# Introduction

One of the enduring concerns of moral philosophy is deciding who or what is deserving of ethical consideration. Although initially limited to "other men," the practice of ethics has developed in such a way that it continually challenges its own restrictions and comes to encompass what had been previously excluded individuals and groups—foreigners, women, animals, and even the environment. Currently, we stand on the verge of another fundamental challenge to moral thinking. This challenge comes from the autonomous, intelligent machines of our own making, and it puts in question many deep-seated assumptions about who or what constitutes a moral subject. The way we address and respond to this challenge will have a profound effect on how we understand ourselves, our place in the world, and our responsibilities to the other entities encountered here.

Take for example one of the quintessential illustrations of both the promise and peril of autonomous machine decision making, Stanley Kubrick's *2001: A Space Odyssey* (1968). In this popular science fiction film, the HAL 9000 computer endeavors to protect the integrity of a deep-space mission to Jupiter by ending the life of the spacecraft's human crew. In response to this action, the remaining human occupant of the spacecraft terminates HAL by shutting down the computer's higher cognitive functions, effectively killing this artificially intelligent machine. The scenario obviously makes for compelling cinematic drama, but it also illustrates a number of intriguing and important philosophical problems: Can machines be held responsible for actions that affect human beings? What limitations, if any, should guide autonomous decision making by artificial intelligence systems, computers, or robots? Is it possible to program such mechanisms

with an appropriate sense of right and wrong? What moral responsibilities would these machines have to us, and what responsibilities might we have to such ethically minded machines?

Although initially presented in science fiction, these questions are increasingly becoming science fact. Researchers working in the fields of artificial intelligence (AI), information and communication technology (ICT), and robotics are beginning to talk quite seriously about ethics. In particular, they are interested in what is now called the ethically programmed machine and the moral standing of artificial autonomous agents. In the past several years, for instance, there has been a noticeable increase in the number of dedicated conferences, symposia, and workshops with provocative titles like "Machine Ethics," "EthicALife," "AI, Ethics, and (Quasi)Human Rights," and "Roboethics"; scholarly articles and books addressing this subject matter like Luciano Floridi's "Information Ethics" (1999), J. Storrs Hall's "Ethics for Machines" (2001), Anderson et al.'s "Toward Machine Ethics" (2004), and Wendell Wallach and Colin Allen's *Moral Machines* (2009); and even publicly funded initiatives like South Korea's Robot Ethics Charter (see Lovgren 2007), which is designed to anticipate potential problems with autonomous machines and to prevent human abuse of robots, and Japan's Ministry of Economy, Trade and Industry, which is purportedly working on a code of behavior for robots, especially those employed in the elder care industry (see Christensen 2006).

Before this new development in moral thinking advances too far, we should take the time to ask some fundamental philosophical questions. Namely, what kind of moral claim might such mechanisms have? What are the philosophical grounds for such a claim? And what would it mean to articulate and practice an ethics of this subject? *The Machine Question* seeks to address, evaluate, and respond to these queries. In doing so, it is designed to have a fundamental and transformative effect on both the current state and future possibilities of moral philosophy, altering not so much the rules of the game but questioning who or what gets to participate.

## The Machine Question

If there is a "bad guy" in contemporary philosophy, that title arguably belongs to René Descartes. This is not because Descartes was a particularly

bad individual or did anything that would be considered morally suspect. Quite the contrary. It is simply because he, in the course of developing his particular brand of modern philosophy, came to associate the animal with the machine, introducing an influential concept—the doctrine of the *bête-machine* or *animal-machine*. "Perhaps the most notorious of the dualistic thinkers," Akira Mizuta Lippit (2000, 33) writes, "Descartes has come to stand for the insistent segregation of the human and animal worlds in philosophy. Likening animals to automata, Descartes argues in the 1637 *Discourse on the Method* that not only 'do the beasts have less reason than men, but they have no reason at all.'" For Descartes, the human being was considered the sole creature capable of rational thought—the one entity able to say, and be certain in its saying, *cogito ergo sum*. Following from this, he had concluded that other animals not only lacked reason but were nothing more than mindless automata that, like clockwork mechanisms, simply followed predetermined instructions programmed in the disposition of their various parts or organs. Conceptualized in this fashion, the animal and machine were effectively indistinguishable and ontologically the same. "If any such machine," Descartes wrote, "had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals" (Descartes 1988, 44). Beginning with Descartes, then, the animal and machine share a common form of alterity that situates them as completely different from and distinctly other than human. Despite pursuing a method of doubt that, as Jacques Derrida (2008, 75) describes it, reaches "a level of hyperbole," Descartes "never doubted that the animal was only a machine."

