right" and universalizable	what are the benefits and harms or consequences of doing/not doing	what is upstanding and admirable	the social community cohesion and enhancement	respect for process of listening and responding	life quality improvement
Awareness of ethical issues and AI function	rights are considered	cost/benefits	how AI can be good	enhancing community	inclusive respect	how AI improves well-being and individual life
Dialogical collaboration	all voices heard; seek who is not heard	utility of cooperation and social disclosure	shared understanding of virtuous things and their acquisition	democracy and respect of diversity of viewpoints	opportunity for understanding and mutual respect	support and respect of mutual well-being
Accountability of AI system	responsible to and for what is right	consequences and utility of AI, security, protection	responsible to be ethical, secure	consistent, reliable, predictable, responsible to each other	seek divergent viewpoints; be comprehensive	do not harm; focus on positivity and doing well/right
Integrity of AI and others	morally consistent; "perfect" duty	responsible moral systems; utility for the overall good	system, industry, and people act ethically; "don't be evil"	reliability, responsibility to all and institutions	listen, respect, share; open to the "better argument" of what is right, good, virtuous, etc.	consistency, honesty, security, freedom

goodness. Like NIST's Cybersecurity Framework,[7] our AI Ethics Framework is not "one size fits all." We recognize that AI systems and contexts have unique purposes for individuals and industry, and that inherent in that are different threats, vulnerabilities, and risk tolerances in interacting with the AI in that technological system.

Accordingly, we propose a rubric that can aid designers in conceptually working through issues of ethical AI systems development (see Table 1). This visual framework or "think sheet" can be used as design principles to work through AI system development and implementation to help ensure ethical proactivity, engagement, and moral goodness throughout the systems' application. The *core functions* and perspectives of the ethical framework are explained here using the first dimension— proactive AI ethics—for brevity sake, but the think sheet is designed to also address the engaged and the good dimensions similarly. This "before, during, and after" (proactive, engaged, and focus on "the good") approach not only safeguards the ethical development of AI but additionally presents opportunities for active ethical reflective processes among the systems designers and in implementation.

We now turn to examples of how to use this AI Ethics Framework from broad consumer-facing contexts to demonstrate the application of the framework for ethical AI systems. Despite the seemingly diverse contexts of the examples, we hope to illustrate that some ubiquitous attributes of ethical consideration are useful for generating policy and procedures for protecting against the singularity.

Case Examples

We apply the aforementioned framework for ethical design of AI to two case studies of AI systems that occur at an individual and social level, namely, social media and personal assistive technology. These two examples effectively juxtapose contexts in which consumers

engage in broad, public communication with a network of friends against contexts of private communication for planning individual daily activities. This offers a spectrum of dialogical applications from broadly social to more private communication in modern technology. The goal is to demonstrate how the AI Ethics Framework may guide the development of ethical systems that not only protect against the singularity, but concomitantly protect privacy, security, and human autonomy in the use of these intelligent systems.

Our analysis of the ethical issues in these two technological contexts may shed light on the generalizability of the AI Ethics Framework broadly to consumer-facing applications of AI. Note that we select these two specific contexts as cases rather than making broad generalizations about all AI systems as a guide to how to use the framework rather than to proscribe solutions, because differing ethical issues, risk assessments, and protections ought to be programmed into social and personal assistive AI systems. For example, AI systems in transportation and smart cities—such as autonomous vehicles use and the security of physical spaces—have their unique contextual particularities that may be dissimilar to our examples but nonetheless can draw from *ethical perspectives* and functions therein. We believe the framework has implications for these contexts as well as larger social economic impacts, which are points for later research.

Social Media News Feeds

Internet-based social media products developed and delivered to consumers by leading companies including Facebook, Twitter, and Google are today collectively host to multitudes of one-to-one and one-to-many human interactions. Together these social media services have become the primary means for human collaboration and communication for many.

Subtly, the designers of the leading social media applications are integrating machine learning systems and weak AI into these technologies at a prodigious rate to take advantage of greater efficiencies in content engagement, digital ad targeting, and ease of interaction with other users. Users, meanwhile, typically are not aware, and do not understand, that AI is implemented into social media applications; how their private information is collected, used, disseminated, and fed back to them; and how it subtly influences their choices in the marketplace, in social relationships, and in their own minds and emotions. Further, the industry's inclination is to not make the changes transparent to consumers. Similarly, the industry prefers not to explain how and why it implements AI, nor does it typically consult with users in a meaningful way about their preferences prior to implementation.[9]

