Spring March 27, 2009

Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots

Yueh-Hsuan Weng, *NCA, Ministry of the Interior, Republic of China*
Chien-Hsun Chen, *National Nano Device Laboratories*
Chuen-Tsai Sun, *National Chiao Tung University*

Int J Soc Robot (2009) 1: 267–282
DOI 10.1007/s12369-009-0019-1

ORIGINAL PAPER

Toward the Human–Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots

Yueh-Hsuan Weng · Chien-Hsun Chen · Chuen-Tsai Sun

Accepted: 7 April 2009 / Published online: 25 April 2009

Abstract Technocrats from many developed countries, especially Japan and South Korea, are preparing for the *human–robot co-existence society* that they believe will emerge by 2030. Regulators are assuming that within the next two decades, robots will be capable of adapting to complex, unstructured environments and interacting with humans to assist with the performance of daily life tasks. Unlike heavily regulated industrial robots that toil in isolated settings, *Next Generation Robots* will have relative autonomy, which raises a number of safety issues that are the focus of this article. Our purpose is to describe a framework for a legal system focused on Next Generation Robots safety issues, including a *Safety Intelligence* concept that addresses robot *Open-Texture Risk*. We express doubt that a model based on Isaac Asimov's Three Laws of Robotics can ever be a suitable foundation for creating an artificial moral agency ensuring robot safety. Finally, we make predictions about the most significant Next Generation Robots safety issues that will arise as the human–robot co-existence society emerges.

Keywords Safety intelligence · Robot legal studies · Roboethics · Robot policy · The three laws of robotics · Robot law · Social robotics

1 Introduction

The Japanese Robot Association[1] predicts that *Next Generation Robots* will generate up to 7.2 trillion yen (approximately 64.8 billion USD) of economic activity by 2025, with 4.8 trillion (43.2 billion USD) going to production and sales and 2.4 trillion (21.6 billion USD) to applications and support. According to the Japanese Ministry of Economy, Trade and Industry (METI), manufacturers will focus on specific markets (e.g., housework, nursing, security), while application and support firms provide maintenance, upgrading, and reselling services similar to today's information technology structure [1]. Also similar to the current IT industry, individual firms will specialize in such areas as education (public, safety, technical, etc.), selling insurance to cover special robot risks, and buying/selling used robots.

The *Fukuoka World Robot Declaration*, issued in February 2004, lists Japanese expectations for Next Generation Robots that co-exist with human beings, assist human beings both physically and psychologically, and contribute to the realization of a safe and peaceful society.[2] However, the declaration falls short in describing what Next Generation Robots should be. In a report predicting the near future (2020–2025) in robot development, the Japanese Robot Policy Committee (RPC, established by METI) created

Y.-H. Weng (✉)
Conscription Agency, Ministry of the Interior, Republic of China, Chengkungling, Taiwan
e-mail: yhweng.cs94g@nctu.edu.tw

C.-H. Chen
7531, Lucas Rd, Richmond, BC, V6Y 1G1, Canada
e-mail: lucemia.cis93g@nctu.edu.tw

C.-T. Sun
College of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctsun@cs.nctu.edu.tw

[1] http://www.jara.jp/.

[2] International Robot Fair 2004 Organizing Office, World Robot Declaration (2004), http://www.prnewswire.co.uk/cgi/news/release?id=117957.

two Next Generation Robots categories: (a) next generation industrial robots capable of manufacturing a wide range of products in variable batch sizes, performing multiple tasks, and (unlike their general industrial predecessors) working with and/or near human employees; and (b) service robots capable of performing such tasks as house cleaning, security, nursing, life-support, and entertainment-all functions to be performed in co-existence with humans in businesses and homes. The report predicts that humans will gradually give Next Generation Robots a growing number of repetitive and dangerous service tasks, resulting in increased potential for unpredictable and dangerous actions [2]. METI describes the danger level in terms of contact degree: "low" for in-home communication or cleaning robots, "mid" for nursing robots, and high for universal humanoid robots capable of performing a wide range of tasks. How well such dangers can be anticipated is closely linked to the amount of autonomous behavior programmed into machines, in a relationship that remains to be determined.

