BUDDHISM AND INTELLIGENT TECHNOLOGY

TOWARD A MORE HUMANE FUTURE

PETER D. HERSHOCK

BLOOMSBURY

BLOOMSBURY ACADEMIC
Bloomsbury Publishing Plc
50 Bedford Square, London, WC1B 3DP, UK
1385 Broadway, New York, NY 10018, USA
29 Earlsfort Terrace, Dublin 2, Ireland

BLOOMSBURY, BLOOMSBURY ACADEMIC and the Diana logo are trademarks of Bloomsbury Publishing Plc

First published in Great Britain 2021

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

ISBN: HB: 978-1-3501-8226-4
PB: 978-1-3501-8227-1
ePDF: 978-1-3501-8228-8
eBook: 978-1-3501-8229-5

Typeset by Newgen KnowledgeWorks Pvt. Ltd., Chennai, India

To find out more about our authors and books visit www.bloomsbury.com and sign up for our newsletters.

Contents

Acknowledgments

The initial research for this book was supported by a nine-month fellowship with the Berggruen Institute. Through the fellowship I was able to spend six months in China and benefitted greatly from a number of the programs organized by the Berggruen Institute's China Center in Qufu, Beijing, and Shanghai. Conversations with Professor Liu Zhe at Peking University helped greatly in seeing Chinese policy on artificial intelligence from a thoughtfully enthusiastic and critical insider's perspective. Conversations with Liu Dele afforded insights not only into the entrepreneurial world in China but also into the spread of Buddhist practice in an increasingly digitalized Chinese society. Professor Roger T. Ames, also at Peking University and emeritus at the University of Hawai`i, was as always an inspiring embodiment of Confucian commitments to harmonious relationality. Finally, much gratitude is due to the Yun Hwa Denomination of World Social Buddhism and its founder, Ji Kwang Dae Poep Sa Nim, without whose exemplification of superlatively embodied compassion and clarity the presentation of Buddhism in this book would not have been possible.

Introduction

We live in revolutionary times. Exponential increases in the power and scope of artificial intelligence (AI) are being realized through the combination of big data and self-improving machine learning algorithms, and the technological and industrial promise is seemingly limitless. With good reason, many people are heralding this as the onset of the Fourth Industrial Revolution. Yet, this new revolution is as metaphysical as it is industrial. Like the Copernican Revolution, which radically decentered humanity in the cosmos five centuries ago, the data- and attention-driven Intelligence Revolution is dissolving once-foundational certainties and opening entirely new realms of opportunity.

Smart cities will be both more efficient and more livable. Smart healthcare has the potential to reach and benefit the half of humanity that now lacks even basic health services. Smart schools will respond to individual student needs, drawing on data generated by millions of teachers and learners worldwide. Yet, it is estimated that over the next twenty years as many as half of all core job tasks are liable to being taken over by artificial agents, even in such white collar professions as law and medicine. The digital economy is proving to be structurally biased toward monopoly. The infrastructure that enables continuous social media connectivity also supports corporate data mining and state surveillance. And smart services and the algorithmic tailoring of experience have potentials for dramatically transforming the meanings of family, friendship, health, work, security, and agency, at first supplementing and then eventually supplanting intelligent human practices. The immense technological promises of the Intelligence Revolution are kin to equally immense perils. This book is an attempt to bring into clearer focus the complex interplay of the technological, societal, and personal dimensions of the Intelligence Revolution and to encourage sustained intercultural and intergenerational ethical collaboration in figuring out how best to align AI with globally shared human values.

In recent years, considerable attention has been given to both the welcome and the worrying prospects of realizing artificial general intelligence or superintelligence. For techno-idealists, the advent of artificial general intelligence will mark an evolutionary leap from carbon to silicon

and will bring an exponential expansion of intelligence and the freeing of consciousness from the constraints of biology. Techno-realists see, instead, potentials for an exponential scaling up of unintended consequences and risks, including threats to the continued existence of humanity. In both cases, attention is focused on the possibility of a *technological singularity*: a historical juncture at which technology acquires infinite value and artificial agents are free either to act upon or ignore human aims and interests.

Estimates regarding the creation of artificial general intelligence vary widely, from a mere quarter century to several centuries, and our chance of arriving at a technological singularity is thus very much a matter of speculation. What is not a matter of speculation—and what should concern us no less profoundly and much more immediately—is the *ethical singularity* toward which we are being hastened by machine agents acting on the intentions of their human creators: a point at which evaluating competing value systems and conceptions of humane intelligence take on infinite value/significance.

Ethics can be variously defined. But, at a minimum, it involves going beyond using our collective human intelligence to more effectively reach existing aims and using it instead to discriminate qualitatively among both our aims and our means for realizing them. In short, ethics is the art of human course correction. The ethical singularity ahead is the point at which the opportunity space for further human course correction collapses—a historical chokepoint at which we will have no greater chance of escaping the consequences of scaling up our often conflicting values than light has of escaping a cosmological black hole.

Navigating through and beyond this ethical singularity toward more humane and equitable futures will require exercising new kinds of collaborative ethical agency. For the first time, the shape of the human-technology-world relationship is not being wrought solely by human hands. Technology is no longer a passive medium through which humanity rearticulates what it means to be human and redefines the world of human experience. Technology is now intelligent: an active and adaptive participant in the human-technology-world relationship. Machine autonomy today is nowhere close to being on par with human autonomy, and may never be so. But machine agency is evolving with ever-increasing speed and it is doing so in ways that often cannot be fully explained. The ethical challenges posed by this new relationship are unprecedented.

Engineering machine evolution can be intoxicating. At gatherings of the scientists, technologists, engineers, and entrepreneurs working at the cutting edges of AI research and development, the excitement is almost palpably radiant. These are brilliantly curious people committed to answering the

intellectual and practical challenges of opening, exploring, and exploiting vastly new realms of human and technological possibility. They are visionaries engaged in turning dreams into realities. For them, bringing ethics to bear on the development of AI is often understood as a matter of computationally achieving acceptable levels of transparency and accountability, assuring compliance with existing laws and regulatory frameworks, and ensuring that machine intelligences demonstrate operational/behavioral respect for accepted societal norms. Discussions tend to focus on developing the tools and techniques needed to achieve these goals and on how best to carry on with the exciting technical and scientific work ahead.

At gatherings of more skeptically inclined philosophers, legal scholars, and social scientists, the atmosphere is much more sober and energies are often directed toward articulating concerns about the moral hazards of adventuring in the realms of possibility that are being opened by the new intelligence industries. In keeping with the principle that is it better to be safe than sorry, these historically cognizant, and often profoundly thoughtful people would have us attend to actual and potential perils concealed in the shadows of technological promise. The most cautious among them urge immediate and rigorous consideration of the existential threat that the advent of artificial superintelligence might pose for humanity. In these gatherings, bringing ethics to bear on the development of AI and machine agencies begins with acknowledging that, good intentions notwithstanding, our technological achievements have often affected the human experience in ways that are malign as well as benign. Discussions tend to focus on preemptive caution, on strategies for slowing the pace of technological change so that we don't find ourselves spun out of control, and on establishing protocols for ensuring that AI, machine learning, and big data are developed and deployed in ways that are aligned with human values and are conducive to both personal and societal well-being.

Both kinds of discussions are necessary and valuable. But we will need something more if we are going to resolve the global predicament into which the Intelligence Revolution is conveying us. If there is one great lesson to be learned from the history of the human-technology-world relationship, it is that new technologies have both anticipated and unanticipated consequences and it is the unintended, ironic consequences which most dramatically alter the human and world dimensions of the relationship. If there is a parallel lesson to be learned from the history of ethics, it is that even the most ardently reasoned claims about universal human values and our options for course correction are historically and culturally conditioned. Significant gulfs already exist, for example, among the approaches to "ethical" or "human-centered" AI that are being taken in the United States, China, and the European Union,

including sharply opposing perspectives on the importance of data privacy and net neutrality, and on how to weight individual, corporate, and state interests in efforts to align AI with human values and societal well-being. The path toward global ethical consensus on intelligent technology is far from apparent.

It is tempting to assume that the promises and perils of now-emerging intelligence technologies and industries can be clearly distinguished, and to conclude that courses of action enabling us to avoid the latter and happily realize the former can be readily identified in advance. Unfortunately, there is no factual warrant for this assumption. On the contrary, the greatest perils of intelligent technology do not appear to be extrinsic to its great promises but rather intrinsic to them. The rise of autonomous vehicles will lower risks of traffic accidents and reduce insurance burdens. But, in doing so, it will also curtail the employment of heavy equipment operators and truck and taxi drivers. Virtual personal assistants will provide smart services drawing on knowledge resources far greater and far more quickly than any human ever could but will at the same time subject their users to profit-generating surveillance and contribute to the atrophy of an increasingly wide range of intelligent human practices.

Thinking of such perils as mere side-effects of realizing the promises of the Intelligence Revolution is both practically and ethically misleading. Side-effects are contingencies that can be controlled for and, in most cases, bypassed on our way to achieving desired ends. They are obstacles that can be gotten around by means of new techniques or practices that enable us to continue forward without making any significant changes in our motivating values, aims, and interests. But in much the same way that mass-produced manufactured goods and climate change are equally primary products of industrialization based on fossil fuel burning when conducted at global scales, the promises and perils of the Intelligence Revolution are analogously primary products of global intelligence industries aimed (as they currently are) at reaping the benefits of machine learning and big data.

The analogy is instructive. Climate change is not a *problem* awaiting technical solution. It a *predicament* that makes evident conflicts within and among prevailing constellations of social, cultural, economic, and political values. Our collective failure to generate the commitments and collaborative action needed to alleviate and eliminate climate change impacts is thus not a technical failure; it is an ethical failure.

Unlike problems, predicaments cannot be solved for the simple reason that the value conflicts that they express make it impossible to determine exactly what would count *as* a solution. Climate change could easily be slowed or stopped by dramatically cutting back and eventually eliminating

carbon emissions. What prevents this are conflicts among environmental values and those that undergird desires for rapid economic growth, political stability, and cultural continuity. Predicaments can only be resolved, where resolution implies both clarity and commitment—*clarity* about how an experienced value conflict has come to be and *commitment* to evaluating and appropriately reconfiguring relevant constellations of values, aims, and practices. Predicament resolution is thus inherently reflexive. It involves changing not only *how* we live but both *why* and *as whom*.

Much like climate change, the Intelligence Revolution is forcing confrontation with the value conflicts that underlie its pairing of benign and malign consequences. But the predicament into which humanity is being enticed by the Intelligence Revolution is particularly complex. Systems of AI are functioning as agents of experiential and relational transformation, actively altering the humanity-technology-world relationship in accord with computational directives to reinforce the readiness of ever more precisely desire-defined individuals to accept ever-greater connective convenience and choice in exchange for granting corporations and states new and evolving powers of control. Historically unprecedented, these systems operate according to a new logic of domination that is not expressed in overt acts of *coercion* but through soliciting voluntary membership in network-enabled regimes of ambiently reinforced *craving*. It is freely spent human attention energy and the data—the traces of human intelligence—that are carried along with it which are being used by competing corporate and state actors to build smart societies, to individually tailor human experience and behavior, and to incentivize reliance on smart services in ways that have the potential to render human intelligence superfluous.

The value tensions made evident by the Intelligence Revolution are at one level intimately personal—tensions between connective freedom and privacy, for example. But they are also political. The Second Industrial Revolution, from the 1870s to the 1940s, was a key driver for scaling up imperial and colonial ambitions in the so-called Great Game played for global control over land and labor. Similarly, the Intelligence Revolution—the Fourth Industrial Revolution—is driving a New Great Game: a competition among commercial, state, and nonstate actors seeking digital control over human attention and, ultimately, dominance in nothing less than the colonization of consciousness itself.