Following this decision, animals have not traditionally been considered a legitimate subject of moral concern. Determined to be mere mechanisms, they are simply instruments to be used more or less effectively by human beings, who are typically the only things that matter. When Kant (1985), for instance, defined morality as involving the rational determination of the will, the animal, which does not by definition possess reason, is immediately and categorically excluded. The practical employment of reason does not concern the animal, and, when Kant does make mention of animality (*Tierheit*), he does so only in order to use it as a foil by which to define the limits of humanity proper. Theodor Adorno, as Derrida points out in the final essay of *Paper Machine*, takes the interpretation one step

further, arguing that Kant not only excluded animality from moral consideration but held everything associated with the animal in contempt: "He [Adorno] particularly blames Kant, whom he respects too much from another point of view, for not giving any place in his concept of dignity (*Würde*) and the 'autonomy' of man to any compassion (*Mitleid*) between man and the animal. Nothing is more odious (*verhasster*) to Kantian man, says Adorno, than remembering a resemblance or affinity between man and animal (*die Erinnerung an die Tierähnlichkeit des Menschen*). The Kantian feels only hate for human animality" (Derrida 2005, 180). The same ethical redlining was instituted and supported in the analytic tradition. According to Tom Regan, this is immediately apparent in the seminal work of analytical ethics. "It was in 1903 when analytic philosophy's patron saint, George Edward Moore, published his classic, *Principia Ethica*. You can read every word in it. You can read between every line of it. Look where you will, you will not find the slightest hint of attention to 'the animal question.' Natural and nonnatural properties, yes. Definitions and analyses, yes. The open-question argument and the method of isolation, yes. But so much as a word about nonhuman animals? No. Serious moral philosophy, of the analytic variety, back then did not traffic with such ideas" (Regan 1999, xii).

It is only recently that the discipline of philosophy has begun to approach the animal as a legitimate subject of moral consideration. Regan identifies the turning point in a single work: "In 1971, three Oxford philosophers—Roslind and Stanley Godlovitch, and John Harris—published *Animals, Men and Morals*. The volume marked the first time philosophers had collaborated to craft a book that dealt with the moral status of nonhuman animals" (Regan 1999, xi). According to Regan, this particular publication is not only credited with introducing what is now called the "animal question," but launched an entire subdiscipline of moral philosophy where the animal is considered to be a legitimate subject of ethical inquiry. Currently, philosophers of both the analytic and continental varieties find reason to be concerned with animals, and there is a growing body of research addressing issues like the ethical treatment of animals, animal rights, and environmental ethics.

What is remarkable about this development is that at a time when this form of nonhuman otherness is increasingly recognized as a legitimate moral subject, its other, the machine, remains conspicuously absent and

marginalized. Despite all the ink that has been spilled on the animal question, little or nothing has been written about the machine. One could, in fact, redeploy Regan's critique of G. E. Moore's *Principia Ethica* and apply it, with a high degree of accuracy, to any work purporting to address the animal question: "You can read every word in it. You can read between every line of it. Look where you will, you will not find the slightest hint of attention to 'the machine question.'" Even though the fate of the machine, from Descartes forward, was intimately coupled with that of the animal, only one of the pair has qualified for any level of ethical consideration. "We have," in the words of J. Storrs Hall (2001), "never considered ourselves to have 'moral' duties to our machines, or them to us." The machine question, therefore, is the other side of the question of the animal. In effect, it asks about the other that remains outside and marginalized by contemporary philosophy's recent concern for and interest in others.

## Structure and Approach

Formulated as an ethical matter, the machine question will involve two constitutive components. "Moral situations," as Luciano Floridi and J. W. Sanders (2004, 349–350) point out, "commonly involve agents and patients. Let us define the class *A* of moral *agents* as the class of all entities that can in principle qualify as sources of moral action, and the class *P* of moral *patients* as the class of all entities that can in principle qualify as receivers of moral action." According to the analysis provided by Floridi and Sanders (2004, 350), there "can be five logical relations between *A* and *P*." Of these five, three are immediately set aside and excluded from further consideration. This includes situations where *A* and *P* are disjoint and not at all related, situations where *P* is a subset of *A*, and situations where *A* and *P* intersect. The first formulation is excluded from serious consideration because it is determined to be "utterly unrealistic." The other two are set aside mainly because they require a "pure agent"—"a kind of supernatural entity that, like Aristotle's God, affects the world but can never be affected by it" (Floridi and Sanders 2004, 377).[1] "Not surprisingly," Floridi and Sanders (2004, 377) conclude, "most macroethics have kept away from these supernatural speculations and implicitly adopted or even explicitly argued for one of the two remaining alternatives."