Since 2000, Japanese [3] and South Korean [4] technocrats have been discussing and preparing for a *human–robot co-existence society* that they believe will emerge by 2030. Based on the content of policy papers and analyses published by both governments, researchers are currently studying potential *Robot Sociability Problems* that—unlike technical problems associated with design and manufacturing entail robot related impacts on human interactions in terms of regulations, ethics, and environments. Regulators are assuming that within the next two decades, robots will be capable of adapting to complex, unstructured environments and interacting with humans to assist with the performance of daily life tasks. Unlike heavily regulated industrial robots that toil in isolated settings, Next Generation Robots will have relative autonomy allowing for sophisticated interactions with humans. That autonomy raises a number of safety issues that are the focus of this article.

In addition to the semi-autonomous robots created by NASA scientists of the United States for exploration of the deep sea and the surface of Mars, in December 2008, the U.S. military reported its plan to devote approximately $4 billion USD within the following two years for the development of "ethical" robots-autonomous robot soldiers which will conform to the laws of warfare [5].

Artificial intelligence (AI) will be the main tool giving robots autonomy, expanded work ranges, and the ability to work in unstructured environments. Changes in human–robot relationships made possible by advancements in AI are likely to exert an impact on human society rivaling that of any other single technological innovation. We will predict the most significant issues of Next Generation Robots safety that will arise as the human–robot co-existence society emerges. The emerging co-existence society and issues in establishing robot law will be respectively discussed in Sects. 2 and 3. In Sects. 4 and 5 we will describe Human-Based Intelligence and our proposal for a *Safety Intelligence* concept to address these issues in light of Isaac Asimov's *Three Laws of Robotics*. In Sect. 6 we will describe the potential development of a *Legal Machine Language* to overcome the considerable shortcomings of Asimov's laws.

2 Human–Robot Co-Existence Society

The public often views robot development in terms of biomorphic machines. The actual situation involves input and innovation from multiple non-engineering fields that pave the way for harmonious interactions between humans and robots of all shapes, sizes, appearances, and capabilities. We will refer to interdisciplinary issues as *Robot Sociability Problems* and to engineering issues as *Robot Technical Problems*.

The Japanese are funding multiple efforts to address robot sociability problems and robot technical problems issues, including the establishment of research committees and physical environments for the testing of robot prototypes. In 1999, the Ministry of International Trade and Industry (MITI, which later became the above-mentioned METI) provided 450 million USD for a five-year research effort called "HRP: The Humanoid Robotics Project." The participants were Japan's major players in robotics: Hirochika Inoue and Susumu Tachi from the University of Tokyo, and representatives from Honda, Fujitsu, Panasonic (Matsushita), Kawasaki, Hitachi, and other corporations. The first (two-year) stage was dedicated to developing a "humanoid robot platform" (HRP-1) and the second to developing HRP-1 applications associated with human–robot co-existence [6].

A separate project was already underway at Waseda University in Tokyo's Shinjuku ward, the 1973 birthplace of WABOT-1, the world's first full-scale biped walking humanoid robot.[3] Named "Innovative Research on Symbiosis Technology for Humans and Robots in an Elderly-Dominated Society",[4] it was sponsored by the Ministry of Education, Culture, Sports, Science and Technology. Another important agreement was finalized in 2001, when Waseda University and Gifu prefecture established a "Wabot-House"[5] technical area in the city of Kakamigahara. The lab consists of three buildings, one for addressing ideas about ideal living spaces for people and various robot types; one focusing on social factors such as daily living needs, medical

[3] http://www.humanoid.waseda.ac.jp/history.html.

[4] http://www.waseda.jp/prj-rt/English/index/html.

[5] http://www.wabot-house.waseda.ac.jp/.

concerns, the natural environment, and other issues associated with human–robot co-existence; and one for determining how robot–human living spaces can be designed to support suitable levels of autonomous robot behavior [7]. Several other cities are now vying to attract robot researchers. Since late 2003, Fukuoka and Kitakyushu (both in Fukuoka prefecture) have shared the distinction of being the world's first Robot Development Empiricism Area (RDEA),[6] created according to national "Special Zone for Structural Reform" legislation [8]. The law addresses rules for road signs, traffic lanes, and traffic regulations, adding flexibility that allows for limited outdoor testing of robots (mostly on sidewalks) for the purpose of collecting empirical data.