The technologies of the first three industrial revolutions—epitomized by the steam engine, scientific mass production, and digitalized computing and communications—scaled up human intentions for over two hundred years before it became evident that continued industrial uses of fossil fuels would irreversibly alter the planetary climate system. Our inability to resolve the

climate predicament and halt or reverse climate impacts over the fifty years since then give us little cause for optimism in the face of the predicament posed by intelligent technology. The Fourth Industrial Revolution is likely to scale up human intentions and value conflicts to the point of irreversible changes in human presence and the dynamics of the anthrosphere over a matter of one or two decades, not one or two centuries.

Given the disparate approaches to "ethical" or "human-centered" AI that are being taken, for example, in the United States, the European Union and China, it's tempting to frame the New Great Game as a competition among different national or regional visions of the "good life" in a "smart society." But this way of seeing things directs critical concern exclusively toward which player(s) could or should win, rather than toward the human/societal impacts of playing the game, regardless of who ends up "winning." Geopolitics is important. Concerns about an AI "arms race" are very real. And the process of establishing global frameworks for ethical AI will undoubtedly be one of craftily negotiated steps forward, backward, sideways, and then forward again. But, framing the New Great Game as a two-dimensional conflict between, for instance, a choice-valuing "West" and a control-valuing "East" cannot do justice to either its geopolitical or its ethical complexity.

Contrary to our historically grounded intuitions, the predicament at the heart of the Intelligence Revolution is inseparable from the fact that the new digital systems of domination are becoming ever more intimately personal the more global their structures of data gathering and machine learning become. This fusion of the personal and the global is not incidental. The Intelligence Revolution would have been impossible without the connectivity explosion that began with the smartphone and global wireless internet access, and that is now being hypercharged by exponential growth in the internet of things. It was the planet-engulfing flood of data released by 24/7 connectivity through portable and increasingly miniaturized wireless devices that made it possible for machine learning to become something other than a tech-lab curiosity. Informed by ever greater volumes, varieties, and velocities of data flows, including data about our day-to-day lives, our loves and our longings, machine learning capabilities have evolved with breathtaking speed, penetrating and increasingly pervading virtually every domain of human endeavor. The personal and the global are being brought into algorithmic communion.

In addition, with nearly seamless worldwide connectivity and a mass transfer of social energy from physical to virtual spaces of interaction, not only is the divide between the private and the public eroding, the personal is becoming nearly indistinguishable from the interpersonal. At the same

time that the geographical coordinates of our bodily presences are becoming globally accessible, our social, political, and economic presences are becoming increasingly delocalized. The Cartesian conviction that "I think, therefore I am" is giving way to a new metaphysics of personal presence summed up in the realization that "as we connect, so we are."

The power implications of this shift are profound. It is already apparent that the search and recommendation algorithms that mediate global connectivity are wielding immense epistemic power, actively shaping informational currents and thus what we encounter and think about and know. The resulting concentration of economic power in key platform and social media providers over the last fifteen years has been nothing short of stunning. Yet, the power to shape connectivity is also ontological power—the power to shape both who we are and who we become. Recognizing this is the first step toward engaging the Intelligence Revolution, not as a technological inevitability but as an ethical challenge and opportunity. Central to the task of meeting this challenge will be envisioning who we need to be *present as* if we are going participate in the Intelligence Revolution as an opportunity for recentering humanity, improvising together in shared commitment to equitable and truly humane futures.

Personal Presence, Ethics, and Global Action

Asking who we need to be *present as* in order to resolve the predicament of AI might be interpreted as a question about how to achieve and sustain individual well-being in a time of dramatic technological change. That, however, is not what I have in mind. Self-help guides to life in a digitally intelligent world may be of great value individually. But the work of establishing more equitable and truly humane trajectories for the Intelligence Revolution will not be accomplished by simply going into individual retreat. Exercising our "exit rights" from social media, online shopping, and smart services might be necessary—at least temporarily—to be able to engage in this complex work. But it will not be sufficient. Opting out completely from the dynamics of network connectivity is, in effect, to opt out of being in a position to inflect or redirect those dynamics. Avoiding a predicament is not a viable means to resolving it.

Alternatively, emphasizing who we need to be *present as* to respond effectively to the ethical singularity toward which we are being impelled by the Intelligence Revolution might suggest a preemptive commitment to virtue ethics—an emphasis on cultivating the character traits needed to lead ethically sound lives and to deliberate effectively about what it might mean to

engineer humane systems of artificial agency. Appeals to virtue ethics have, in fact, formed one of the major currents in recent attempts to frame ethics policies for research and development in AI and robotics. However, as will be discussed in Chapter Five, there are very good reasons to avoid relying exclusively on any single ethical tradition. Moreover, as resources for truly global predicament resolution, systems of virtue ethics have the liability of being built around culturally specific and often conservative commitments to virtues or character traits that are very much historically conditioned. The ethical deliberation involved in truly global predicament resolution must be both intercultural and improvisational.

A New Metaphysics: The Centrality of Interdependence

The aim of focusing on qualities of personal presence is to prepare ourselves for the personal challenges of intercultural and improvisational ethical deliberation. But the underlying reasons for stressing qualities of personal presence have to do with what we might call the everyday metaphysics and economics of the Intelligence Revolution, and in particular their erasure of clear boundaries between the global and the personal, between the public and the private, and between control and choice.

As was just intimated, one of the guiding premises of this book is that the most pressing challenges of the present and coming decades—climate change and the transformations associated with the Intelligence Revolution central among them, but including as well the degradation of both natural and urban environments; the persistence of global hunger; and rising inequalities of wealth, income, risk, and opportunity—cannot be responded to adequately and resolved sustainably unless we are able to grapple with and resolve the complex value conflicts that they objectively express. In the terms used earlier, we are in the midst of an era-defining shift from the dominance of problem solution to that of predicament resolution: a shift from the primacy of the technical to that of the ethical.

The problem-to-predicament shift, however, is also evidence of the end of an era in which changes and challenges could be met and managed in the context of a world that remained essentially familiar. Predicament resolution is not a matter of simply choosing among ends and interests like destinations listed on an airline departures board. It involves improvising holistic and ultimately irreversible shifts of existential direction: a process of responsive and responsible self-transformation that, because it alters the totality of our relational possibilities, is always also world-altering. Put in the language of social science, we are in the midst of an era-defining transmutation of the interplay of structure and agency.

As earlier alluded, this ongoing transmutation of the world is no less dramatic than that brought about by the Copernican Revolution some five hundred years ago. Copernicus was confronted with evidence that something was profoundly and fundamentally amiss in prevailing assumptions about the structure of the cosmos. There seemed no way to reconcile observations of the movements of celestial bodies like the stars, planets, and the Moon with the commonsense and apparently divinely ordained belief that the Earth was at the center of the universe. To realize, as Copernicus did, that the Earth orbited the Sun, rather than the other way around, meant realizing that our everyday experience of the Sun rising in the east and setting in the west is simply an illusion of perspective. This realization was as much metaphysical as it was physical, and it shattered a host of previously foundational certainties. Most immediately, these were intellectual and religious. But over time, the casualties also came to include long-standing social, political, and economic certainties. In this spreading absence of certainty, there opened entirely new spaces of opportunity.

Today, instead of a decentering of the Earth, we are undergoing a decentering of the individual: the dissolution of once-evident certainties about the primacy of personal independence and national sovereignty. It is becoming increasingly—and at times, painfully—apparent that continued belief in the primacy of the individual contributes to our being incapable of navigating the ethical straits we must pass through in order to resolve the global predicaments we face. Although we may persist in taking individual persons, religions, cultures, corporations, or nations as our points of reference and continue seeing global dynamics as revolving around them, we have strong evidence now that the centrality of the individual is an illusion. It is not the individual that is basic but the relational.

Just as daybreaks continue to appear to us as sun rising events rather than as Earth-turnings, centuries after Copernicus proved otherwise, most of us will "naturally" continue to experience ourselves as individuals. Hopefully, however, we will be as successful in acting on our new knowledge of the primacy of relationality as we have been in acting on our knowledge about the movements of celestial bodies. Acting on the knowledge that the Earth revolves around the Sun, we have sent probes to all of the solar system's planets and have even landed a robot surveyor on an asteroid the size of a few city blocks in an orbit between that of Mars and Jupiter. These are astonishing feats. Through the decentering of the individual, we are being introduced not only to entirely new kinds of challenges but also to entirely new kinds of opportunities. Once we begin acting consistently on our knowledge that the existence of individuals is a perspectival illusion, we will gain similarly astonishing new capabilities.

A New Economics: The Value of Attention

As long as the economics of the Intelligence Revolution remain on their current course, however, realizing these new capabilities equitably will not be easy. Although we continue to speak about a global information economy, big data has made this obsolete. Information is too cheap and abundant to serve either as real-world currency or as sought-after commodity. The core activity of emerging intelligence industries and their computational factories is processing the data circulating through global connectivity networks, discovering patterns in how we direct our attention, inferring from these patterns our personal values and interests, and then using these inferences to first anticipate and then shape human intentions. The core product of these industries and their factories is human behavior.

It is the systematic attraction and exploitation of attention that now drives global circulations of goods and services. In this new attention economy, revenue growth is directly proportional to the effectiveness of machine agencies in discerning what we value and predicting what we can be induced to want. It is an economy in which major platform, social media, and internet service providers are empowered to profit from algorithmically tailoring our increasingly connection-defined human experience. And the holy grail of this new economy—which I will call the Attention Economy 2.0—is not free energy or cheap labor. It is *total attention-share*.

We are, of course, being compensated for allowing our patterns of attention to be mapped and profitably modified. As consumers, in exchange for our attention, we are provided with individually tailored and yet seemingly infinite arrays of choice—for tangible goods and services, for the intangible experiences afforded by film and music, and for social connection. Yet, granted that sustained attention is the single indispensable resource required to engage in creative relational transformation, this is not a costless or value-neutral exchange. As machine and synthetic intelligences are fed ever more revealing access to real-time streams of behavioral, biometric, and other data, the algorithmic tailoring of digitally mediated experience will become ever more seamlessly effective in crafting experiential and connective destinies that are intimately expressive of our own yearnings and desires.

This might be considered by some of us as a dream come true: the technological provision of access to unique experiential paradises for each and every one of us. Others might consider it a nightmare in the making. As we will later see in greater detail, the immense commercial benefits of big data, machine learning, and AI depend on state-sanctioned perforations of the boundary between the public and the private—a marriage of the *attention*

economy and the *surveillance state*. Like traditional arranged marriages that have the primary purpose of alloying and securing family fortunes, the technology-mediated marriage of commercial and state interests is not a love match. It is a highly calculated union through which a small number of global elites gain decisive (and for the most part invisible) control over where and how we connect, with whom, and for what purposes.

This concentration of power, and the funneling up of wealth and opportunity that accompany them, is cause enough for serious global worry. But the deeper concern—and what warrants seeing this already ongoing process of technological transformation as powerfully predicament-generating—is the fact that our ever greater and more individuated *privileges to choose* come at the cost of granting commercial and political actors ever more extensive *rights to control*. As noted earlier, this is not control via *coercion* but via *craving*. The technological route to paradise opened before us is one along which our experiential options will become both wider in scope and ever more acutely desirable but only to the extent that we trade away our "exit rights" from the connective—and thus relational—destinies that are being crafted for us with tireless machine ingenuity: a forfeiture, eventually, of experiential and relational wilderness in exchange for compulsively attractive digital captivity.

To be clear, the claim here is not that digital captivity is our inevitable human destiny. That is one direction—albeit a likely one at present—in which we might be carried by the digitally mediated fusion of dynamic global structures and individual agency. In fact, the rhetorical opposition of "wilderness" and "captivity" is no better in doing justice to the complexity of the transformations in which we find ourselves caught up than is the reduction of competitions in the New Great Game to a bipolar "Cold War" of algorithmic proliferation. Because of the intimate ways in which attention factors into these transformations and competitions, we are not caught up in them contingently; we are caught up both willingly and constitutively. We are captivated, at once enthralled and complicit together in the machine-enabled process of being remade.