Examining news feeds in social media, we see the functions and principles at play—but not always in a manner compliant with the ethical framework. There often exists a lack of awareness on the consumers' part, which leads to distrust; dialog and communicative respect are missing, and opportunities for accountability of the AI system and respect for the integrity of the user are lacking.[9] We argue that due to the lack of ethical action on the part of many developers of AI systems in the context of social media platforms, users are at risk of unintended and unethical breach of privacy and data security as well as disrespect for persons' sense of self and their views or emotions. One need not look further than the now infamous Facebook's study that demonstrated negative emotional contagion via manipulation of news feeds, which illustrates this point.[10]

While there are many instances of social media applications by the companies listed above and others, here we examine news feeds, which are prevalent as a feature commonly implemented in many social media platforms. These can be personalized, easily and comprehensively presenting each user with a series of content that he or she is likely to most wish to see as determined using advanced machine learning algorithmic technology, a class of weak AI. Individual privacy is obviously invaded; despite their subscriptions to terms of use agreements, many users may be unaware of the nature and extent to which news feeds are manipulative and can additionally can present "fake news" or biased reports to users.

To examine the privacy issues further, it is important to note that AI embedded into a news feed feature may, over time, assess specific information—for instance, that a particular user's two greatest interests are her immediate family and her educational institution. The AI might make these assessments through analysis of the user's past interactions within the news feed and on other media such as other websites through the use of tracking technologies like cookies. After making these assessments, the AI might rank the content shared by this user's immediate family and educational institution highest in her news feed so that she is likely to see and engage with the content. Such ranking can, over time, maximize her potential engagement with the news feed platform, which is in the interest of the social media platform provider.

If users are aware of these systems and how they interact with private information, then users may come to believe it is to their benefit to dialogically engage these systems to be more aligned with their best interests, such as being informed about their communities. Similarly, users may personally decide to engage in dialogical participation with both the platform's systems and other users to foster social engagement. They might also assume control over what is shared and how to go about changing their interests and preferences. Such actions

would uphold the core ethical functions within our proposed ethical framework related to identifying potential issues, awareness of systems, dialogical interaction, and accountability and integrity of systems' interaction with users and their data. The ethical principles or ends that could be upheld are furthering community, the social good, fostering honesty and other virtuous character in participating in creating and sharing news, and feeling empowered in their communicative interactions with others and society to create a better life or world.

It is essential for users to know that a technology recommends content for them to see, because news feed platforms and other social media services often constitute their primary medium for social interaction—which can necessarily determine what information a user will consume, thus influencing her social perspective. With knowledge that an AI technology is behind ranking the news feed, users may, for example, learn that there are ways to change the ranking or become more empowered to turn off such a technology should the option exist. Considering the scale of reach, dialogical interplay, and capacity for future innovation in the news feed context, as well as the potential for security and privacy infractions, it is critical for the designers of social media news feed technologies to inform their users of the integration of AI technology for such tasks as content ranking and moderation.

Various companies are already transparent about their use of and forward investment in such technologies as machine learning and AI, but they often do so in ways and contexts that do not make that fact readily apparent to users. Instead of making such announcements at industry conferences or on blog posts, for example, firms should communicate implementations of AI in social media through such formats as relatable data use policies and contextual consent messaging to users. Wherever possible, consistency across the industry, for example, through trade association member alignment efforts or other voluntary industry agreements, should be sought to minimize demands on the user's attention to such details. Communication ethics and dialogical engagement are fundamental to the further development of AI systems.

We secondarily note that users must be empowered to actively engage with the AI and its development. In the absence of this capacity, users cannot communicate implicit concerns they may have with the technology—whether those concerns may relate to privacy, autonomy, or security, among others. Facebook, for instance, has engaged directly with users in the past on such issues as privacy. For instance, it launched its Ask Our CPO feature in 2013, enabling any user to voice concerns directly to the company's principal privacy official. This kind of openness and transparent practice serves the *core functions* of accountability and integrity as well as upholds the *ethical perspectives* related to rights, individual autonomy, and communicative ethics.

Third, we would note that news feed AI should be accountable in the context of the rights of the user as articulated in the fourth core function. In particular, respect for individual privacy and fairness are critical to developing a fair technology. Other aspects of this core function lead into the integrity function and include the upholding of honesty, trust, confidentiality, and iterative improvement for the overall good of humans and society. In sum, the AI Ethics Framework we propose can support analysis of the use of AI systems in news feeds and related social media affordances that work similarly deeper and more comprehensively, and be more likely to ensure human well-being as this technology develops.

Voice Assistive Systems

Voice assistive systems (for instance, Amazon's Alexa or Apple's Siri) are in wide use throughout the world today. These systems can improve efficiency and enhance well-being. With such benefits, however, also come risks; these systems often monitor and track their users in ways that can implicate individual privacy, security, safety, and autonomy. It is thus vital that the development of AI systems be privacy aware and ethical in data collection, use, storage, and dissemination.

