CHAPTER 7

AI AND THE FUTURE

The changes wrought by advances in printing in fifteenth-century Europe offer a historical and philosophical comparison to the challenges of the age of AI. In medieval Europe, knowledge was esteemed but books were rare. Individual authors produced literature or encyclopedic compilations of facts, legends, and religious teachings. But these books were a treasure vouchsafed to a few. Most experience was lived, and most knowledge was transmitted orally.

In 1450, Johannes Gutenberg, a goldsmith in the German city of Mainz, used borrowed money to fund the creation of an experimental printing press. His effort barely succeeded—his business floundered, and his creditors sued—but by 1455, the Gutenberg Bible, Europe's first printed book, appeared. Ultimately, his printing press brought about a revolution that reverberated across every sphere of Western, and eventually global, life. By 1500, an estimated nine million printed books circulated in Europe, with the price of an individual book having plummeted. Not only was the Bible widely distributed in the languages of day-to-day life (rather than Latin), the works of classical authors in the fields of history, literature, grammar, and logic also began to proliferate.[1]

Before the advent of the printed book, medieval Europeans accessed knowledge primarily through community traditions—participating in harvesting and seasonal cycles, with their accumulation of folk wisdom; practicing faith and observing its sacraments at places of worship; joining a guild, learning its techniques, and being admitted to its specialized networks. When new information was acquired or new ideas arose (news from abroad, an innovative farming or mechanical invention, novel theological interpretations), it was transmitted either orally through a community or manually through hand-copied manuscripts.

As printed books became widely available, the relationship between individuals and knowledge changed. New information and ideas could spread quickly, through more varied channels. Individuals could seek out information useful to their specific endeavors and teach it to themselves. By examining source texts, they could probe accepted truths. Those with strong convictions and access to modest resources or a patron could publish their insights and interpretations. Advances in science and mathematics could be transmitted quickly, at continental scale. The exchange of pamphlets became an accepted method of political dispute, intertwined with theological dispute. New ideas spread, often either toppling or fundamentally reshaping established orders, leading to adaptations of religion (the Reformation), revolutions in politics (adjusting the concept of national sovereignty), and new understandings in the sciences (redefining the concept of reality).

Today, a new epoch beckons. In it, once again, technology will transform knowledge, discovery, communication, and individual thought. Artificial intelligence is not human. It does not hope, pray, or feel. Nor does it have awareness or reflective capabilities. It is a human creation, reflecting human-designed processes on human-created machines. Yet in some instances, at awesome scale and speed, it produces results approximating those that have, until now, only been reached through human reason. Sometimes, its results astound. As a result, it may reveal aspects of reality more dramatic than any we have ever contemplated. Individuals and societies that enlist AI as a partner to amplify skills or pursue ideas may be capable of feats—scientific, medical, military, political, and social—that eclipse those of preceding periods. Yet once machines approximating human intelligence are regarded as key to producing better and faster results, reason alone may come to seem archaic. After defining an epoch, the exercise of individual human reason may find its significance altered.

The printing revolution in fifteenth-century Europe produced new ideas and discourse, both disrupting and enriching established ways of life. The AI revolution stands to do something similar: access new information, produce major scientific and economic advances, and in so doing, transform the world. But its impact on discourse will be difficult to determine. By helping humanity navigate the sheer totality of digital information, AI will open unprecedented vistas of knowledge and understanding. Alternatively, its discovery of patterns in masses of data may produce a set of maxims that become accepted as orthodoxy across continental and global network platforms. This, in turn, may diminish humans' capacity for skeptical inquiry that has defined the current epoch. Further, it may channel certain societies and network-platform communities into separate and contradictory branches of reality.