> Alternative (1) maintains that all entities that qualify as moral agents also qualify as moral patients and vice versa. It corresponds to a rather intuitive position, according to which the agent/inquirer plays the role of the moral protagonist, and is one of the most popular views in the history of ethics, shared for example by many Christian Ethicists in general and by Kant in particular. We refer to it as the standard position. Alternative (2) holds that all entities that qualify as moral agents also qualify as moral patients but not vice versa. Many entities, most notably animals, seem to qualify as moral patients, even if they are in principle excluded from playing the role of moral agents. This post-environmentalist approach requires a change in perspective, from agent orientation to patient orientation. In view of the previous label, we refer to it as non-standard. (Floridi and Sanders 2004, 350)

Following this arrangement, which is not something that is necessarily unique to Floridi and Sanders's work (see Miller and Williams 1983; Regan 1983; McPherson 1984; Hajdin 1994; Miller 1994), the machine question will be formulated and pursued from both an agent-oriented and patient-oriented perspective.

The investigation begins in chapter 1 by addressing the question of machine moral agency. That is, it commences by asking whether and to what extent machines of various designs and functions might be considered a legitimate moral agent that could be held responsible and accountable for decisions and actions. Clearly, this mode of inquiry already represents a major shift in thinking about technology and the technological artifact. For most if not all of Western intellectual history, technology has been explained and conceptualized as a tool or instrument to be used more or less effectively by human agents. As such, technology itself is neither good nor bad, it is just a more or less convenient or effective means to an end. This "instrumental and anthropological definition of technology," as Martin Heidegger (1977a, 5) called it, is not only influential but is considered to be axiomatic. "Who would," Heidegger asks rhetorically, "ever deny that it is correct? It is in obvious conformity with what we are envisioning when we talk about technology. The instrumental definition of technology is indeed so uncannily correct that it even holds for modern technology, of which, in other respects, we maintain with some justification that it is, in contrast to the older handwork technology, something completely different and therefore new. . . . But this much remains correct: modern technology too is a means to an end" (ibid.).

In asking whether technological artifacts like computers, artificial intelligence, or robots can be considered moral agents, chapter 1 directly and

quite deliberately challenges this "uncannily correct" characterization of technology. To put it in a kind of shorthand or caricature, this part of the investigation asks whether and to what extent the standard dodge of the customer service representative—"I'm sorry, sir, it's not me. It's the computer"—might cease being just another lame excuse and become a situation of legitimate machine responsibility. This fundamental reconfiguration of the question concerning technology will turn out to be no small matter. It will, in fact, have significant consequences for the way we understand technological artifacts, human users, and the presumed limits of moral responsibility.

In chapter 2, the second part of the investigation approaches the machine question from the other side, asking to what extent machines might constitute an Other in situations of moral concern and decision making. It, therefore, takes up the question of whether machines are capable of occupying the position of a moral patient "who" has a legitimate claim to certain rights that would need to be respected and taken into account. In fact, the suspension of the word "who" in quotation marks indicates what is at stake in this matter. "Who" already accords someone or something the status of an Other in social relationships. Typically "who" refers to other human beings—other "persons" (another term that will need to be thoroughly investigated) who like ourselves are due moral respect. In contrast, things, whether they are nonhuman animals, various living and nonliving components of the natural environment, or technological artifacts, are situated under the word "what." As Derrida (2005, 80) points out, the difference between these two small words already marks/makes a decision concerning "who" will count as morally significant and "what" will and can be excluded as a mere thing. And such a decision is not, it should be emphasized, without ethical presumptions, consequence, and implications.

Chapter 2, therefore, asks whether and under what circumstances machines might be moral patients—that is, someone to whom "who" applies and who, as a result of this, has the kind of moral standing that requires an appropriate response. The conceptual precedent for this reconsideration of the moral status of the machine will be its Cartesian other—the animal. Because the animal and machine traditionally share a common form of alterity—one that had initially excluded both from moral consideration—it would seem that innovations in animal rights philosophy

would provide a suitable model for extending a similar kind of moral respect to the machinic other. This, however, is not the case. Animal rights philosophy, it turns out, is just as dismissive of the machine as previous forms of moral thinking were of the animal. This segregation will not only necessitate a critical reevaluation of the project of animal rights philosophy but will also require that the machine question be approached in a manner that is entirely otherwise.