This is not as pessimistic a claim as might first appear. The hope informing this book is that our involvement in this process will be both critical and caring. The "alternating current" of data flows and algorithmic responses that are streaming constantly through the intelligence-gathering infrastructure of global network connectivity are dissolving the validity conditions for claims about the opposition of free will and determinism. But even as keen attention to this process forces questioning the relationship of freedom and choice, it also invites creatively exploring the meanings of freedom beyond choice and of agency beyond individuality.

It is hard to fully grasp the fact that exercising greater freedoms of choice as individual agents is consistent with ever greater and more concentrated powers of structurally mediated control. This is especially true for those of us raised to see freedom of choice as a foundational value. The manipulative powers being acquired through the Intelligence Revolution are unprecedented in their precision and reach. In terms of extensive reach, these are powers to affect everything from consumption patterns to voting behavior. In terms of intensive reach, they are powers to affect everything from our emotional lives and dating behavior to our patterns of curiosity. Yet, our intimate implication in the process of generating these powers, and our awareness of the predicament of freedom in a world of "wish-fulfilling" technologies, also position us to play active roles in altering the dynamics of the new attention economy and shaping the course of the Intelligence Revolution.

It is admittedly tempting to argue that the scale and ambient nature of the systems involved render us individually powerless—incapable of personally determining how things turn out, even for ourselves or our families. Global, history-making forces are at work. But this apparent lack of individual *power* is nevertheless compatible with acquiring the relational *strength* needed to make a historical difference through collaborative deliberation and action. In the context of today's recursively evolving human-technology-world relationship, the difference between being implicated in a system of smart *manipulation* and being implicated in a system for intelligent and perhaps liberating *self-discipline* is ultimately not a technical matter of design. It is a profoundly ethical matter of attentive quality and values.

The Importance of Diversity for a Humanely Oriented Intelligence Revolution

A second guiding premise of this book is that global predicament resolution can only be carried out as an interpersonal, intercultural, and intergenerational process and that this implies the centrality of diversity as an ethical value and imperative. In other words, our possibilities for humanely redirecting the dynamics of the Intelligence Revolution are predicated on appreciating the necessity of concerted and resolutely coordinated global action in which differences are engaged as the basis for mutual contribution to sustainably shared flourishing. The book's pivotal question is: who do we need to be present as to engage globally in the work of shared predicament resolution?

One of the distinguishing features of the response offered will be my appeal to classical philosophy and, in particular, to Buddhist thought and personal ideals. Given the unprecedented nature of the ethical labor ahead, appealing to premodern philosophical traditions is perhaps counterintuitive. Our natural inclination is to turn to our own contemporaries for guidance. The logic for this is clear. Contemporary ethicists are able to take into account the nature of our present circumstances and challenges, formulating normative perspectives on personhood from within—and in direct response to—current patterns of technological mediation. Yet, as logical as it might be to appeal to our contemporaries for guidance, it can also be a liability.

If designing and implementing new technologies is simultaneously a process of designing and implementing new norms and processes for being (or becoming) human, the industrial and intelligence revolutions are as fundamentally processes of remaking ourselves as they are processes of crafting new intentional prostheses and strategic environments. Every human-technology-world system reveals new possibilities of presence and action while concealing others. As we will be exploring in detail, technologies are emergent systems of material and conceptual practices that embody and deploy both strategic and normative values: intentional environments that at once shape how we do things and why. As such, maturing technologies have the effect of naturalizing certain forms of agency and qualities of presence. Granted this, there is no escaping the fact that the ethical systems which have evolved in the specific contexts of modern histories of human-technology-world relations will (even if only indirectly) constrain our imaginations of who we can and should be as ethical agents.

The purpose of looking to premodern philosophical traditions is not to entertain the revival of a premodern way of life. We can no more revive past ways of life than we can return to our youths, nor should we wish to do so. Rather, the purpose is to make visible potentially valuable conceptions of exemplary human presence that the evolutionary history of the modern human-technology-world relation has written over in zealous expression of its own salvific inevitability. Carefully engaged, premodern, and indigenous traditions can afford us critical sanctuary. They open perspectives on thinking together toward shared futures not only from "before" the conceptual bifurcation of the individual/personal and the collective/social but also from "before" the technological marvel of an intelligence-gathering infrastructure that profitably fuses personalization and popularization, creating an immaterial alloy of uniqueness of choice and universality of connection that, when algorithmically sharpened, may prove capable of severing the roots of human responsibility and creativity.

Turning to Buddhism for critical insight regarding the risks and most apt responses to intelligent technology is, admittedly, an especially counterintuitive move. In part, it is simply a reflection of my own background. My doctoral training was in Asian and intercultural philosophy with a focus on Chinese Buddhist traditions, and I have maintained a daily Buddhist meditation practice for almost forty years. But there are three more substantive rationales for turning to Buddhist thought and practice for insight.

First, Buddhism was founded as a practical response to the conflicts, troubles and suffering that result when the interdependent origins of all things are ignored in attachment to individual, self-centered existence—a tradition of keen and critical attunement to relational qualities and dynamics, rather than to individual agents and their actions. As such, it offers distinctive conceptual resources for exploring both the experiential and structural ramifications and risks of intelligent technology. Second, as a tradition rooted practically in attention training, it opens prospects both for developing a much-needed, critical phenomenology of the attention economy and for engaging in disciplined and yet freedom-securing resistance to it. Finally, in ways that are particularly useful in understanding the risks of a human-technology-world relationship shaped by evolutionary algorithms creatively intent on actualizing conflicting human values, the Buddhist concept of karma uniquely highlights the inseparability of fact and value and thus the ultimate nonduality of metaphysics and ethics.

Given the popular association of karma with everything from stories of past lives and inescapable fates to bland truisms like "what goes around, comes around," it is perhaps useful to offer a preliminary characterization of the Buddhist teaching of karma. Stated as a conditional, the Buddhist teaching of karma is that if we pay sufficiently close and sustained attention to our own life experiences, it becomes evident how abiding patterns of our values, intentions, and actions invariably occasion consonant patterns of experienced outcomes and opportunities. That is, we discover that we live in a world that is irreducibly meaning laden—a world in which experienced realities always imply responsibility and in which what we conventionally objectify as matter is ultimately the definition of a point of view: an emergent function of what has mattered to us. Karmic ethics is thus an ethics of predicament resolution, an ethics concerned critically not only with how and why we think and act as we do but also with the qualities of consciousness and intention embodied in our thinking and acting. It is, in short, an ethics of commitment to realizing progressively liberating patterns of attention.

The Plan of the Book

Given the centrality of Buddhism to my interpretation of the Intelligence Revolution and my convictions about who we need to be present as to navigate through and beyond the ethical singularity it is precipitating, Chapter One offers a concise introduction to Buddhist thought. For readers unfamiliar with Buddhism, this will provide a body of considerations within which the various narrative threads about intelligent technology presented in the succeeding chapters will be able to productively resonate, much as the wooden body of an acoustic guitar amplifies and accentuates resonances when its strings are plucked or strummed.

One of the practical implications of a karma is that sustainably alleviating and/or eliminating conflict, trouble, and suffering is only possible on the basis of a clarity about how things have come to be as they are. That is, histories matter. Real possibilities for course correction in the context of the Intelligence Revolution depend on understanding, not only the factual historical confluences but also the aspirational narratives out of which intelligent technology has emerged. Chapter Two offers an overview of the history of AI that stresses key streams of research and development and their breakthrough mergers over the last decade.

Chapters Three opens by presenting working definitions of intelligence and technology and explores how intelligent technology is transforming the human lifeworld as corporate and state interests combine in the playing of the New Great Game for digital supremacy. Building on this structural overview of the Intelligence Revolution, Chapter Four looks prospectively forward toward the human-technology-world relationship that current applications of big data, machine learning, and AI seem likely to bring into being over the next ten to fifteen years. Chapter Five considers whether the existential risks, environmental threats, and inequalities being generated by the Intelligence Revolution can be dealt with adequately from within the horizons of existing ethical systems and then makes use of Buddhist resources to argue on behalf of the necessity of an intercultural or "ethical ecosystem" approach to addressing the global predicament of intelligent technology. Granted the soundness of that argument, the question of who we need to be present as to contribute to resolving the intelligence predicament can be restated more precisely as "who do we need to be present as to engage in diversity-enhancing ethical improvisation?"

Chapter Six offers a preliminary response to that question, exploring first the Confucian virtue of relational intimacy and the Socratic virtue of rational integrity and then their blending within the Buddhist ethical ideal

of attentive and responsive virtuosity. Chapter Seven investigates Buddhist personal ideals of uncompelled and compassionately engaged presence and the centrality of improvisation in Buddhist ethics. The final two chapters reflect on how to move forward practically to open prospects for a humane turn in the Intelligence Revolution, realizing the conditions for turning aspirational ideals into realities employing a Middle Path strategy of resistance and redirection focused on data governance and education.

It is important to stress at the outset that the extensive use of Buddhist concepts to understand the dynamics and risks of the Intelligence Revolution, and to frame a set of personal ideals suited to engaging those dynamics and risks, is not meant to be exclusionary. Buddhist resources are very helpful in laying out not only why ethical improvisation is needed to resolve the predicament posed by intelligent technology but also why it is imperative to ecologically integrate ethical perspectives native to a wide range cultural traditions and historical periods. Rather than a Buddhist ethics of technology, what is presented in the pages to follow is a Buddhism-inflected rationale for the indispensability of ethical diversity in responding to the challenges of intelligent technology.

*

The transformation of the human-technology-world relationship by intelligent technology is not a blind process set in motion by transcendent forces, and what it means for our futures is our shared, human responsibility. Intelligent technology is exposing us to deepening structural risks of mass unemployment and underemployment; of precipitously deepening inequalities of wealth, income, and opportunity; and of outsourcing morally charged decision-making to autonomous machine agents. As currently oriented, the dynamics of intelligent technology are liable to fix in place and amplify conflicts among human values and intentions of the kind that are at the roots of global predicaments like climate change, the persistence of hunger in a world of food excess, and our collective human failure to realize conditions of dignity for all in an era of the greatest wealth generation in history. Machine intelligences are poised to diligently transform the world as servants, savants, soldiers, and solicitors scaling the best and worst of human intentions, mirroring back to us our own patterns of attention and inattention.

The Intelligence Revolution cannot carry humanity forward into a more humane and equitable world as long as *we* are incapable of consistently humane and equitable conduct. Realizing the liberating potentials of intelligent technology will depend on whether we succeed in liberating ourselves from our tendencies to embody conflicting values, intentions, and

actions. Given humanity's historical track record, the prospects of realizing such liberating potentials might seem poor. But humanity has never been faced with such clear assurance of ultimate responsibility for the futures we share or such clear imperatives for shouldering that responsibility collaboratively. Given this, sober optimism is perhaps not a fruitless exercise in unreasonable or empty hope.

1

Buddhism: A Philosophical Repertoire

The history of Buddhism is roughly as long as the history of imagining intelligent artificial beings. But whereas the history of artificial intelligence (AI) as a scientific and technological quest begins in earnest only at the dawn of the modern era, Buddhist practices, institutions, and systems of thought were evolving with great vitality and with substantial social, political, and cultural impacts across South, Central, and East Asia as early as the second century. Most of this long history is of no particular relevance for who we need to be present as to engage in collaborative global predicament resolution. But some historical context is useful in appreciating the contemporary relevance of Buddhist personal ideals and their conceptual roots.

Although Buddhism is customarily presented as a "world religion," this is somewhat misleading. Prior to the invitation of Buddhist teachers from various parts of Asia to the 1893 World Parliament of Religions in Chicago, few Buddhists would have identified themselves generically as practitioners of "Buddhism." Instead, most Buddhists would have identified with one or more lines of transmission, passing through a specific teacher, text, or temple, that connected them to Buddhism's founding figure, Siddhartha Gautama, who is generally referred with the honorific title the "Enlightened One" or Buddha. Traveling across premodern Asia, one would have passed through dramatically differing "ecologies of enlightenment" thriving at scales ranging from the local to the regional, each one of which would have been characterized by a distinctive set of personally transmitted practices and supporting teachings.