Voice assistive technologies, computers, and people are key interfaces with AI. Prior to the advent of voice assistive technologies, the widespread use of AI in everyday consumer life required human–computer interactions, such as making consumer purchases that then influenced your news feed or web searches. Voice assistive technologies now use AI that infuses your former computer interactions into the AI system, and in a more human-like and friendly way. This helps you to listen to your music on command or to request an action or to purchase an item online without having to go to your computer, tablet, or phone. People share private consumer information with this new AI system for their own convenience, perhaps not aware of the privacy concerns and conceivably even more unaware that they are unwittingly assisting the AI to iteratively improve. Alexa listens ubiquitously, and developers currently are analyzing such things as not only the content of information shared, but the tone and intonation of the requests. This seems very personal and private.

These technologies are seemingly at the other end of the public engagement spectrum from social media applications whose role is to portray to the user global content in a singular way. Voice assistive technologies tend to be used by a single human interlocutor who asks or directs something and involve direct one-on-one dialog to a specific end, and thus constitute a wholly different contextual experience for the human.

That said, voice assistive technologies raise many of the same concerns as our prior example of social media news feeds. Specifically, users tend to be unaware of how and why AI is active in these systems. To be transparent and respectful of human rights to privacy and self-determination, engineers of AI systems ought to provide guidance to users as to the fact that AI technology is being used to power the technology. Following the *core functions* and *ethical perspectives* outlined in the AI Ethics Framework we propose can assist with the development of the systems and education about their use in human lives. Two *core functions* that have some degree of overlap are the function of dialog and the perspective of communication ethics. Engineers and industry should implement ways to understand user feedback about this dialogical technology so that improvements can be made over time. In particular, developers should address any concerns over negative psychological or social impacts these technologies may have on their users, such as the revelation of health issues; personal, political, or religious preferences; and the like. Voice assistive technologies should uphold privacy laws and information practices and maintain the security of the interactions between humans and the systems. Upholding these regulations meets the AI Framework's accountability and integrity core functions and additionally serves to protect fundamental rights such as protection from harm and promotion of the social good.

It is important to note the distinctions between social media news feeds and voice assistance systems on more than their face. In regard to user privacy, social media interactions are often construed as either public or limited to specific sharing groups. That is, the goal is to share. In contrast, users might have heightened privacy expectations in their interactions with voice assistance technologies because they are usually used in private contexts. The voice AI interface typically seems more personal and often has a name and a gender associated with it. It is not perceived as connected to a broader social network platform but more as a "person" loaded with encyclopedic knowledge and musical prowess and functioning. Whether or not "he" or "she" can keep secrets is yet another question. It is beyond the scope of this article to discuss the media affordances and anthropomorphism of technologies, but note that that is another area of AI system ethics to be explored. For the engineer, the practice of privacy principles becomes simpler if the user's data need not be shared with other parties for purposes of functionality of the service.

But the case of private information sharing is important here for this analysis. Users might reveal certain information that is especially sensitive given the one-on-one nature of the technology. For instance, users might train the voice assistant to know when they wake, use the shower, or open the garage to leave home. As the AI systems integrated into voice assistive technologies use this information to help users become more efficient in their daily life, ethical AI should not share this sensitive information unnecessarily. Wherever possible, decentralized computing should be applied to minimize privacy harms and to protect individual rights. We would further assert that the development of voice assistive technologies can make use of the inherent dialogical nature of the technology itself. That is, if the humans using the technology are aware of the system's development and help to identify ethical issues within the system and its role in human life, the *core functions* of dialog and integrity are upheld. Additionally upheld are the *ethical perspectives* of utility, dialog, and human flourishing.

Looking to the future, we recommend further analysis of current AI systems at the societal and physical environmental levels, such as autonomous vehicles and smart cities in addition to the social media and voice assistive technology cases presented here. We selected our two examples to highlight the individual and the social worlds interacting via technology. We recognize that the personal and the social worlds are impossible to perfectly parse—both practically and psychologically—because individuals are a product of interaction with others and their social worlds. But they are also influenced by the larger environment: a total ecosystem of interactions of people, things, and culture. In proposing our AI Ethics Framework, we attempted to draw from historical and current ethical theories as well as widespread *core functions* important to upholding and supporting ethical human existence. We contend that while AI systems at personal, social, societal, and physical levels share common features, algorithms underlying the AI systems have some important differences. Specifically, risk analyses programmed into the algorithms for autonomous vehicles is of a different ethical complexity and nature compared to algorithms for social networking sites. However, using our AI Ethics Framework, we hope to demonstrate that the underlying ethics involved across contexts are relevantly similar; across many contexts is the need for transparency, equality and fairness, credibility, authenticity, and protection of civil liberties such as privacy, autonomy, and respect for persons.