Robot researchers in the two cities receive special tax dispensation and are exempt from Japanese radio law, meaning they do not have to apply for special certification for experiments using the 5 GHz wireless frequency range. A second goal of the Wabot-House group is to establish Gifu prefecture as one of several centers of the Japanese robot industry [9, 10]; other prefectures opening some of their city streets to robots in the interest of attracting manufacturers are Kanagawa (International Rescue Complex project) [11] and Osaka (RoboCity CoRE) [12]. A huge "robot city" is considered essential to planning for and creating a human–robot co-existence society, since the qualities of artificial environments that match the technical requirements of robot functions must be identified. In this regard, researchers are studying the potential for robots to serve as bridges between physical and virtual worlds. As Google's Vinton Cerf observes regarding the Internet:

> Virtual and real worlds will merge. Virtual interactions will have real world consequences. Control of the electrical grid and power generation systems could be made to appear to be part of a virtual environment in which actions in the virtual space affect actions in the real space. If your air conditioner is attached to the Internet, your utility might turn it off to prevent a brownout. Educational environments that mix real and virtual instruments and places will enrich the learning experience of school children [13].

If we add Next Generation Robots to Cerf's scenario, they will simultaneously act as virtual agents and physical actors, with overlapping boundaries that allow for the movement of many Next Generation Robots within a robot city.

[6]The cities of Fukuoka and Kitakyushu were designated as Robot Development Empiricism Research Areas by the Japanese government in November, 2003. The first experiments in using robots in public spaces were conducted in February, 2004. In 2002, Fukuoka and Busan (South Korea) co-sponsored an international "Robocup" competition and conference. See http://www.robocup.or.jp/fukuoka/ and http://www.island-city.net/business/it/it.html.

Before that day arrives, several important policy and regulatory issues must be settled to prevent a legal crisis. We will review some environmental issues first before looking at safety-related and other legal concerns.

An important concept in this area of robot research is *affordance*-the quality of an object or environment that allows an individual to perform an action. Psychologist James J. Gibson, who introduced the term [14], defined affordances as including all latent "action possibilities" in an environment, objectively measurable and independent of an individual's ability to recognize them, but always in relation to the actor and therefore dependent on the actor's capabilities. Since humanoid shapes are currently considered most suitable for a human–robot co-existence society, affordances for humanoid robots will be much more complex than those for industrial robots. Whereas industrial robots are limited to using relatively simple arm-like mechanisms to grab, move, and place objects in a restricted space, robots with legs may someday perform tasks using all four of its limbs—for instance, leaving one's physical home or business to run errands. For Next Generation Robots, affordance issues will involve service applications, their effects on industrial planning, functional imagination (if owners can do certain tasks, why not their robots?), and human–robot psychology [15]. Arguably, the central question is this:

- Should robots be designed to essentially "do anything" using all of their action possibilities?
- Should they be entrusted with providing nursing care in the absence of human caregivers?
- Should they be allowed to use their huge power requirements to feed grapes to their reclining owners?
- Should they be allowed to use their huge power requirements to feed grapes to their reclining owners?
- Should they be capable of sexual relations with humans [16]?

The range of possibilities raises the specter of a complex licensing system to control how Next Generation Robots are used.

Power consumption and generation is one item on a long list of environmental concerns that need to be addressed, preferably before the human–robot co-existence society becomes a reality. We are notoriously messy creatures, putting up hundreds of satellites into orbit above the earth and letting them stay up there long after their original purposes are exhausted. How can we prevent the abandonment of robots used to explore extreme environments that are too dangerous for humans? Furthermore, anyone who has tried to recycle a personal computer or peripheral knows that it is not as easy as placing them in a curbside recycling bin. Robot technology is a complex technical domain that will require a combination of ingenuity and strict enforcement in order to avoid disposal problems that are sure to arise when millions of robots and robot parts break down or wear out.