The differences among Buddhist traditions are in part a function of their historical depth. Although scholars continue to debate the dates of the Buddha's life, Buddhist practices and teachings date back to at least the fifth century BCE. As these practices and teachings spread from their origin in the Himalayan foothills across most of Eurasia over the next twelve hundred years, substantial differentiation would be expected as a matter of course. But in addition—and in contrast with religions built on foundations of divine revelation—Buddhism valued adaptive variation. For example, while the earliest strata of teachings attributed to the historical Buddha have remained important to the present day, the Buddhist canon remained open

for well over a thousand years after first being put into writing in the first century BCE.

This openness was quite self-conscious. Buddhism originated in the context of dramatic rural-to-urban migration and small-scale industrialization on the Indian subcontinent and spread along with expanding trade throughout South Asia and then into Central, East, and Southeast Asia. An important factor in the rapidity of this spread was the Buddha's insistence that his teachings be conveyed in local languages and in ways that were both accessible to different audiences and responsive to local concerns.

In addition, unlike religious traditions based on divinely revealed truths that must be accepted on faith, Buddhist teachings were presented as perspectives on the human experience that could be verified personally—instructions or guidelines for truing or better aligning our ways of being present. Instead of theoretical reflections or metaphysical declarations, Buddhist teachings laid out a therapeutic system for understanding and alleviating psychological and social malaise. In sum, Buddhism originated as a purposefully evolving repertoire of practices for personally realizing liberating forms of relationality.

Buddhist Origins in Predicament Resolution

It is significant that Buddhist teachings originated in response to the challenges of day-to-day life—including those of sickness, old age, and death—during a time of rapid societal change, mass migration, and expanding trade. As is true today in urban centers around the world, when people from different cultural and moral communities find themselves living in close company, values conflicts are practically unavoidable. Buddhist teachings thus responded, at least in part, to the personal and cultural predicaments that we experience when compelled to make conscious and often uncomfortable decisions about which of our customs and identities to abandon and which new customs and identities to adopt in their place. In such disconcerting contexts, we often discover that what we had assumed to be universal commonsense is not actually common to everyone and that many of our most deeply ingrained habits of thought, feeling, speech, and action are as likely to trigger discord as they are to bring expected results.

In its most succinct formulation, the Buddhist therapeutic system is built around just four "truths" or "realities" (*satya*): (1) the presence of conflict, trouble, and suffering (*dukkha*; *duḥkha*) in the human experience; (2) the origination and persistence of conflict, trouble, and suffering as a function of value-infused patterns of causes and conditions; (3) the

presence of possibilities for disrupting and dissolving those patterns; and (4) the existence of a pathway or method for accomplishing this personally through the embodied realization of moral clarity (*śīla*), attentive mastery (*samādhi*), and wisdom (*paññā*; *prajñā*).[1] These four truths/realities are not transcendentally derived eternal declarations about the origins and purpose of the world or our places in it. They are points of embarkation for revising *how we are present*.

Interdependence: The Primacy of Relationality

The pivotal insight informing the Buddhist therapeutic system is that all things and beings arise interdependently (*paṭiccasamuppāda*; *pratītyasamutpāda*). Thanks to the environmental crises triggered by human industrial activity over the last half century, this claim is not as strikingly novel today as it was during the Buddha's lifetime. We are now at least intellectually familiar with the idea that everything in nature is deeply interconnected, but also that we are intimately interconnected with one another and with nature. The interdependence of economies around the world is now taken for granted, and awareness of the global predicament of human-induced climate change has made practical engagement with patterns of deep interconnection a transnational moral and political imperative.

At the time Siddhartha Gautama had this insight—the culmination of his six-year search to discover how to sustainably alleviate experiences of conflict, trouble, and suffering—it ran very much counter to all prevailing religious and philosophical convictions about the nature of reality. These convictions were arrayed along a spectrum. At one end were convictions about the ultimate reality of "spirit/mind." At the other end were convictions about the ultimate reality of "matter/body." In between was a shifting array of hybrid or dualist views. The so-called Middle Path forwarded by the Buddha ran perpendicular or oblique to this spectrum rather than taking up a position somewhere along it. In effect, it was a method for moving beyond reductions of reality to some primal 'this' or 'that', as well as beyond combinations thereof. In philosophical terms, the Middle Path denies validity both to metaphysical monism (reality ultimately consists in just one kind of thing) and to metaphysical pluralism (reality ultimately consists in many kinds of things). It is a path of realizing, both personally and progressively, that what is ultimately real is relationality.[2]

At one level, to see all things as interdependent is to see how causality is always in some degree mutual rather than linear or one way. Mutual causality is, of course, at the heart of the contemporary science of ecology, which explores interdependence as the basis of healthy biological organization in

natural ecosystems. In ecosystems, each species—plant, animal, insect, and microbial—makes some special contribution to the vibrancy and resilience of the system as an emergent and relationally sustained whole. Each species both affects and is affected by all of the other species in the system.

This way of describing things suggests, however, that individual plants, animals, insects, and microbes are basically separate entities that have various kinds of contingent relationships with one another. It suggests, for instance, that a tiger captured in the wild and relocated to a zoo remains "the same" tiger. This is the thinking that underlies the American saying that "you can take the girl out of the country, but you can't take the country out of the girl." From a Buddhist perspective, this idiom rightly points toward the mutual implication—the mutual enfolding—of persons and places. But according to its usual interpretation, it wrongly insists that the girl from the countryside who moves to the city always remains essentially a "country girl." This implies some degree of real independence. The Buddhist concept of interdependence goes substantially deeper: ultimately, there is nothing at all that exists independently, in and of itself.

To begin appreciating the radical sweep of this claim, consider trees. When I was a first-year graduate student preparing to introduce the idea of interdependence to local elementary students in a Philosophy for Children class, I asked my 7-year-old son the apparently simple question, "What is a tree?" His immediate response was to point out the living room window. Rather than defining what it is to be a tree, he offered me an example. Rephrasing, I asked him what it "took" to be a tree. Without hesitating, he described trees as having trunks and lots of branches and leaves, thus distinguishing trees from bushes and other plants. With a bit of prompting, he added that trees also had roots, like the one he'd tripped over at the playground the day before, and that to have trees you needed soil, sunshine, and rain.

To find out how imaginatively he was able to expand the horizons of what it takes to have trees, I then asked what would happen if you put a tree up in orbit—say, in the international space station. Would a tree remain a "tree" in zero gravity? Would it continue to have a canopy of skyward-lifted leaves supported by a more or less vertical trunk? Would its roots still grow "downward" if there were no longer any gravitational 'up' or 'down'? With these questions, he really got into the spirit of things. To have trees, you need the Earth. But the Earth wouldn't be the Earth without the Sun. With his library-fed imagination shifting into high gear, he described how the Earth orbits the Sun as part of one solar system among millions and millions of other solar systems in the Milky Way galaxy, which is itself just one galaxy among millions and millions of galaxies. To have a tree, he excitedly and confidently concluded, you needed the whole universe!

An "obvious" rejoinder is that while it's true that Earth's trees in some sense depend on the rest of the universe to exist as they do, the reverse surely can't be true. We can easily imagine trees (or, for that matter, human beings) ceasing to exist due to the destruction of the Earth by a massive meteor strike and this being just an infinitesimally minor event in a known universe that is ninety-three billion light years in diameter, with each light year equaling six trillion miles. But this imaginary scenario depends on taking up a kind of "view from nowhere" from which to observe the persistence of the cosmos without the Earth. If trees and humans were indeed eliminated, could it really be claimed that what remains is still the universe as we now know it? As sentient organisms, trees and humans are not just *objects in* the universe, they are *perspectives on* a world that only exists as such through them. To eliminate a perspective is to eliminate all that appears to exist objectively from it. Or, to put this somewhat more dramatically: any apparently objective universe exists only as a function or result of a specific point of view.

It is interesting to note here that something like this interdependence of the observed world and perspectives of observation is central to the physics of relativity and quantum mechanics that revolutionized modern science in the early twentieth century. As we will see in discussing karma in more detail, what Buddhism distinctively insists on is the sentient—that is, the feeling or affective—nature and quality of perspectival presence. Interdependence can be of different qualities, different "flavors." Thus, while realizing the interdependent or relational nature of all things is the essence of Buddhist wisdom, what makes that realization therapeutic is its marriage with compassion.

Emptiness (*śūnyatā*) and the Conceptual Nature of 'Things'

The statement that so-called objective reality is actually a function of perspective or point of view might be interpreted as an idealist claim about the ultimate reality of thought or experience: a denial of materiality. But idealist insistence on the independent existence of mind or spirit is not consistent with the insight that reality is ultimately relational. The Buddhist Middle Path, we should recall, runs athwart both idealist and materialist reductions. If the world prior to the assumption of perspectives *on* it can be characterized at all, it is as open or ambiguous. As the second century Indian Buddhist philosopher Nāgārjuna put it: the interdependence of all things means that they are ultimately empty (*śūnya*) of any kind of independent, essential, or abiding "self-nature" (*svabhāva*). To be unrelated is not to be at all.[3] Realizing the emptiness (*śūnyatā*) of all things means realizing their mutual relevance. To be is to mean something for others.

According to the main streams of Buddhist philosophy, we do not normally perceive the world this way because our experience is linguistically and conceptually conditioned. Words and concepts single out for attention different aspects of experience, attributing to them what amount to individual and at least relatively fixed identities. As I walk outside my front door and circle the house, I see an octopus tree, a mango tree, a lime tree, and an avocado tree. Above, there are some cumulus clouds. And across the valley is the extinct volcano, Mount Tantalus. Each of these appears to be an entirely separate entity. Walking is not the same as running; cooking is not the same as eating. But what these various words refer to are not independently existing entities or processes. They refer to different depths, qualities, and kinds of interaction—different patterns of dynamic relationality. The "Tantalus" of steeply pitched and thickly forested cinder cone that I clamber up has a weighty gravity totally absent in the "Tantalus" of winding roads and hidden residential cul-de-sacs that I explore in my pickup truck.

Reality, as it is constituted through the interactions of an ant with its surroundings, is not the same as reality constituted through the interactions of a human with the "same" surroundings. There are no 'clouds' as we know them in the ant's world, no 'mountains' or 'trees', and certainly no 'atoms' and 'molecules.' But the world of human interactivity does not include—except abstractly or through some technological proxy—the complex interplay of chemical signatures that are crucial to the ant's senses of location and presence. The cliff face that we can *see* is impossible to scale is, for the ant, a readily navigable landscape. Although it is a common belief that scientific and mathematical knowledge are getting better and better at revealing world "as it truly is," all that we are really justified in claiming is that scientific and mathematical descriptions bring into clearer and more detailed focus what human interactivity has been bringing into being *as* the world. The scope of our knowledge of the world may in some measurable sense exceed that of ants, but the human world is not ultimately any more real than the world of ant experience.

Concepts are distillations of stable patterns of interaction or relationality. Languages are media for interpersonal and intergenerational clarifications and elaborations of these patterns. That is, words do not refer to things in the world, but to *relational conventions*. Making distinctions between 'this' and 'that', or between what something 'is' and what it 'is-not', are not acts of *discovery*; they are acts of interest- or value-driven *disambiguation*. What each language makes evident—even the languages of mathematics and logic—is what Mahayana Buddhist philosophers have called "provisional" (*saṁvṛti*) reality, the world as we have enacted it, and not an "ultimate" (*paramārtha*) reality existing independently of our knowing relationships with it.