AI systems take part in driving our social development and threaten to outpace our ability to understand its social impact. This presents a serious threat to human autonomy, self-determination, and democratic self-governance as well as to conceptualizations of a "good life."

The time to respond proactively is now. We presented here an Ethical Framework for developing AI

systems and policies that is ethically grounded in its *core functions* and an implementation process that is proactive, engaged, and for the good of humanity.

Our ultimate aim in the development of this framework is to create a communicative ethics tool to use for AI systems' construction and deployment that is ethical in its content as well as in the process of use. It reflects communicative ethical principles of human autonomy, fairness, social justice, and most importantly, a practice of communicative discourse ethics that can benefit individual consumers while improving AI systems. We suggest this framework may inform initiatives that govern AI's use and development, with a particular eye toward privacy protections, privacy awareness and choice, and most of all, self-determination in the ethical development and use of AI.

The AI Ethics Framework introduced in this article reflects the changing nature of society and social engagement that is demanding novel ethical thinking. Our design principles within the framework suggest ways to ensure building a future where humans and AI systems collaboratively and ethically engage. Our analysis illustrates that while some AI applications are already working toward better transparency, fairness, and inclusion, more can be done to protect privacy, enhance security, and engage dialogically. Further delving into this framework and investigation of additional examples will allow careful examination of the ethical pros and cons of AI systems' development trajectory and hopefully can have significant impact in the promotion of human flourishing. We view this framework as a first step and as evolving through dialogical communication with those working with AI systems—from both the user's and the developer's perspectives. More research will be needed to refine and further expand the principles, core functions, and contexts discussed in this article.

It is simply a matter of time before society faces challenges of the singularity. The more that can be done proactively and ethically to protect human dignity, autonomy, and security, the better chances there will be for humans to achieve the good life they both seek and deserve as a fundamental right. ■

References

1. H.J. Parkinson, "Wake Up! Amazon, Google, Apple and Facebook Are Running Our Lives," *The Guardian*, 12 May 2017; https://www.theguardian.com/commentisfree/2017/may/12/wake-up-amazon-google-apple-facebook-run-our-lives.
2. J. Falconer, "What Is the Technological Singularity?," *The Next Web*, 19 June 2011.
3. N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014.
4. J. Habermas, *Moral Consciousness and Communicative Action*, MIT Press, 1991.
5. D.B. Rasmussen, "Human Flourishing and the Appeal to Human Nature," *Social Philosophy and Policy*, vol. 16, no. 1, 1999, pp. 1–43.
6. T.W. Bynam, "Flourishing Ethics," *Ethics and Information Technology*, vol. 8, no. 4, 2006, pp. 157–173; https://link.springer.com/article/10.1007%2Fs10676-006-9107-1.
7. "Framework for Improving Critical Infrastructure Cybersecurity," draft version 1.1, National Institute of Standards and Technology, 10 Jan. 2017.
8. ACM Code of Ethics and Professional Conduct, ACM; https://www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct.
9. C. Newton. "It's Time for Facebook's News Feed to Explain Itself," *The Verge*, 26 Jan. 2018.
10. V. Goel, "Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry," *The New York Times*, 29 June 2014.

Dawn E. Schrader is an Associate Professor of Communication Ethics and Moral Psychology at Cornell University. She received her doctorate in human development from Harvard University. A research psychologist, she combines theory development in moral psychology with practical applications such as policy development and educational practices that enhance personal and professional ethical judgment, responsibility, integrity, and action. Schrader focuses on personal and social transformation through the creation of guidelines, climates, and systems that lead to more ethical and civil interactions and democratic cultural contexts in a changing technological world. Current research projects center on the psychology of privacy including privacy awareness and risks, women's online private information sharing and social aggression, and the ethics of technology and artificial intelligence systems on human life. She can be contacted at dawn.schrader@cornell.edu.

Dipayan Ghosh is a Fellow at New America and the Shorenstein Center at the Harvard Kennedy School. His research focuses on digital privacy and security and the impact of the new information age on society. Until recently, Ghosh worked on global privacy and public policy issues at Facebook where he led strategic efforts to address public policy and technical concerns around privacy, security, telecommunications policy, and ethical design of algorithms. He was a technology and economic policy advisor in the Obama White House. Ghosh received a PhD in electrical engineering and computer science at Cornell University and completed postdoctoral study at the University of California, Berkeley. He holds a bachelor's in electrical engineering from the University of Connecticut. He can be contacted at dipayan_ghosh@hks.harvard.edu.