The third and final chapter responds to the critical complications and difficulties that come to be encountered in chapters 1 and 2. Although it is situated, in terms of its structural placement within the text, as a kind of response to the conflicts that develop in the process of considering moral agency on the one side and moral patiency on the other, this third part of the investigation does not aim to balance the competing perspectives, nor does it endeavor to synthesize or sublate their dialectical tension in a kind of Hegelian resolution. Rather, it approaches things otherwise and seeks to articulate a thinking of ethics that operates beyond and in excess of the conceptual boundaries defined by the terms "agent" and "patient." In this way, then, the third chapter constitutes a *deconstruction* of the agent–patient conceptual opposition that already structures, delimits, and regulates the entire field of operations.

I realize, however, that employing the term "deconstruction" in this particular context is doubly problematic. For those familiar with the continental tradition in philosophy, deconstruction, which is typically associated with the work of Jacques Derrida and which gained considerable traction in departments of English and comparative literature in the United States during the last decades of the twentieth century, is not something that is typically associated with efforts in artificial intelligence, cognitive science, computer science, information technology, and robotics. Don Ihde (2000, 59), in particular, has been critical of what he perceives as "the near absence of conference papers, publications, and even of faculty and graduate student interest amongst continental philosophers concerning what is today often called *technoscience*." Derrida, however, is something of an exception to this. He was, in fact, interested in both sides of the animal-machine. At least since the appearance of the posthumously published *The Animal That Therefore I Am*, there is no question regarding Derrida's interest in the question "of the living and of the living animal" (Derrida 2008, 35). "For me," Derrida (2008, 34) explicitly points out, "that

will always have been the most important and decisive question. I have addressed it a thousand times, either directly or obliquely, by means of readings of all the philosophers I have taken an interest in." At the same time, the so-called father of deconstruction (Coman 2004) was just as interested in and concerned with machines, especially writing machines and the machinery of writing. Beginning with, at least, *Of Grammatology* and extending through the later essays and interviews collected in *Paper Machine*, Derrida was clearly interested in and even obsessed with machines, especially the computer, even if, as he admitted, "I know how to make it work (more or less) but I don't know *how* it works" (Derrida 2005, 23).

For those who lean in the direction of the Anglo-American or analytic tradition, however, the term "deconstruction" is enough to put them off their lunch, to use a rather distinct and recognizably Anglophone idiom. Deconstruction is something that is neither recognized as a legitimate philosophical method nor typically respected by mainstream analytic thinkers. As evidence of this, one need look no further than the now famous open letter published on May 9, 1992, in the *Times* of London, signed by a number of well-known and notable analytic philosophers, and offered in reply to Cambridge University's plan to present Derrida with an honorary degree in philosophy. "In the eyes of philosophers," the letter reads, "and certainly among those working in leading departments of philosophy throughout the world, M. Derrida's work does not meet accepted standards of clarity and rigour" (Smith et al. 1992).

Because of this, we should be clear as to what deconstruction entails and how it will be deployed in the context of the machine question. First, the word "deconstruction," to begin with a negative definition, does not mean to take apart, to un-construct, or to disassemble. Despite a popular misconception that has become something of an institutional (mal)practice, it is not a form of destructive analysis, a kind of intellectual demolition, or a process of reverse engineering. As Derrida (1988, 147) described it, "the de- of *de*construction signifies not the demolition of what is constructing itself, but rather what remains to be thought beyond the constructionist or destructionist schema." For this reason, deconstruction is something entirely other than what is understood and delimited by the conceptual opposition situated between, for example, construction and destruction.

Second, to put it schematically, deconstruction comprises a kind of general strategy by which to intervene in this and all other conceptual oppositions that have and continue to organize and regulate systems of knowledge. Toward this end, it involves, as Derrida described it, a double gesture of *inversion* and conceptual *displacement*.