The differences among things are not intrinsic to *them*, but to *our interactions with them*. Or more accurately stated, the only boundaries that obtain among things are boundaries that we have imposed. Like the horizons we see as we turn slowly in a circle on a beach, these boundaries do not reveal absolute features of the world; they reveal features of our own perspectival presence. At every scale and in every domain—perceptual, cognitive, and emotional—everything we experience is the result of what we have elicited from our environments as actionable and as valuable. What *exists* for us is what *matters* for us. Indeed, matter is simply the definition of a point of view. Every existence marks the presence of a specific horizon of relevance.

To make this less abstract, consider the question of who comes first, parents or children? For most of us, the answer seems obvious. We might appreciate the logical tangle involved in trying to answer the structurally similar question about which comes first, the chicken or the egg. Nevertheless, our parents certainly preceded us and we certainly precede our own children. But in actuality, no one can exist *as* a parent prior to conceiving or adopting a child. Parenthood and childhood are coeval; neither can exist before or without the other. Moreover, what it means to be a mother, father, son, or daughter is not fixed. In Buddhist terms, parents and children are empty of any intrinsic and abiding self-nature. Our individual presences *as* fathers, mothers, sons, and daughters reveal what each of us identifies with as "me" and takes on as "my" roles in the dynamics of the family, and this will change dramatically over time. What matters and what it means to be a son at age eight or eighteen is not the same as at age 65 or 70. What "father," "mother," "son," and "daughter" refer to in any given family at any given time are value-expressing horizons of relevance, responsibility, and readiness.

Taken together, the Buddhist teachings of interdependence and emptiness do not offer a metaphysical claim about the nature of reality or "how things really are." Rather than instructing us as to *what* things are, they offer guidance in *how* we should see things in order to author our own liberation from conflict, trouble, and suffering. In other words, their purpose is exhortation, not revelation—encouraging us to depart from disputations about what 'is' or 'is not' real, true, mundane, or divine and to concern ourselves with the emancipatory significance of evaluating relational qualities.

Karma: The Meaning-Articulating Dynamics of Intentional Presence

By not seeing "reality" as something *discovered*, but rather as something *conferred*—a process of bringing or carrying together (Latin: *con+ferre*)—we effectively give precedence to ethics over both epistemology and metaphysics.[4] Who we *should* be present *as* has precedence over who we presently *are*, or

experience conflict, trouble, and suffering. Given this, and given that our experience is indeed karmically ordered, the fact that we *do* experience conflict, trouble, and suffering has to be understood as evidence of tensions between enacted values and intentions that individually would bring about desirable consequences, but that in combination generate troubling relational and experiential cross talk. In short, our experiences of *duḥkha* are evidence of the predicament-laden nature of our experience. Changing the *way* things are changing is ultimately not a matter of altering this or that aspect of our objective circumstances. It involves the much harder labor of predicament resolution.

To anticipate why the concept of karma might be particularly relevant for clarifying the risks of a craving- and control-driven attention economy, as well as the more general value conflicts accentuated by the Intelligence Revolution, consider how our internet experience is now shaped by self-improving algorithms (decision-making procedures) that use patterns of our searches, preferences, and purchases to determine and structure the web content made available to us. These algorithms are designed to continuously shape and reshape our online experiences to bring them into ever closer accord with what we would like, based on our own expressed aims and interests. In short, the values embodied in our connective behavior are used recursively to configure and progressively reconfigure our online realities. In effect, the computational factories of the Intelligence Revolution are functioning as karmic engines.

In theory, the web affords us access to everything digital. In actuality, algorithmic filtering is crafting the web content we experience so that it becomes ever more suitably and uniquely our own. The values we express and act upon set the horizons and content of our future experience. To play with a phrasing gleaned from the Mahāyāna Buddhist text, the Diamond Sutra: our 'everything' is not everything, even though we have no reason not to refer to it as "everything." It *is* all that we are being enabled to experience, even though it does *not* include everything that could be referred to *as* "everything." Algorithms are functioning as arbiters of our "digital karma."

The teaching of karma is that our offline lives, just like our online lives, are shaped by patterns among our values and intentions. The values that inform our thinking, speaking, and acting establish our life headings, specifying or bringing into actionable focus only some aspects of all that is (or could have been) present. But our computationally engineered karma—the shaping of experience that is being undertaken by ambient machine learning systems—is not geared to present desired outcomes in ways that also afford opportunities for changing our values, intentions, and actions in response to any troubling patterns of interdependence in which they might be implicating us. On the

contrary, these surrogate machine agents are geared toward figuring out how to give us only and always what we want.

The ethical concern is not just that we are being enabled to live in "filter bubbles" (Pariser 2012), insulated from having to confront others with different values and interests. Conflict avoidance is an all-too-common tactic for managing our experience and requires no digital assistance whatsoever. The concern is that decisions about whether to avoid others with different views are being made for us. If they continue developing as they are at present, the intelligence industries and their computational factories have the potential to become so sure and seamless in anticipating and providing us with what we like and desire that we will lose the opportunity to make and learn from our own mistakes.

The technologically achieved impossibility of making mistakes and being disappointed in choices we have made might initially seem appealing. But reflecting for a moment on what it would mean to become effectively "locked in" to the patterns of values and interests we had as teenagers or toddlers should suffice to see the very real dangers involved. What is at stake, ultimately, is the opportunity to engage in the intelligent human practice of predicament resolution, developing the creativity and moral maturity needed to participate effectively in intercultural and intergenerational ethical improvisation.

Buddhist Practice: The Teaching of the Three Marks

The purpose of Buddhist practice is to change who we are present as in order to be effective in realizing more liberating patterns of change and freeing ourselves from *duḥkha*. The so-called teaching of the three marks (*tilakkhaṇa*; *trilakṣaṇa*) is one of the most succinct formulations of how to do so. Although it is often represented as a set of metaphysical doctrines about how things *are*, the teaching of the three marks is more aptly understood as a set of practical directives for dissolving the conditions for conflict, trouble, and suffering: seeing all things *as* troubled or troubling (*duḥkha*); *as* impermanent (*anicca*; *anitya*); and *as* without essence/self (*anattā*; *anatman*). It is, in other words, a method for revising *how* we are present: a method for relinquishing the fixed standpoints from which we have habitually demarcated 'me' and 'mine' from 'you' and 'yours', qualitatively transforming our dynamic interdependence from the inside out.

Seeing All Things as Troubling. The teaching of the three marks suggests that seeing all things as marked by conflict, trouble, and suffering is a method for eliminating the conditions that give rise to these undesirable

experiences. But for most of us, accustomed as we are to evaluating things on the basis of our individual perspectives, the invitation to see all things as marked by *duḥkha* does not make intuitive sense as a way to eliminate the conditions that give rise to it. In fact, it can seem like an invitation to adopt the relentlessly pessimistic view that "life is suffering." This interpretation, however, is consistent neither with traditional presentations of the teaching of the three marks nor with the aim of Buddhist practice.

To begin with, it is undeniable that the flavors of a hearty American breakfast of bacon, eggs, and toast are pleasurable. Falling in love is a glorious feeling. A walk along the seashore, listening to waves washing up and then percolating down into the sand is deeply relaxing. Empirically, it is simply not the case that "life is suffering." The point of seeing all things *as* troubled/troubling is not to foster abject pessimism. It is to practice shifting attention according to context and across scales of interdependence, verifying the falsity of assuming that if "I'm okay, you must be okay." It is, in short, a practice of cultivating ethical awareness.

To begin appreciating how this works, consider the mundane example of having bacon and eggs for breakfast. Only a slight shift of perspective is needed to see that this breakfast means something different for humans, hogs, and chickens. Our pleasure while eating bacon comes at the cost of the hog's life. While chickens do not have to die for us to eat their eggs, if they are not free-range chickens, their industrially constrained lives are spent more or less immobilized as biological "machines" in "factories" engineered to produce eggs at the highest possible rates and for the lowest possible unit cost. Industrial agribusiness treats animals not as ends in themselves but as mere means to meeting human needs for nutrition and pleasure, doing so in ways that satisfy corporate desires for profit.

Even if we dismiss the rights of animals to moral consideration, however, our breakfast of industrially produced bacon and eggs can still be seen as troubling. The more successfully the goals of industrial agribusiness are met, the more difficult it is for family farms to remain viable as businesses without adopting similar cost-cutting practices. Moreover, when eating bacon and eggs is promoted as a nutritional ideal for the purpose of increasing agribusiness profits, the resulting negative health impacts can be quite pronounced. Eating bacon and eggs in moderation can be both pleasurable and nutritious. But being indoctrinated by corporate advertising to eat bacon and eggs every day can easily have health-compromising, if not life-threatening, impacts. Seeing a bacon and egg breakfast as *duḥkha* is a method for bringing ethical considerations to bear on the pleasures of eating and the structural dynamics of meeting daily nutritional needs.

The point of seeing all things as implicated in patterns of interdependence characterized by conflict, trouble, and suffering is not to stop enjoying life. The point is to expand the horizons of responsibility, within the compass of which we make life decisions, determining what to do and what to refrain from doing. Our actions are never simply ways of *effecting* our own aims and interests. Our actions are always also *affecting* others and the world around us in ways that will eventually affect us in return. This experiential feedback can occur because others interpret and directly respond to our actions. Or it can occur through systemic transformations like those brought about by environmental degradation and climate change. Either way, acting on others and the world around us is acting on ourselves. There are no agents that are not also patients of their own actions.

This means that at least when other sentient agents are involved, there is always an emotional aspect to the feedback loop joining intentional conduct and subsequent experiential outcomes and opportunities. Our behavior is *felt* and responded to by others as wanted or as unwanted, as pleasant or as unpleasant, as consonant or as dissonant with their own nature and interests. In fact, even an inanimate piece of wood will "resist" our artisanal efforts if care is not taken to align our interests and actions with the wood's grain—the record of its formation and deformation over time. We can, of course, use power tools in carving wood. We can override others' interests by employing physical, social, legal, or other powers in excess of their capacities for resistance. We can force things to go our way. But as we will see very clearly in discussing the human-technology-world relationship, the karmic result will be experiences of new and deeper kinds of resistance and resentment, and thus multiplying opportunities, if not compulsions, to exercise still greater power.

Alternatively, we can refrain from exerting power over others. Doing so will enable us in many instances to enjoy experiences of resonant mutuality—a sharing of aims and interests accompanied by feelings of accelerated movement in desirable directions. But this clearly is not always the case. Consider, for example, getting a workplace promotion, or winning a prestigious scholarship, or simply attracting someone's romantic interest. In each case, even if our happiness was not actively sought at others' expense, there will often be others who will be deeply disappointed when things go decisively our way and not theirs. When one company gains overwhelming market dominance, others are at risk of failing to remain solvent; when one country achieves overwhelming military might, others fear aggression and act accordingly. To see all things *as* troubled or troubling is not to indulge in abject pessimism; it is to see all experience in terms of mutually affecting, meaningfully felt, and hence ethically charged relationships.

Seeing All Things as Impermanent. Seeing all things as impermanent undermines expectations that good situations will last forever. The pleasing sense of fullness after a fine meal passes away into new hunger. The sensual blossoming of romantic love does not last a lifetime. Health is interrupted by illness. Getting your dream job is not the same as doing that job day in and day out, month after month, year after year. At some point, to quote the old blues song, we discover that "the thrill is gone."

Seeing all things as impermanent is, at a deeper level, to practice seeing all things as processes—as emergent phenomena within always ongoing relational dynamics, rather than as stably existing entities that are either shoved about by the winds or currents of change or as abiding agents that somehow manage to ride them. The Greek philosopher Heraclitus famously proclaimed the impossibility of stepping into "the same" river twice since the waters comprising it are always flowing onward. To this, Buddhism adds that we as observers and the bridges or rocky promontories from which we view the river are also always "flowing on." Bridges and rocky promontories may "flow" more slowly than humans, but they age and change character no less certainly than we do over time. What sets humans and other sentient beings apart are our capacities for affecting the pace, direction, and qualitative dynamics of change. We are not merely beings who are *subject to* change; we are ever-becoming *subjects of* change.