AI may better or—if wrongly deployed—worsen humanity, but the mere fact of its existence challenges and, in some cases, transcends fundamental assumptions. Until now, humans alone developed their understanding of reality, a capacity that defined our place in the world and relationship to it. From this, we elaborated our philosophies, designed our governments and military strategies, and developed our moral precepts. Now AI has revealed that reality may be known in different ways, perhaps in more complex ways, than what has been understood by humans alone. At times, its achievements may be as striking and disorienting as those of the most influential human thinkers in their heydays—producing bolts of insight and challenges to established concepts, all of which demand a reckoning. Even more frequently, AI will be invisible, embedded in the mundane, subtly shaping our experiences in ways we find intuitively suitable.

We must recognize that AI's achievements, within its defined parameters, sometimes rank beside or even surpass those that human resources enable. We may comfort ourselves by repeating that AI is artificial, that it has not or cannot match our conscious experience of reality. But when we encounter some of AI's achievements—logical feats, technical breakthroughs, strategic insights, and sophisticated management of large, complex systems—it is evident that we are in the presence of another experience of reality by another sophisticated entity.

Accessed by AI, new horizons are opening before us. Previously, the limits of our minds constrained our ability to aggregate and analyze data, filter and process news and conversations, and interact socially in the digital domain. AI permits us to navigate these realms more effectively. It finds information and identifies trends that traditional algorithms could not—or at least not with equal grace and efficiency. In so doing, it not only expands physical reality but also permits expansion and organization of the burgeoning digital world.

Yet, at the same time, AI subtracts. It hastens dynamics that erode human reason as we have come to understand it: social media, which diminishes the space for reflection, and online searching, which decreases the impetus for conceptualization. Pre-AI algorithms were good at delivering "addictive" content to humans. AI is excellent at it. As deep reading and analysis contracts, so, too, do the traditional rewards for undertaking these processes. As the cost of opting out of the digital domain increases, its ability to affect human thought—to convince, to steer, to divert—grows. As a consequence, the individual human's role in reviewing, testing, and making sense of information diminishes. In its place, AI's role expands.

The Romantics asserted that human emotion was a valid and indeed important source of information. A subjective experience, they argued, was itself a form of truth. The postmoderns took the Romantics' logic a step further, questioning the very possibility of discerning an objective reality through the filter of subjective experience. AI will take the question considerably further, but with paradoxical results. It will scan deep patterns and disclose new objective facts—medical diagnoses, early signs of industrial or environmental disasters, looming security threats. Yet in the worlds of media, politics, discourse, and entertainment, AI will reshape information to conform to our preferences—potentially confirming and deepening biases and, in so doing, narrowing access to and agreement upon an objective truth. In the age of AI, then, human reason will find itself both augmented and diminished.

As AI is woven into the fabric of daily existence, expands that existence, and transforms it, humanity will have conflicting impulses. Confronted with technologies beyond the comprehension of the nonexpert, some may be tempted to treat AI's pronouncements as quasi-divine judgments. Such impulses, though misguided, do not lack sense. In a world where an intelligence beyond one's comprehension or control draws conclusions that are useful but alien, is it foolish to defer to its judgments? Spurred by this logic, a re-enchantment of the world may ensue, in which AIs are relied upon for oracular pronouncements to which some humans defer without question. Especially in the case of AGI (artificial general intelligence), individuals may perceive godlike intelligence—a superhuman way of knowing the world and intuiting its structures and possibilities.

But deference would erode the scope and scale of human reason and thus would likely elicit backlash. Just as some opt out of social media, limit screen time for children, and reject genetically modified foods, so, too, will some attempt to opt out of the "AI world" or limit their exposure to AI systems in order to preserve space for their reason. In liberal nations, such choices may be possible, at least at the level of the individual or the family. But they will not be without cost. Declining to use AI will mean not only opting out of conveniences such as automated movie recommendations and driving directions but also leaving behind vast domains of data, network platforms, and progress in fields from health care to finance.

At the civilizational level, forgoing AI will be infeasible. Leaders will have to confront the implications of the technology, for whose application they bear significant responsibility.