> We must proceed using a double gesture, according to a unity that is both systematic and in and of itself divided, according to a double writing, that is, a writing that is in and of itself multiple, what I called, in "The Double Session," a *double science*. On the one hand, we must traverse a phase of *overturning*. To do justice to this necessity is to recognize that in a classical philosophical opposition we are not dealing with the peaceful coexistence of a *vis-à-vis*, but rather with a violent hierarchy. One of the two terms governs the other (axiologically, logically, etc.), or has the upper hand. To deconstruct the opposition, first of all, is to overturn the hierarchy at a given moment. . . . That being said—and on the other hand—to remain in this phase is still to operate on the terrain of and from the deconstructed system. By means of this double, and precisely stratified, dislodged and dislodging, writing, we must also mark the interval between inversion, which brings low what was high, and the irruptive emergence of a new "concept," a concept that can no longer be, and never could be, included in the previous regime. (Derrida 1981, 41–43)

The third chapter engages in this kind of double gesture or double science. It begins by siding with the traditionally disadvantaged term over and against the one that has typically been privileged in the discourse of the status quo. That is, it initially and strategically sides with and advocates patiency in advance and in opposition to agency, and it does so by demonstrating how "agency" is not some ontologically determined property belonging to an individual entity but is always and already a socially constructed subject position that is "(presup)posited" (Žižek 2008a, 209) and dependent upon an assignment that is instituted, supported, and regulated by others. This conceptual inversion, although shaking things up, is not in and of itself sufficient. It is and remains a mere revolutionary gesture. In simply overturning the standard hierarchy and giving emphasis to the other term, this effort would remain within the conceptual field defined and delimited by the agent–patient dialectic and would continue to play by its rules and according to its regulations. What is needed, therefore, is an additional move, specifically "the irruptive emergence of a new concept" that was not and cannot be comprehended by the previous system. This will, in particular, take the form of another thinking of *patiency* that is not programmed and predetermined as something derived from or the mere

counterpart of agency. It will have been a kind of primordial patiency, or what could be called, following a Derridian practice, an *arche-patient* that is and remains in excess of the agent–patient conceptual opposition.

## Questionable Results

This effort, like many critical ventures, produces what are arguably questionable results. This is precisely, as Derrida (1988, 141) was well aware, "what gets on everyone's nerves." As Neil Postman (1993, 181) aptly characterizes the usual expectation, "anyone who practices the art of cultural criticism must endure being asked, What is the solution to the problems you describe?" This criticism of criticism, although entirely understandable and seemingly informed by good "common sense," is guided by a rather limited understanding of the role, function, and objective of *critique*, one that, it should be pointed out, is organized according to and patronizes the same kind of instrumentalist logic that has been applied to technology. There is, however, a more precise and nuanced definition of the term that is rooted in the traditions and practices of critical philosophy. As Barbara Johnson (1981, xv) characterizes it, a critique is not simply an examination of a particular system's flaw and imperfections that is designed to make things better. Instead "it is an analysis that focuses on the grounds of that system's possibility. The critique reads backwards from what seems natural, obvious, self-evident, or universal, in order to show that these things have their history, their reasons for being the way they are, their effects on what follows from them, and that the starting point is not a given but a construct, usually blind to itself" (ibid.). Understood in this way, critique is not an effort that simply aims to discern problems in order to fix them or to ask questions in order to provide answers. There is, of course, nothing inherently wrong with such a practice. Strictly speaking, however, criticism involves more. It consists in an examination that seeks to identify and to expose a particular system's fundamental operations and conditions of possibility, demonstrating how what initially appears to be beyond question and entirely obvious does, in fact, possess a complex history that not only influences what proceeds from it but is itself often not recognized as such.

This effort is entirely consistent with what is called *philosophy*, but we should again be clear as to what this term denotes. According to one way

of thinking, philosophy comes into play and is useful precisely when and at the point that the empirical sciences run aground or bump up against their own limits. As Derek Partridge and Yorick Wilks (1990, ix) write in *The Foundations of Artificial Intelligence*, "philosophy is a subject that comes running whenever foundational or methodological issues arise." One crucial issue for deciding questions of moral responsibility, for example, has been and continues to be *consciousness*. This is because moral agency in particular is typically defined and delimited by a thinking, conscious subject. What comprises consciousness, however, not only is contentious, but detecting its actual presence or absence in another entity by using empirical or objective modes of measurement remains frustratingly indeterminate and ambiguous. "Precisely because we cannot resolve issues of consciousness entirely through objective measurement and analysis (science), a critical role exists for philosophy" (Kurzweil 2005, 380). Understood in this way, philosophy is conceptualized as a supplementary effort that becomes inserted into the mix to address and patch up something that empirical science is unable to answer.