Ambulatory robots will consume enormous amounts of energy. The International Energy Association (IEA) took that into consideration when making their predictions for future energy needs, and reported that if governments stick to their current policies, electric power consumption will double by 2030, and the percentage of electricity in total energy consumption will rise from 17 to 22 percent. Whereas western countries may find ways to generate "green power" for robots (e.g., fuel cells), developing countries will have little choice but to continue using the least expensive ways to generate electric power. In most cases that means burning coal, thereby increasing the quantity of greenhouse gasses released into the atmosphere [17]. Clearly, the emergence of a human–robot co-existence society makes our search for clean energy sources and ways of sharing them with developing countries-even more imperative.

3 Robot Law

Future robot-related issues will involve human values and social control, and addressing them will require input from legal scholars, social scientists, and public policy makers, using data from researchers familiar with robot legal studies. Levy [18] argues convincingly that a new legal branch of *Robot Law* is required to deal with a technology that by the end of this century will be found in the majority of the world's households. Here we will review the main issues expected to emerge in the fast-arriving era of human–robot co-existence in terms of four categories: robot ethics, rights, policy, and safety.

3.1 Robot Ethics

Determining how robotics will emerge and evolve requires agreement on ethical issues among multiple parties, in the same manner as nuclear physics, nanotechnology, and bioengineering. Creating consensus on these issues may require a model similar to that of the Human Genome Project for the study of Ethical, Legal and Social Issues (ELSI) sponsored by the US Department of Energy and National Institutes of Health.[7] Each agency has earmarked 3–5 percent of its financial support for genome research to ethical issues. ELSI's counterpart across the Atlantic is the European Robotics Research Network (EURON), a private organization devoted to creating resources for and exchanging knowledge about robotics research.[8] To create a systematic assessment procedure for ethical issues involving robotics research and development, a EURON committee has written and published *Roboethics Roadmap* [19], a collection of articles outlining potential research pathways, and speculating on how each one might develop. Due to the rate of rapid change occurring in the technology, EURON does not promote the collection as a guideline to state-of-the-art robotics or a declaration of principles such as those emerging from Japan and Korea. Instead, the *Roadmap* is billed as a review of topics and issues aimed at those individuals and regulatory bodies that will eventually determine robot policies—legislatures, academic institutions, public ethics committees, industry groups, and the like. It is important to note that the Roadmap focuses on human centered rather than robot or artificial intelligence centered ethics, perhaps due to its "near future urgency" perspective that addresses the next decade while contemplating foreseeable long term developments. For this reason, *Roboethics Roadmap* does not consider potential problems associated with robot consciousness, free will, and emotions.

According to the *Roadmap* authors, most members of the robotics community express one of three attitudes toward the issue of roboethics:

- *Not interested*: they regard robotics as a technical field and don't believe they have a social or moral responsibility to monitor their work.
- *Interested in short-term ethical questions*: they acknowledge the possibility of "good" or "bad" robotics and respect the thinking behind implementing laws and considering the needs of special populations such as the elderly.
- *Interested in long-term ethical concerns*: they express concern for such issues as "digital divides" between world regions or age groups. These individuals are aware of the gap between industrialized and poor countries and the utility of developing robots for both.

The authors of this paper are in the third category, believing that social and/or moral questions are bound to accompany the emergence of a human–robot co-existence society, and that such a society will emerge sooner than most people believe. Furthermore, we agree with the suggestions of several *Roboethics Roadmap* authors that resolving these ethical issues will require agreement in six areas:

1. Are Asimov's Three Laws of Robotics (discussed in Sect. 5) usable as guidelines for establishing a code of roboethics?
2. Should roboethics represent the ethics of robots or of robot scientists?
3. How far can we go in terms of embodying ethics in robots?
4. How contradictory are the goals of implementing roboethics and developing highly autonomous robots?
5. Should we allow robots to exhibit "personalities"?
6. Should we allow robots to express "emotions"?

[7] http://www.ornl.gov/sci/techresources/Human_Genome/research/elsi.shtml.

[8] http://www.euron.org/.

This list does not include the obvious issue of what kinds of ethics are correct for robots. Regarding "artificial" (i.e., programmable) ethics, some *Roadmap* authors briefly touch on needs and possibilities associated with robot moral values and decisions, but generally shy away from major ethical questions. We consider this unfortunate, since the connection between artificial and human-centered ethics is so close as to make them very difficult to separate. The ambiguity of the term *artificial ethics* as used in the EURON report ignores two major concerns:

- How to program robots to obey a set of legal and ethical norms while retaining a high degree of autonomy.
- How to control robot-generated value systems—or morality.