The need for an ethic that comprehends and even guides the AI age is paramount. But it cannot be entrusted to one discipline or field. The computer scientists and business leaders who are developing the technology, the military strategists who seek to deploy it, the political leaders who seek to shape it, and the philosophers and theologians who seek to probe its deeper meanings all see pieces of the picture. All should take part in an exchange of views not shaped by preconceptions.

At every turn, humanity will have three primary options: confining AI, partnering with it, or deferring to it. These choices will define AI's application to specific tasks or domains, reflecting philosophical as well as practical dimensions. For example, in airline and automotive emergencies, should an AI copilot defer to a human? Or the other way around? For each application, humans will have to chart a course; in some cases, the course will evolve, as AI capabilities and human protocols for testing AI's results also evolve. Sometimes deference will be appropriate —if an AI can spot breast cancer in a mammogram earlier and more accurately than a human can, then employing it will save lives. Sometimes partnership will be best, as in self-driving vehicles that function as today's airplane autopilots do. At other times, though—as in military contexts—strict, well-defined, well-understood limitations will be critical.

AI will transform our approach to what we know, how we know, and even what is knowable. The modern era has valued knowledge that human minds obtain through the collection and examination of data and the deduction of insights through observations. In this era, the ideal type of truth has been the singular, verifiable proposition provable through testing. But the AI era will elevate a concept of knowledge that is the result of partnership between humans and machines. Together, we (humans) will create and run (computer) algorithms that will examine more data more quickly, more systematically, and with a different logic than any human mind can. Sometimes, the result will be the revelation of properties of the world that were beyond our conception—until we cooperated with machines.

AI already transcends human perception—in a sense, through chronological compression or "time travel": enabled by algorithms and computing power, it analyzes and learns through processes that would take human minds decades or even centuries to complete. In other respects, time and computing power alone do not describe what AI does.

ARTIFICIAL GENERAL INTELLIGENCE

Are humans and AI approaching the same reality from different standpoints, with complementary strengths? Or do we perceive two different, partially overlapping realities: one that humans can elaborate through reason and another that AI can elaborate through algorithms? If this is the case, then AI perceives things that we do not and cannot—not merely because we do not have the time to reason our way to them, but also because they exist in a realm that our minds cannot conceptualize. The human quest to know the world fully will be transformed—with the haunting recognition that to achieve certain knowledge we may need to entrust AI to acquire it for us and report back. In either case, as AI pursues progressively fuller and broader objectives, it will increasingly appear to humans as a fellow "being" experiencing and knowing the world—a combination of tool, pet, and mind.

This puzzle will only deepen as researchers near or attain AGI. As we wrote in chapter 3, AGI will not be limited to learning and executing specific tasks; rather, by definition, AGI will be able to learn and execute a broad range of tasks, much like those humans perform. Developing AGI will require immense computing power, likely resulting in their being created by only a few well-funded organizations. Like current AI, though AGI may be readily distributable, given its capacities, its applications will need to be restricted. Limitations could be imposed by only allowing approved organizations to operate it. Then the questions will become: who controls AGI? Who grants access to it? Is democracy possible in a world in which a few "genius" machines are operated by a small number of organizations? What, under these circumstances, does partnership with AI look like?

If the advent of AGI occurs, it will be a signal intellectual, scientific, and strategic achievement. But it does not have to occur for AI to herald a revolution in human affairs.

AI's dynamism and capacity for emergent—in other words, unexpected—actions and solutions distinguish it from prior technologies. Unregulated and unmonitored, AIs could diverge from our expectations and, consequently, our intentions. The decision to confine, partner with, or defer to it will not be made by humans alone. In some cases, it will be dictated by AI itself; in others, by auxiliary forces. Humanity may engage in a race to the bottom. As AI automates processes, permits humans to probe vast bodies of data, and organizes and reorganizes the physical and social worlds, advantages may go to those who move first. Competition could compel deployment of AGI without adequate time to assess the risks—or in disregard of them.