Most importantly for Buddhist practice, seeing all things as impermanent also entails recognizing that no bad experiences last forever and that no situations are truly intractable. If change is always ongoing, there can never be any real question about whether change is possible. What is uncertain are only the direction, pace, and quality of change. Seeing all things as impermanent is thus training to be aware that it is always possible to change *how* things are changing. Even when we find ourselves apparently stuck in a troubling pattern of interaction, the *meaning* of this pattern—the experiential consequences of its apparent persistence, perhaps in spite of our sincere efforts to break free of it—is never fixed. The significance of our situation is always negotiable.

It follows from this that whenever we claim there is no way for us to change some situation, we are in fact proclaiming that we are only subject to—and not also the responsible and responsive subjects of—our experience. Seeing all things as impermanent involves seeing how intractability never belongs to our circumstances, but to the fixed nature of the positions we assume within them. The experience of being stuck where and as we are announces our failure or refusal to attend to our situation as one in which we are openly and dynamically implicated. Karmically, the feeling that there is nothing we can do to change our situation is self-justifying evidence of our own resistance to changing who we are present as.

attentive gravity. The self—the presumed 'owner' of experience—is only virtually (and not actually) present.

This Buddhist practice of seeing our personal presence as a composition of five ever-shifting clusters of mutually dependent experiential factors has been compared to the so-called bundle theory of self that was forwarded by David Hume (1711–1776) in his *Treatise on Human Nature*, more recent versions of which have been articulated by thinkers like Daniel Dennett, Derek Parfit, and Owen Flanagan. But whereas these so-called reductionist theories of self effectively deny the presence of a lasting moral agent, equating being without-self with being without-responsibility, the Buddhist concept of no-self or being without-self is intimately allied with teachings about the centrality of karma and compassion in the therapeutic system of Buddhist practice.[8] In fact, the Buddha was often confronted by critics who had interpreted the practice of seeing all things as without-self in reductionist terms and who then pointedly challenged him to explain how karma could operate in the absence of an agent that persists over time and is the moral patient of its own past actions. It's instructive that in these encounters, the Buddha resolutely insisted on according primacy to therapy (what works) rather than to theory (what can be explained) and often remained silent as a way of letting critics know that their questions were unanswerable in the terms that they were posed.

Among these questions were: "Do I really exist now? Did I exist in the past, and if so who was I and how did I live? Will I live again in future lives and who will I be? Where did I come from and where am I ultimately bound?" These questions are unanswerable because responding to them involves committing to some *duḥkha*-generating point of view. "I have a self. I have no self. It is precisely by means of self that I perceive self. It is precisely by means of self that I perceive not-self. It is precisely by means of not-self that I perceive self. This very self of mine—the knower who is sensitive here and there to the ripening of good and bad actions—is constant, everlasting, eternal, not subject to change, and will stay just as it is for eternity" (*Sabbasava Sutta,* MN 2). However different these points of view may be theoretically, they are alike in being therapeutically counterproductive.[9] Adopting any of these positions only results in being caught up in birth, aging, and death; in pain, distress, and despair; and in conflict, trouble, and suffering.

Understanding the teaching of no-self in its original soteriological context is important insurance against concluding that being without-self prohibits having purposes, values, or bases for making and keeping commitments. Being without-self is not *being absent*. It is training to be present so that "in the seen there will only be the seen; in the heard, only the heard; in the sensed,

only the sensed; and in the cognized, only the cognized." Being present in this way, "no 'you' is *with that*; when no 'you' is *with that*, no 'you' is *in that*; and when there is no 'you' either *with that* or *in that*, there is no 'you' here, there or in between the two, and it is precisely this that is the end of *duḥkha*" (Bāhiya Sūtta, *Udana* 1.10). Being without-self is *not* being present *as subject to* suffering, conflict, and trouble. It is not the absence of continuous intentional moral agency, but only that of the abiding moral agent who might be permanently damaged by conflict or scarred by experience. The practice of being without-self is one of becoming present as needed for the emergence of liberating agency.

Some sense of what this practice involves can be gleaned from reflecting, for example, on the moment one first learned to ride a bicycle. I learned from my older brother, who had me sit on the bike and begin pedaling as he steadied the handlebar, walking and then jogging next to me for a few yards before releasing me to continue on my own. The unfailing result was a rapid loss of balance and a crash onto the concrete sidewalk or the grass beside it. But then everything clicked. My brother let go of the handlebar and suddenly there was no thinking about peddling and keeping the handlebars steady and staying on the sidewalk, no worrying about falling. "I" vanished and all that was left was balance in movement, peddling freely and easily, faster and faster. At moments like this, we are neither in control nor out of control. We are present only as fully embodied and wholly active (in this case, bike riding) agency.

This achievement of total immersion in an activity can occur in almost any kind of activity—in doing art, in sports, in musical performance, in manual or mental labor. It is generally the result of intense effort applied over time in a consistent practice, resulting in what the psychologist Mihaly Csikszentmihalyi (1990) has termed optimal experience or "flow"—an exhilarating realization of activity so deeply and thoroughly concentrated as to bring about one's complete absorption in it. Usually, however, this is not a transferable achievement. "Flow" is a domain-specific, challenge-handling, and striving-generated achievement. A virtuosic musician or world-class athlete may be able to somewhat regularly disappear into the "flow" of peak performance and yet be wholly incapable of doing so outside the concert hall or off the playing field. In other parts of their lives, they may continue to find themselves blocked or caught by their circumstances, acting neither wisely nor compassionately. The purpose of Buddhist practice is to be present without-self in all circumstances, realizing ever-deepening accord with one's circumstances, and responding as needed to bring about more liberating relational dynamics, benefitting all involved.

An Ethics of Compassionate Relational Virtuosity

If ethics is the art of human course correction, Buddhist ethics is course correction based on the integral embodiment of wisdom (Pali: *paññā*; Skt.: *prajñā*), moral clarity/discipline (Pali: *sīla*; Skt.: *śīla*), and attentive mastery (Pali and Skt.: *samādhi*). It is their joint cultivation that enables skillfully and successfully directing relational dynamics away from the affective distortions of conflict, trouble, and suffering, and toward nirvana (Pali: *nibbāna*), the consummate aim of the so-called Eightfold Path of Buddhist practice: engaging in correct/corrective action, speech, and livelihood as part of cultivating moral clarity (*śīla*); correct/corrective effort, mindfulness/remembrance, and attentive poise as aspects of cultivating attentive mastery (*samādhi*); and correct/corrective views and intentions as dimensions of cultivating wisdom (*prajñā*).

Nirvana literally means "blown out" or "cooled down." The first modern European and American interpreters of Buddhism, grappling with the foreignness of teachings that stressed being without-self and realizing the emptiness of all things, can perhaps be forgiven for concluding that Buddhism was a nihilistic religion in which salvation amounted to being snuffed out like a candle flame. In fact, the extinguishing of a candle flame was a common metaphor used by many early Buddhists to explain nirvana. But the point of the metaphor was to direct attention to the *process* by means of which the "flames" of conflict, trouble, and suffering are put out. Just as removing a burning candle's wick or the oxygen surrounding it will result in its flame disappearing, the "flames" of *duḥkha* will be extinguished by removing the conditions of belief in independent existences and captivation by clinging forms of desire. Thus, rather than the goal of Buddhist practice, a destination to be arrived at, nirvana is its therapeutic orientation.

It is one of the distinguishing features of Buddhist ethics that its ultimate aim, the "good" toward which Buddhist practice is oriented, is not positively characterized and remains resolutely undefined. Buddhist ethics offers no conceptual maps or fixed principles for arriving rationally at the "good life" or building a "good society." Instead, consistent with the practices of seeing all things as implicated in conflict, trouble, and suffering, as changing and as lacking any fixed identities, implicit to Buddhist ethics is an acceptance of the fact that the course corrections that may be warranted in any particular situation could not have been determined in advance. Indeed, the very possibility of attaining nirvana is predicated on the karmic fact that the way things are changing is conditional and always open to change. Buddhist ethics consists in *skillful* course correction in the absence of a "moral telescope" that might allow us to see in advance where we should be going.

This skill depends on the insights and wisdom that emerge through the continuous and deepening practices of seeing all things as interdependent, as karmically configured, and as without-self. Buddhist wisdom is not something achieved *through* acquiring specific bodies of knowledge or through enduring the perspective-widening processes of aging and maturation. It is an achievement *of* steadfastly relinquishing the horizons of relevance, responsibility, and readiness that until now have *defined* who we *are* and thus limited who we have been capable of being *present as*.[10] Buddhist ethics involves *furthering* that process. In carrying out relational course corrections, wisdom does not function as a "moral telescope," but as a "moral compass."

Successful course correction also requires a very clear understanding of current conditions. It is one thing to know that every situation we find ourselves in is an expression of some karmically configured pattern of interdependence and that relational turbulence is ultimately the result of conflicting values and intentions. It is another to discern and correctly read the currents of intention and value that are implicated in *this* situation and exactly *how* they are affecting relational dynamics. It is not possible for a sailor to keep on course without being keenly sensitive to even the most subtle shifts in currents and winds. That is not possible if he or she is daydreaming or drunk. Cultivating and maintaining keen *sensitivity* to karmic currents and the winds of passions and desires is crucial to Buddhist ethics.

As part of the moral discipline involved in cultivating and maintaining moral sensitivity, all Buddhist practitioners take five vows: to refrain from harming or killing others, from speaking in hurtful ways, from sexual impropriety, from using intoxicants to the point of heedlessness, and from taking what was not freely given. In much the same way that basic hygiene practices like regularly washing our hands and cleaning our homes can prevent us from catching and spreading many common contagious diseases, refraining from these actions works as a kind of karmic hygiene that ensures basic "moral health." But maintaining a clean body and home is no guarantee of optimal health. In addition, an exercise regime and supportive diet may be needed as well. Keeping the five precepts is good, but realizing liberating relational dynamics will require also adequately reading and responding to the karmic currents implicated, for example, in emotional, cognitive, social, cultural, or political conflicts and turbulence: the exercise of attentive capacities that are almost athletic in their focus and flexibility.

To embody wisdom and enact moral clarity requires attentive mastery. We will later discuss the roles played by focus- and flexibility-oriented meditation practices in realizing Buddhist ideals of personal presence. Here, anticipating critical engagement with the dynamics of the attention economy,

it is enough to stress that attention training is integral to the processes of physical, emotional, and intellectual dehabituation that are needed to be freely responsive. The Pali and Sanskrit term for attention, *manasikāra*, simply means awareness that is concentrated or resolutely focused. This implies that one can be attentive with different degrees of concentration or focus. We can devote half our attention to cooking and half to conversing. But in addition to how much attention we are paying to our situation, Buddhism makes a distinction qualitatively between being attentive in ways that bind us to or that free us from conflict, trouble, and suffering.

It is possible, even without training, to be keenly attentive to our present circumstances. Young children avidly awaiting the ice-cream cone being prepared for them and adolescents in the throes of video game ecstasy are both clearly capable of highly concentrated attention. What is not so clear is whether they are freely attentive or compulsively so. Without training, our attention is readily and *involuntarily* attracted or distracted. In particular, we are especially susceptible to unwisely having our attention captured by the superficial, craving-inducing aspects of things (*ayoniśomanasikāra*). This, as we will see, is crucial to the workings of the new attention economy being realized through intelligent technology. Yet, with training, our attention can also be wisely concentrated—directed freely and *intentionally* in ways that are both sensitive to the interdependent origins of things and consistent with *truing* relational patterns (*yoniśomanasikāra*).