This is, however, a limited and arguably nonphilosophical understanding of the role and function of philosophy, one that already assumes, among other things, that the objective of any and all inquiry is to supply answers to problems. This is, however, not necessarily accurate or appropriate. "There are," Slavoj Žižek (2006b, 137) argues, "not only true or false solutions, there are also false questions. The task of philosophy is not to provide answers or solutions, but to submit to critical analysis the questions themselves, to make us see how the very way we perceive a problem is an obstacle to its solution." This effort at reflective self-knowledge is, it should be remembered, precisely what Immanuel Kant, the progenitor of critical philosophy, advances in the *Critique of Pure Reason*, where he deliberately avoids responding to the available questions that comprise debate in metaphysics in order to evaluate whether and to what extent the questions themselves have any firm basis or foundation: "I do not mean by this," Kant (1965, Axii) writes in the preface to the first edition, "a critique of books and systems, but of the faculty of reason in general in respect of all knowledge after which it may strive independently of all experience. It will therefore decide as to the possibility or impossibility of metaphysics in general, and determine its sources, its extent, and its limits." Likewise, Daniel Dennett, who occupies what is often considered to be the opposite

end of the philosophical spectrum from the likes of Žižek and Kant, proposes something similar. "I am a philosopher, not a scientist, and we philosophers are better at questions than answers. I haven't begun by insulting myself and my discipline, in spite of first appearances. Finding better questions to ask, and breaking old habits and traditions of asking, is a very difficult part of the grand human project of understanding ourselves and our world" (Dennett 1996, vii).

For Kant, Dennett, Žižek, and many others, the task of philosophy is not to supplement the empirical sciences by supplying answers to questions that remain difficult to answer but to examine critically the available questions in order to evaluate whether we are even asking about the right things to begin with. The objective of *The Machine Question*, therefore, will not be to supply definitive answers to the questions of, for example, machine moral agency or machine moral patiency. Instead it will investigate to what extent the way these "problems" are perceived and articulated might already constitute a significant problem and difficulty. To speak both theoretically and metaphorically by way of an image concerning vision ("theory" being a word derived from an ancient Greek verb, *θεωρέω*, meaning "to look at, view, or behold"), it can be said that a question functions like the frame of a camera. On the one hand, the imposition of a frame makes it possible to see and investigate certain things by locating them within the space of our field of vision. In other words, questions arrange a set of possibilities by enabling things to come into view and to be investigated as such. At the same time, and on the other hand, a frame also and necessarily excludes many other things—things that we may not even know are excluded insofar as they already fall outside the edge of what is able to be seen. In this way, a frame also marginalizes others, leaving them on the exterior and beyond recognition. The point, of course, is not simply to do without frames. There is and always will be a framing device of some sort. The point rather is to develop a mode of questioning that recognizes that all questions, no matter how well formulated and carefully deployed, make exclusive decisions about what is to be included and what gets left out of consideration. The best we can do, what we have to and should do, is continually submit questioning to questioning, asking not only what is given privileged status by a particular question and what necessarily remains excluded but also how a particular mode of inquiry already makes, and cannot avoid making, such decisions; what

assumptions and underlying values this decision patronizes; and what consequences—ontological, epistemological, and moral—follow from it. If the machine question is to be successful as a philosophical inquiry, it will need to ask: What is sighted by the frame that this particular effort imposes? What remains outside the scope of this investigation? And what interests and investments do these particular decisions serve?

In providing this explanation, I do not wish to impugn or otherwise dismiss the more practical efforts to resolve questions of responsibility with and by autonomous or semiautonomous machines and/or robots. These questions (and their possible responses) are certainly an important matter for AI researchers, robotics engineers, computer scientists, lawyers, governments, and so on. What I do intend to point out, however, is that these practical endeavors often proceed and are pursued without a full understanding and appreciation of the legacy, logic, and consequences of the concepts they already mobilize and employ. The critical project, therefore, is an important preliminary or prolegomena to these kinds of subsequent investigations, and it is supplied in order to assist those engaged in these practical efforts to understand the conceptual framework and foundation that already structures and regulates the conflicts and debates they endeavor to address. To proceed without engaging in such a critical preliminary is, as recognized by Kant, not only to grope blindly after often ill-conceived solutions to possibly misdiagnosed ailments but to risk reproducing in a supposedly new and original solution the very problems that one hoped to repair in the first place.