An AI ethic is essential. Each individual decision—to constrain, partner, or defer—may or may not have dramatic consequences, but in the aggregate, they will be magnified. They cannot be made in isolation. If humanity is to shape the future, it needs to agree on common principles that guide each choice. Collective action will be hard, and at times impossible, to achieve, but individual actions, with no common ethic to guide them, will only magnify instability.

Those who design, train, and partner with AI will be able to achieve objectives on a scale and level of complexity that, until now, have eluded humanity—new scientific breakthroughs, new economic efficiencies, new forms of security, and new dimensions of social monitoring and control. Those who do not have such agency in the process of expanding AI and its uses may come to feel that they are being watched, studied, and acted upon by something they do not understand and did not design or choose—a force that operates with an opacity that in many societies is not tolerated of conventional human actors or institutions. The designers and deployers of AI should be prepared to address these concerns—above all, by explaining to non-technologists what AI is doing, as well as what it "knows" and how.

AI's dynamic and emergent qualities generate ambiguity in at least two respects. First, AI may operate as we expect but generate results that we do not foresee. With those results, it may carry humanity to places its creators did not anticipate. Much like the statesmen of 1914 failed to recognize that the old logic of military mobilization, combined with new technology, would pull Europe into war, deploying AI without careful consideration may have grave consequences. These may be localized, such as a self-driving car that makes a life-threatening decision, or momentous, such as a significant military conflict. Second, in some applications, AI may be unpredictable, with its actions coming as complete surprises. Consider AlphaZero, which, in response to the instruction "win at chess," developed a style of play that, in the millennia-long history of the game, humans had never conceived. While humans may carefully specify AI's objectives, as we give it broader latitude, the paths AI takes to accomplish its objectives may come to surprise or even alarm us.

Accordingly, AI's objectives and authorizations need to be designed with care, especially in fields in which its decisions could be lethal. AI should not be treated as automatic. Neither should it be permitted to take irrevocable actions without human supervision, monitoring, or direct control. Created by humans, AI should be overseen by humans. But in our time, one of AI's challenges is that the skills and resources required to create it are not inevitably paired with the philosophical perspective to understand its broader implications. Many of its creators are concerned primarily with the applications they seek to enable and the problems they seek to solve: they may not pause to consider whether the solution might produce a revolution of historic proportions or how their technology may affect various groups of people. The AI age needs its own Descartes, its own Kant, to explain what is being created and what it will mean for humanity.

Reasoned discussion and negotiation involving governments, universities, and private-sector innovators should aim to establish limits on practical actions—like the ones that govern the actions of people and organizations today. AI shares attributes of some regulated products, services, technologies, and entities, but it is distinct from them in vital ways, lacking its own fully defined conceptual and legal framework. For example, AI's evolving and emergent properties pose regulatory challenges: what and how it operates in the world may vary across fields and evolve over time—and not always in predictable ways. The governance of people is guided by an ethic. AI begs for an ethic of its own—one that reflects not only the technology's nature, but also the challenges posed by it.

Frequently, existing principles will not apply. In the age of faith, courts determined guilt during ordeals in which the accused faced trial by combat and God was believed to dictate victory. In the age of reason, humanity assigned guilt according to the precepts of reason, determining culpability and meting out punishment consistent with notions such as causality and intention. But AIs do not operate by human reason, nor do they have human motivation, intent, or self-reflection. Accordingly, their introduction complicates existing principles of justice being applied to humans. When an autonomous system operating on the basis of its own perceptions and decisions acts, does its creator bear responsibility? Or does the fact that the AI acted sever it from its creator, at least in terms of culpability? If AI is enlisted to monitor signs of criminal wrongdoing, or to assist in judgments of innocence and guilt, must the AI be able to "explain" how it reached its conclusions in order for human officials to adopt them?