To the extent that Buddhist ethics consists in the goalless, nirvana-oriented practice of integrally cultivating wisdom, moral clarity, and attentive mastery, it is hard to place readily or without remainder into one of the standard categories of ethics grounded on definitive and generalized judgments regarding personal character (virtue ethics), duties (deontological ethics), or the consequences of actions (utilitarianism). Given Buddhism's ethical insistence on pairing wisdom with compassion, a closer fit might be care ethics, with its emphasis on situationally apt attentive responsiveness. But Buddhist compassion is not reducible to the natural inclinations to care about and for others that are invoked by care ethics, much less to the abstractly mandated responses to suffering that are typically framed with reference to personal virtues or duties, or derived through a consequentialist calculus of harms and happiness. Rather, Buddhist compassion is exemplified in the ongoing intentional practice of dissolving the karmic causes and conditions of shared conflict, trouble, and suffering—a necessarily improvisational labor of shared predicament resolution in steadfast pursuit of increasingly liberating relational outcomes and opportunities.

What makes Buddhist ethics so difficult to place (and, potentially, so relevant today) is the fact that it offers only an open-ended training

technology is human-centered to consider how best to ensure that humanity and intelligent technology are as *humanely* interdependent as possible.

Prior to the eighteenth century, the word "humane" was simply a variant for "human." It was only in the eighteenth century, roughly at the beginning of the First Industrial Revolution, that "humane" came to signify kindness, compassion, and benevolence—qualities that were for the first time seen as attributable, not only to people but to actions, processes, and institutions. Given the open-ended nature of Buddhist ethics, these qualities should not be considered either definitive or exhaustive of a superlative human-technology-world relationship. But at least provisionally, they afford a footing on which to frame explicitly qualitative (if not fully normative) concerns about intelligent technology. As the following chapters will make evident, the Fourth Industrial Revolution—an Intelligence Revolution—is well underway and rapidly accelerating, almost miraculously scaling up human values and intentions. Given the stakes involved, it is fortunate that whether it will prove to be a truly humane revolution is still to be determined.

2

Artificial Intelligence: A Brief History

When we think of the conflicts of interests arising with the emergence of intelligent technology, we think of the future. Can machine intelligences be developed in ways that align with human values? Will artificial intelligence (AI) surpass human intelligence one day? How will the Intelligence Revolution change the lives of our children and grandchildren? Will the internet of things make us more secure or more vulnerable? These are all forward-looking questions. The premise of this chapter is that if we want to be able to influence what the Intelligence Revolution will mean for the human experience, we need also to look backward.

One of the core insights of the Buddha was that we can only effectively and sustainably resolve the conflicts, troubles, or suffering that we are experiencing on the basis of first understanding how things have come to be as they are. A "snapshot" of the present is not enough, no matter how wide-angled the lens or how detailed the image. We need a "film," and ideally one with "footage" shot from many different angles. Histories matter.

Servants of Our Own Making: Dreams of Artificial Beings and Mechanizing Reason

Humans are tool-makers. We are not unique in this. Rudimentary tool-making and tool-use are known among at least several other species. What has been unique is the extraordinary inventiveness of our tool-making and the range of uses to which we have put our tools. But tools only do so much. They can extend or amplify our own efforts, but they will not do our work for us. Speculatively, it's not much of a stretch to imagine that recognizing the limitations of tools might have spurred the development of draft animal domestication and slavery practices, both of which became common at roughly the same time in the large agricultural societies developing in Mesopotamia some five thousand five hundred years ago.

But while draft animals, slaves, and servants can be made to do tool-using work, they also need to be fed and kept healthy. Resistance and revolt

are always possible. So, it's perhaps not surprising that the tool-making imagination would eventually entertain the possibility of creating tool-using beings capable of tirelessly and contently doing one's bidding. The earliest evidence of such an imagined perfection of the tool-making art—the creation of artifacts capable of doing all the work normally undertaken by draft animals and servants—is in the Greek epic, the *Iliad*.[1] There we find brief but tantalizing descriptions of "self-propelled chairs" and "golden attendants" crafted by none other than Hephaistos—the tool-wielding and tool-making god of sculptors and blacksmiths.

Similar stories of artificial servants or companions were fairly common in the works of classical Greek and Roman writers and persisted into early modern times.[2] Often, as in Ovid's *Metamorphoses*, written some two thousand years ago, these artificial beings are humanly crafted statues brought to life by their maker's loving desire and the grace of the gods. Among these stories, *The Fairie Queene*, published by Edmund Spenser in 1590, is unique in featuring an "iron man" that is granted by an immortal to one of the epic poem's protagonists, not to satisfy his personal desires but as a sword-wielding assistant in the noble—if often violent—work of dispensing justice. But in all these premodern tales, even if these artificial beings were crafted by human hands out of clay, stone, or metal, they needed to be animated or "inspired" by the gods.

The first material evidence of imagining that it might be within human reach to build a functioning artificial servant is a set of drawings by Leonard da Vinci. Drafted around 1495, these drawings of a mechanical knight depict inner works comprising an array of pulleys and gears that could be set in motion without divine animation. Interestingly, like the "iron man" who would appear a century later in *The Fairie Queene* and in the dreams of many of those who are funding AI research today, Leonardo's mechanical knight was designed for martial labor.

Human tool-making ingenuity was not up to the task of making anything even remotely like iron men or mechanical knights until well into the last century. Mechanizing mental labor turned out to be much easier than mechanizing physical labor. About the same time Spenser was penning *The Fairie Queene*, a new "curriculum" model of education was being forwarded in which knowledge was a quantifiable good that could be analyzed into component parts for delivery by means of standardized lessons in competitively graded short courses.[3] This new approach to learning was premised on the ideas that reasoning is based on logic and that all forms of knowledge should aspire to the crystalline purity and certainty of mathematical proofs. This association of the commanding heights of human

reasoning and intelligence with mathematics—more an exception than a rule across most of human history—proved decisive in setting the course of efforts to build machine intelligence.

By the mid-seventeenth century, machines for performing mathematical operations like addition, multiplication, subtraction, and division were being built in France and Germany, and a "logic demonstrator" was constructed in 1777 by Charles Stanhope (1753–1816) that proved machines could generate logical proofs. Half a century later, after building a "difference engine" that could carry forward the results of a calculation to succeeding operations—an elementary form of machine memory—Charles Babbage (1791–1891) drafted plans for an "analytical engine" that featured a logic unit and an integrated memory, the design of which anticipated the engineering logic and circuitry of the first mainframe computers that were eventually built in the late 1930s.

A major shortcoming of the calculators and logical devices built through the early twentieth century was their reliance on mechanically transmitted energy. Computers made of relatively heavy metal parts require a great deal of energy to set and keep in motion and then suffered from mechanical strain and heat buildup. In effect, the precision limits of machining and assembly effectively set caps on processing speeds and operational complexity. The construction of general purpose programmable computers with substantial working memory became possible only with the inventions of vacuum tubes and solid state electronics that have no moving parts.

Significantly for the course of computing history, it was the paroxysm of the Second World War that put electronic computing on a development fast track. Making advances in weapons design and manufacturing were key military priorities—as were advances in communication and code-breaking—and electronic computers were critical for carrying out the complex calculations involved. Pursuing military/strategic advantage through computational artifice has remained a key driver of basic computing, communications, and AI research ever since.

Modeling Thought: The Research Origins of the Intelligence Revolution

One of the major contributors to Allied efforts to advance computer science was the British mathematician Alan Turing (1912–1954). He was also one of the first scientists to maintain that building artificial general intelligence was within human reach. His core insight was that any act of reasoning that

could be converted into a set of algorithms or rule-bound decision-making procedures could be simulated by a sufficiently complex electronic device. While all the operations carried out by such a device could, in principle, be carried out by unassisted human beings at the relatively slow speed of electrochemical exchanges in the brain, the machine's electronic substrate would enable these operations to be carried out at near light speed. Any reasoning that could be formally encoded could also be automated and accelerated.

Over the next decade, remarkably productive mergers were crafted among advances made in what had previously been the largely separate academic research fields of engineering, logic, neurophysiology, evolutionary theory, and cognitive science. By the mid-1950s, the basic principles of cybernetics and the role of feedback mechanisms had been laid out by Norbert Wiener (1894–1964), and growing numbers of mathematicians and computer scientists were beginning to wonder how to best approach building a general purpose AI. A high-level seminar on machine learning was hosted in Los Angeles in 1955, followed by a profoundly influential summer research program on AI at Dartmouth College in 1956, and by a 1958 conference on the mechanization of thought processes hosted at the National Physical Laboratory of the United Kingdom.[4] With these conferences, the Intelligence Revolution can be said to have begun in earnest. AI was no longer seen as the stuff of dreams but as a research agenda worthy of substantial, dedicated investment.

Over the first generation of serious AI research, two major approaches emerged. One approach, building on presumptions about the close relationship among mathematics, logic, and reasoning, was "neat" in the sense that it aimed for precision in programming and in solving well-defined problems. The successes were striking. Digitally computable programs were written that were able to generate proofs for algebra word problems and mathematical theorems, sometimes doing so more elegantly than had previously been done by humans. Systems were built that automatically produced relatively crude but still functional and cost-effective, translations between natural languages like Russian and English. And machine vision and robotics developed to the point that artificial "agents" could carry out simple building procedures and navigate through obstacles in a controlled environment.

But as impressive as the advances made by "neat" research were, with the exception of translation machines and search engines operating on unique databases, the AI that was resulting could easily be dismissed as capable of handling nothing more than "toy" problems. As the philosopher Hubert Dreyfus (1972) pointed out, this was partly because those working

reason at human levels, the funding winter came to an end. In competitive response, the American and British governments quickly rebooted funding for AI research.

The most notable result of this investment thaw was the emergence of a billion-dollar industry in so-called expert systems. Originally developed in the mid-1960s, expert systems combined a detailed "knowledge base" with an "inference engine" designed on the basis of interviews with human experts in fields relevant to decision-making in the target knowledge domain. Early successes in analyzing chemical compounds and matching disease symptoms with antibiotic prescriptions had proved the validity of the concept. But it was not until the 1980s that it became possible to build general-purpose inference engines that could be "fueled" with domain-specific human expertise and large and fluid data sets. This proved to be a remarkably powerful way of addressing a range of business needs, including monitoring and managing inventory, diagnosing operational bottlenecks, scheduling and guiding equipment maintenance, and evaluating credit applications.

Yet, even well-designed expert systems were susceptible to breaking down when given unusual inputs (the "brittleness" problem), and it was difficult to map out in advance all the preconditions involved in successful action in real-world contexts (the "qualification" problem). And while these systems worked well in decision-making contexts where a few hundred inference rules would suffice, in complex contexts where thousands of rules might be needed and/or where a constantly evolving model of the knowledge domain was required, effective and reliable expert systems were much harder to deliver. Moreover, the growth of personal computing in the early 1990s and the development of more intuitive interface architectures fostered growing decentralization of business computing applications, and by mid-decade the boom in expert systems had largely gone bust.

Although the period from the late 1980s into the mid-1990s is sometimes referred to as a "second AI winter," seen another way it was a period of fruitful convergence and commingling among various streams of AI research. Hearkening back to Hubert Dreyfus's argument that human cognition is fundamentally embodied and environmentally situated, an "actionist" or embodied approach to machine intelligence developed around the idea that intelligence is rooted in sensory-motor coupling with an ever-changing world and in a proprioceptive sense of being present in that world. This yielded significant gains in robotic intelligence. Advances were also made in applying new theoretical work on convolutional and recurrent networks, which greatly improved machine learning performance and proved that there were ways of bypassing extensive supervised training while maintaining full functionality. Other conceptual advances included so-called

mixture-of-expert architectures, the application of probability theory and decision theory to AI, the use of "fuzzy logic," and the development of evolutionary algorithms that could rewrite themselves in adaptive response to their informational environments.