At what point and in what contexts in the technology's evolution it should be subject to internationally negotiated restrictions is another essential subject of debate. If attempted too early, the technology may be stymied, or there may be incentives to conceal its capabilities; if delayed too long, it may have damaging consequences, particularly in military contexts. The challenge is compounded by the difficulty of designing effective verification regimes for a technology that is ethereal, opaque, and easily distributed. Official negotiators will inevitably be governments. But forums need to be created for technologists, ethicists, the corporations creating and operating AIs, and others beyond these fields.

For societies, the dilemmas AI raises are profound. Much of our social and political life now transpires on network platforms enabled by AI. This is especially the case for democracies, which depend upon these information spaces for the debate and discourse that form public opinion and confer legitimacy. Who or what institutions should define the technology's role? Who should regulate it? What roles should be played by the individuals who use AI? The corporations that produce it? The governments of the societies that deploy it? As part of addressing such questions, we should seek ways to make it auditable—that is, to make its processes and conclusions both checkable and correctable. In turn, formulating corrections will depend upon the elaboration of principles responsive to AI's forms of perception and decision making. Morality, volition, even causality do not map neatly onto a world of autonomous AIs. Versions of such questions arise for most other elements of society, from transportation to finance to medicine.

Consider AI's impact on social media. Through recent innovations, these platforms have rapidly come to host vital aspects of our communal lives. Twitter and Facebook highlighting, limiting, or outright banning content or individuals—all functions that, as we discussed in chapter 4, depend on AI—are testaments to their power. In particular, democratic nations will be increasingly challenged by the use of AI in the unilateral, often opaque promotion or removal of content and concepts. Will it be possible to retain our agency as our social and political lives increasingly shift into domains curated by AI, domains that we can only navigate through reliance upon that curation?

With the use of AIs to navigate masses of information comes the challenge of distortion—of AIs promoting the world humans instinctually prefer. In this domain, our cognitive biases, which AIs can readily magnify, echo. And with those reverberations, with that multiplicity of choice coupled with the power to select and screen, misinformation proliferates. Social media companies do not run news feeds to promote extreme and violent political polarization. But it is self-evident that these services have not resulted in the maximization of enlightened discourse.

AI, FREE INFORMATION, AND INDEPENDENT THOUGHT

What, then, should our relationship with AI be? Should it be cabined, empowered, or a partner in governing these spaces? That the distribution of certain information—and, even more so, deliberate disinformation—can damage, divide, and incite is beyond dispute. Some limits are needed. Yet the alacrity with which harmful information is now decried, combated, and suppressed should also prompt reflection. In a free society, the definitions of *harmful* and *disinformation* should not be the purview of corporations alone. But if they are entrusted to a government panel or agency, that body should operate according to defined public standards and through verifiable processes in order not to be subject to exploitation by those in power. If they are entrusted to an AI algorithm, the objective function, learning, decisions, and actions of that algorithm must be clear and subject to external review and at least some form of human appeal.

Naturally, the answers will vary across societies. Some may emphasize free speech, possibly differently based on their relative understandings of individual expression, and may thus limit AI's role in moderating content. Each society will choose what it values, perhaps resulting in complex relations with operators of transnational network platforms. AI is porous—it learns from humans, even as we design and shape it. Thus not only will each society's choices vary, so, too, will each society's relationship with AI, its perception of AI, and the patterns that its AIs imitate and learn from human teachers. Nevertheless, the quest for facts and truth should not lead societies to experience life through a filter whose contours are undisclosed and untestable. The spontaneous experience of reality, in all its contradiction and complexity, is an important aspect of the human condition—even when it leads to inefficiency or error.

AI AND INTERNATIONAL ORDER

Globally, myriad questions demand answers. How can AI network platforms be regulated without inciting tensions among countries concerned about their security implications? Will such network platforms erode traditional concepts of state sovereignty? Will the resulting changes impose a polarity on the world not known since the collapse of the Soviet Union? Will small nations object? Will efforts to mediate such consequences succeed, or have any hope of success at all?