In short, at a conceptual level, rather than a second "winter," the period from the late 1980s to the late 1990s was perhaps something more like a protracted spring "cold snap" with lots of new growth going on just out of sight. In retrospect, it is easy to see that what was preventing AI from really coming into its own was not a dearth of innovative science and engineering but a sufficiently rich information environment—an environment with enough data radiance to nurture and sustain the practical embodiment of machine intelligence. That, however, was just around the corner. The information transmission and generation grid known as the internet was scaling rapidly up from being a network used by military and academic elites into a general purpose infrastructure capable of mediating the mutual adaptation of machine and human intelligences in the complex and diverse informational domains of economic, social, political, and cultural conduct.

A New Informational Infrastructure: The Internet, Personal Computer, and Smartphone

The basic design for the transmission grid of this new infrastructure had been commissioned by DARPA and launched in 1969 as the Advanced Research Projects Agency Network or ARPANET. The original motivation for building this "packet switching" system was to have a secure, decentralized, and node-to-node communication network that could withstand nuclear weapons assault—a system capable of sustaining military and governmental communications under worst-case scenarios. It did not take long, however, for the broader potentials of this network architecture to be realized. Connection to ARPANET grew rapidly among American universities and defense agencies and contractors, and international links were established in 1973. A year later, the "internet" was born with the formation of Telnet, the first commercial internet service provider (ISP).

Other ISPs quickly followed. But for the next fifteen years, the internet was still largely used to connect universities, research centers, and governmental agencies. Readily accessible, commercial dialup internet service was launched in 1989, and in the following year, the basic language and text transfer protocols used in developing websites as we now know them were invented (the hypertext markup language or HTML and the hypertext transfer protocol or HTTP). Two years later, the World Wide

Web was inaugurated as a truly public space when the first open-access web servers were turned on.

Web browsers came onto the market over the next few years, along with the first websites for selling goods over the internet, including Amazon and eBay. High speed cable access to the internet became commercially available in 1996. Google was launched in 1998 and its algorithmic search engine quickly became the most widely engaged machine intelligence in the world. To give a sense of the rapidity of the changes taking place at the time, in just the eighteen-month period from December 1998 to August 2000, the number of households in the United States with personal computers rose from 42 percent to 51 percent and the number with internet access nearly doubled from 26 percent to 42 percent. Roughly 80 percent of all children in the United States were suddenly using computers at school, and 20 percent of all Americans were accessing daily news online (US Census Special Report, 2000). While questions continued to be asked about whether the World Wide Web would ever become an environment suitable for profitable commercial activity, at the close of the 1990s, the internet was unquestionably established as a crucial and expanding dimension of the communication infrastructure for the twenty-first century.

The increasing power and decreasing size of microprocessors that were crucial to the personal computer revolution were at the same time enabling both the miniaturization and expanded functionality of mobile communications devices. With the 2002 rollout of the Blackberry, a handheld, internet-linked device with a small but functional keyboard for composing text messages, the era of 24/7 email and internet access was born. The introduction of the first Apple iPhone in 2007—which featured touch screen operation and support for both Web 2.0 (user-generated internet content) and third-party applications—revolutionized mobile communications and triggered dramatic growth in the variety and use of digital social media.

Seen at the level of fiber-optic cables, satellites, and server farms, the phenomenal growth of the digital network infrastructure of the internet was a triumph of physical engineering. But what this physical network made possible was a networking of machine and human intelligences in digital environments that fostered a coevolutionary intelligence explosion. With vast troves of data and aided by spectacular gains in computing speed and memory, artificial neural networks and machine learning algorithms were suddenly poised for unprecedented successes. Over a handful of years, after nearly half a century of concerted effort and slow progress, machine vision and speech recognition suddenly improved to levels that first rivaled and then surpassed human capabilities. From solving "toy" problems and learning to play games like checkers and chess, AI was suddenly able to

"graduate," leaving laboratory "schools" to start real on-the-job training. Artificial servants, savants, seers, and soldiers were no longer merely the stuff of dreams.

Artificial Agency and the Goal of Intentional Partnership

Like the internet, the idea of building digitally embodied forms of agency that blend deep machine learning, unlimited information reach, and a natural language interface began as a DARPA brainchild. Although it received almost no media coverage, in 2003, DARPA initiated a project aimed at developing a Cognitive Agent that Learns and Organizes (CALO). It was the largest single AI project that had ever been funded: a $150 million dollar, five-year effort that involved some 350 people at SRI International, a leading technology research and development corporation associated with Stanford University.

The details of the project are instructive. DARPA's mandate was ambitious: develop a personal assistant that could learn onsite and in real time to assist military personnel execute their duties across a range of activity domains from supply management to command post. Such an assistant would incorporate the decision-recommendation capabilities of expert systems, the search and learning capabilities of software agents based on neural networks akin to those in Deep Blue (the computer system that defeated world chess champion Gary Kasparov in 1997), the connectivity needed to carry out commands in a full spectrum of real-world environments, and natural language processing abilities sufficiently advanced to allow completely hands-free partnership. In short, the virtual personal assistant sought by DARPA was one that could be seamlessly integrated into the military workplace and that could not only learn to provide requested information but also to anticipate what information might be relevant and when, offering decision options, managing routine tasks, and carrying out user commands immediately.

According to the SRI website, the aim of the CALO project was "to create cognitive software systems … that can reason, learn from experience, be told what to do, explain what they are doing, reflect on their experience, and respond robustly to surprise."[5] It was successful enough that SRI exercised its legal right to pursue further research aimed at building a virtual assistant that could be marketed to the public. The commercial potential was obvious. SRI spun off a separate company in 2007 to continue working on a commercially viable personal virtual assistant. Two years later, Siri Incorporated, this spinoff company, premiered a virtual assistant software that could be installed on

any smartphone. Within a matter of months, the company was purchased by Apple for an amount estimated to be in the neighborhood of $200 million.

As it had been developed at Siri Incorporated, the virtual personal assistant was envisioned as a comprehensive "do engine" and not just a "search engine." Linked to a suite of more than forty web services, it could suggest alternative travel plans in the case of a cancelled flight, make rental car reservations, book tables at restaurants based on user-stipulated preferences in consultation with a range of restaurant rating sites, and pull together a list of news stories on topics of personal interest in the last ten days. Unlike expert systems that relied on a structured database of knowledge, the virtual assistant could draw on a range of internet-accessible databases and learn how to deploy information from them in completely uncontrolled and unstructured environments.

In 2011, with considerable fanfare, Apple launched Siri—a stripped down version of the virtual assistant with no "do engine" functionality—as a key feature of its new iPhone 4. In the years since, Apple has incrementally added capabilities to and refined Siri, and a host of other virtual personal assistants have come on the market. Most of these have been general purpose "chatbots" that are low on "do engine" or execution capabilities—Amazon's Alexa, Microsoft's Cortana, and Google's Now being among the most well-known. Other virtual assistants have been designed with greater action-capability but for use in specific contexts—for example, travel-related services.

This is changing. Disenchanted with Apple's decision to market a dumbed-down version of the assistant that they had developed, roughly a third of the original Siri team left Apple in 2011 to form Viv Incorporated. Their mission was simple but enterprisingly visionary: make AI a "utility" like water or electricity—a necessity of daily life in the twenty-first century. After five years of development, the company launched its new virtual assistant in spring 2016. According its website, "Viv is an artificial intelligence platform that enables developers to distribute their products through an intelligent, conversational interface. It's the simplest way for the world to interact with devices, services and things everywhere. Viv is taught by the world, knows more than it is taught, and learns every day."[6]

Unlike expert systems, Viv is not programmed to perform specific tasks or supplied with a fixed knowledge base. Armed with state-of-the-art natural language processing and constructed around the deep learning architecture behind the successes of Deep Blue and Alpha Star (a computer system that learned on its own how to play the multiplayer strategic game StarCraft II at grandmaster level), Viv is able to interpret a user's intention and to write a program for executing that intention by assembling all the required resources from as many different digital and real environments as necessary. And it is able to do this in a matter of milliseconds. One might inform Viv, for

production capabilities into the things we interact with daily, ranging from our pens, running shoes, and medicine bottles to our cars and our refrigerators. This infusion of connectivity into everyday objects—the creation of the so-called internet of things—is producing a world of ever more enticingly "enchanted objects" (Rose 2015). Running shoes embedded with internet-connected sensors keep track of your pace and the routes you run and calculate how many calories you burn. The caps of your elderly parents' "enchanted" medicine bottles change color to remind them when they need to take their medications, while at the same time sending records of whether they are doing so to their health care providers. It's estimated that by 2025, the average, connected person will interact with some 4,800 such devices per day and that this will result in the generation of 163 zettabytes of data globally per year (Reinsel, et.al. 2017). To put this in perspective, in a single year, humanity will produce enough data to make a high definition video lasting longer than the five-billion-year history of the Earth.

Drawing on data gathered from credit/debit card purchases, web searches, and text, image, and video postings to social media, algorithmic intelligences have become remarkably effective at personalizing product and service advertising and pricing. As might be imagined, this is a very valuable skill.[10] But, in addition to running recommendation engines, algorithms nourished by big data are also skilled at producing consumer credit ratings; organizing airline flight schedules; making "risk assessments" and "evidenced-based" recommendations regarding bail, sentencing, and parole; and finding patterns of disease treatment effectiveness that have until now eluded human recognition.

All of the data that is now being uploaded into the "cloud" by our use of internet service providers, social media platforms, online retailing, credit/debit cards, smartphone payment apps, navigation devices, and the internet of things also falls as data "rain" that can be channeled back into AI development and "deep learning," further energizing and extending the reach and effectiveness of algorithmic agency. In short, the more we make use of virtual assistants and deploy algorithmic agencies, the more transparently and powerfully they will be able to respond to our expressed needs and desires, but also the more precisely they will be able to interpret and anticipate our actions and intentions.

The Fourth Industrial Revolution: A Revolution in the Cloud

The confluence of evolutionary machine learning algorithms and big data is changing both the pace and character of the Intelligence Revolution.

Although it is still apt to regard it as an industrial revolution, it is industrial in a new way. Asked to think about an industrial revolution, although we know that the factories of today are not like their nineteenth and twentieth century forebears, most of us will still envision blunt-faced, utilitarian structures built of brick or concrete, bustling with activity and exhaling plumes of fossil fuel smoke. Inside them, we would expect to find skeletal constructs of steel and brass set into cacophonous motion by steam engines or electrical generators, devouring raw materials and step-by-step transforming them into conveyor belt–conducted parades of identical finished products. These are valid imaginations. Until quite recently, most industrial production was carried out according to a marvelously visible logic of moving mechanical parts. You could see it taking place.

Asked to reflect on the presence of AI, machine learning, or big data in our lives, most of us will think first about computers and smartphones and the recommendation "engines" and navigation services we access through them. Or, we might think about self-driving cars or robotic surgeons. In short, we are inclined to think in modern industrial terms about material objects and processes: physical machines that somehow manage to behave intelligently. The "factories," "machines," and "products" specific to the Intelligence Revolution, however, are neither strictly located nor directly visible. It is true that there are data server farms and cloud computing campuses housed in structures that can be as large as six million square feet in floor area and that individually can require millions of gallons of water daily for cooling purposes. But these buildings and the equipment in them are not the actual factories of the emerging AI industries. They are analogous at best to the brick and mortar shells in which factory equipment was operated in the heyday of the Machine Age, many of which have now been gainfully repurposed as innovation centers or commercial complexes.

Walking along a server farm's seemingly interminable, identical aisles of head-high racks of lightly glowing equipment faces, you will not *see* any of the work being done. The Intelligence Revolution is industrial, but its factories and machines are computational. They are not built out of concrete or metal but out of mathematical and logical codes. It is these codes that "magically" instruct the movement of electrical energy through circuits so finely etched into silicon substrates that more than twenty-five million transistors can be fitted into a single square millimeter. There are causal processes at work, but we are not in a position to physically witness them. Open up a smartphone or a tablet computer and you will not see machine intelligence at work. You will not witness the transformation of any raw materials into finished products. The industrial factories and machines proper to the Intelligence Revolution