As AI's capabilities continue to increase, defining humanity's role in partnership with it will be ever more important and complicated. One can contemplate a world in which humans defer to AI to an ever-greater degree over issues of ever-increasing magnitude. In a world in which an opponent successfully deploys AI, could leaders defending against it responsibly decide not to deploy their own, even if they were unsure what evolution that deployment would portend? And if the AI possessed a superior ability to recommended a course of action, could policy makers reasonably refuse, even if the course of action entailed sacrifice of some magnitude? For what human could know whether the sacrifice was essential to victory? And if it was, would the policy maker truly wish to gainsay it? In other words, we may have no choice but to foster AI. But we also have a duty to shape it in a way that is compatible with a human future.

Imperfection is one of the most enduring aspects of human experience, especially of leadership. Often, policy makers are distracted by parochial concerns. Sometimes, they act on the basis of faulty assumptions. Other times, they act out of pure emotion. Still other times, ideology warps their vision. Whatever strategies emerge to structure the human-AI partnership, they must accommodate. If AI displays superhuman capabilities in some areas, their use must be assimilable into imperfect human contexts.

In the security realm, AI-enabled systems will be so responsive that adversaries may attempt to attack before the systems are operational. The result may be an inherently destabilizing situation, comparable to the one created by nuclear weapons. Yet nuclear weapons are situated in an international framework of security and arms-control concepts developed over decades by governments, scientists, strategists, and ethicists, subject to refinement, debate, and negotiation. AI and cyber weapons have no comparable framework. Indeed, governments may be reluctant to acknowledge their existence. Nations—and probably technology companies—need to agree on how they will coexist with weaponized AI.

The diffusion of AI through governments' defense functions will alter international equilibrium and the calculations that have largely sustained it in our era. Nuclear weapons are costly and, because of their size and structure, difficult to conceal. AI, on the other hand, runs on widely available computers. Because of the expertise and computing resources needed to train machine-learning models, creating an AI requires the resources of large companies or nation-states. Because the application of AIs is conducted on relatively small computers, AI will be broadly available, including in ways not intended. Will AI-enabled weapons ultimately be available to anyone with a laptop, a connection to the internet, and an ability to navigate its dark elements? Will governments empower loosely affiliated or unaffiliated actors to use AI to harass their opponents? Will terrorists engineer AI attacks? Will they be able to (falsely) attribute them to states or other actors?

Diplomacy, which used to be conducted in an organized, predictable arena, will have vast ranges of both information and operation. The previously sharp lines drawn by geography and language will continue to dissolve. AI translators will facilitate speech, uninsulated by the tempering effect of the cultural familiarity that comes with linguistic study. AI-enabled network platforms will promote communication across borders. Moreover, hacking and disinformation will continue to distort perception and evaluation. As complexity increases, the formulation of implementable agreements with predictable outcomes will grow more difficult.

The grafting of AI functionality onto cyber weapons deepens this dilemma. Humanity sidestepped the nuclear paradox by sharply distinguishing between conventional forces—deemed reconcilable with traditional strategy—and nuclear weapons, deemed exceptional. Where nuclear weapons applied force bluntly, conventional forces were discriminating. But cyber weapons, which are capable of both discrimination and massive destruction, erase this barrier. As AI is mapped onto them, these weapons become more unpredictable and potentially more destructive. Simultaneously, as they move through networks, these weapons defy attribution. They also defy detection—unlike nuclear weapons, they may be carried on thumb drives—and facilitate diffusion. And in some forms, they can, once deployed, be difficult to control, particularly given AI's dynamic and emergent nature.

This situation challenges the premise of a rules-based world order. Additionally, it gives rise to an imperative: to develop a concept of arms control for AI. In the age of AI, deterrence will not operate from historical precepts; it will not be able to. At the beginning of the nuclear age, the verities developed in discussions between leading professors (who had government experience) at Harvard, MIT, and Caltech led to a conceptual framework for nuclear arms control that, in turn, contributed to a regime (and, in the United States and other countries, agencies to implement it). While the academics' thinking was important, it was conducted separately from the Pentagon's thinking about conventional war—it was an addition, not a modification. But the potential military uses of AI are broader than those of nuclear arms, and the divisions between offense and defense are, at least currently, unclear.

In a world of such complexity and inherent incalculability, where AIs introduce another possible source of misperception and mistake, sooner or later, the great powers that possess high-tech capabilities will have to undertake a permanent dialogue. Such dialogue should be focused on the fundamental: averting catastrophe and, in so doing, surviving.

AI and other emerging technologies (such as quantum computing) seem to be moving humans closer to knowing reality beyond the confines of our own perception. Ultimately, however, we may find that even these technologies have limits. Our problem is that we have not yet grasped their philosophical implications. We are being advanced by them, but automatically rather than consciously. The last time human consciousness was changed significantly—the Enlightenment—the transformation occurred because new technology engendered new philosophical insights, which, in turn, were spread by the technology (in the form of the printing press). In our period, new technology has been developed, but remains in need of a guiding philosophy.

AI is a grand undertaking with profound potential benefits. Humans are developing it, but will we employ it to make our lives better or to make our lives worse? It promises stronger medicines, more efficient and more equitable health care, more sustainable environmental practices, and other advances. Simultaneously, however, it has the capability to distort or, at the very least, compound the complexity of the consumption of information and the identification of truth, leading some people to let their capacities for independent reason and judgment atrophy.

Other countries have made AI a national project. The United States has not yet, as a nation, systematically explored its scope, studied its implications, or begun the process of reconciling with it. The United States must make all these projects national priorities. This process will require people with deep experience in various domains to work together—a process that would greatly benefit from, and perhaps require, the leadership of a small group of respected figures from the highest levels of government, business, and academia.

Such a group or commission should have at least two functions:

1. Nationally, it should ensure that the country remains intellectually and strategically competitive in AI.
2. Both nationally and globally, it should study, and raise awareness of, the cultural implications AI produces.

In addition, the group should be prepared to engage with existing national and subnational groups.

We write in the midst of a great endeavor that encompasses all human civilizations—indeed, the entire human species. Its initiators did not necessarily conceive of it as such; their motivation was to solve problems, not to ponder or reshape the human condition. Technology, strategy, and philosophy need to be brought into some alignment, lest one outstrip the others. What about traditional society should we guard? And what about traditional society should we risk in order to achieve a superior one? How can AI's emergent qualities be integrated into traditional concepts of societal norms and international equilibrium? What other questions should we seek to answer when, for the situation in which we find ourselves, we have no experience or intuition?

Finally, one "meta" question looms: can the need for philosophy be met by humans *assisted* by AIs, which interpret and thus understand the world differently? Is our destiny one in which humans do not completely understand machines, but make peace with them and, in so doing, change the world?

Immanuel Kant opened the preface to his *Critique of Pure Reason* with an observation:

> Human reason has the peculiar fate in one species of its cognitions that it is burdened with questions which it cannot dismiss, since they are given to it as problems by the nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason.[2]

In the centuries since, humanity has probed deeply into these questions, some of which concern the nature of the mind, reason, and reality itself. And humanity has made great breakthroughs. It has also encountered many of the limitations Kant posited—a realm of questions it cannot answer, of facts it cannot know fully.

The advent of AI, with its capacity to learn and process information in ways that human reason alone cannot, may yield progress on questions that have proven beyond our capacity to answer. But success will produce new questions, some of which we have attempted to articulate in this book. Human intelligence and artificial intelligence are meeting, being applied to pursuits on national, continental, and even global scales. Understanding this transition, and developing a guiding ethic for it, will require commitment and insight from many elements of society: scientists and strategists, statesmen and philosophers, clerics and CEOs.

This commitment must be made within nations and among them. Now is the time to define both our partnership with artificial intelligence and the reality that will result.