In this article we will respectively call these *Type 1* and *Type 2* artificial ethics. Since both will be created and installed by humans, the boundary between them will be exceptionally fluid.

Visually-impaired people depend on guide dogs to navigate their living environment. Given their serious responsibility, guide dogs must absolutely obey orders given by their owners. However, the dogs received instruction in "Intelligent Disobedience" which trains the dog to act against the orders of its master in emergency cases to ensure the person's safety. Initially, the dogs were trained to make these decisions according to human-centered value systems or what we have called Type 1 artificial ethics earlier. Nevertheless, through repeated training in disobedience through various kinds of situations, the decisions of the dogs show a blending of its own value system with the inculcated human-centered value system. As such, Intelligent Disobedience is what we call Type 2 artificial ethics, in which value is not absolutely human-centered.

Susan Leigh Anderson also holds the similar ideas [20], such as:

> It might be thought that adding an ethical dimension to a machine is ambiguous. It could mean either (a) in designing the machine, building in limitations to its behavior according to an ideal ethical principle or principles that are followed by the human designer, or (b) giving the machine ideal ethical principles, or some examples of ethical dilemmas together with correct answers and a learning procedure from which it can use the principle[s] in guiding its own actions.

The South Korean government is putting the finishing touches on a *Robot Ethics Charter*; when published,[9] it may stand as the world's first official set of ethical guidelines for robotics. According to that country's Ministry of Commerce, Industry and Energy (MC-IE), the Charter will present criteria for robot users and manufacturers, and guidelines for ethical standards to be programmed into robots. The standards are being established in response to a plan announced by the Ministry of Information and Communication to put a robot in every South Korean home by 2020. The Charter's main focus appears to be social problems—for example, human control over robots and the potential for human addiction to robot interaction. However, the document will also deal with a number of legal issues, including protections for data acquired by robots and machine identification for determining responsibility distribution.

In an April 2007 presentation at an international "Workshop on Roboethics" held in Rome, an MCIE representative gave three reasons explaining why his government felt a need to write a *Robot Ethics Charter* [21]: the country's status as a testing ground for robots (similar to its experience with IT electronics); a perceived need for preparation for a world marked by a strong partnership between humans and robots; and social demands tied to the country's aging population and low birth rate. The inclusion of guidelines for the robots themselves may be interpreted as tacit acknowledgement of Asimov's Three Laws of Robotics as well as concern over the implementation of Type 1 ethics in robot control systems.

3.2 Robot Rights

There are many barriers to overcome before we can produce human-based intelligence robots capable of making autonomous decisions and having limited "self-awareness". Still, futurists who believe that such a day will come in this century are contemplating issues that might emerge. In 2006, the Horizon Scanning Centre (part of the United Kingdom's Office of Science and Innovation) published a white paper with predictions for scientific, technological, and health trends for the middle of this century [22]. The authors of the section entitled "Utopian Dream, or Rise of the Machines" raise the possibility of robots evolving to the degree that they eventually ask for special "robo-rights" [23]. In this paper we will limit our discussion to how current human legal systems, in which rights are closely tied to responsibilities, will affect early generations of non-industrial robots.

Whenever an accident occurs involving humans, the person or organization that must pay for damages can range from individuals (responsible for reasons of user error) to product manufacturers (responsible for reasons of poor product design or quality). Rights and responsibilities will need to be spelled out for two types of Next Generation Robots. The system for the first type—Next Generation Robots lacking artificial intelligence-based "self-awareness"—will be straightforward: 100 percentage human-centered, in the

[9] http://www.reuters.com/article/lifestyleMolt/idUSSEO16657120070507.

same manner that dog owners must take responsibility for the actions of their pets. In other words, robots in this category will never be given human-like rights or rights as legal entities.

The second type consists of Next Generation Robots programmed with some degree of "self-awareness", and therefore capable of making autonomous decisions that can result in damage to persons or property. Nagenborg, Capurro, Weber and Pingel [24] argue that all robot responsibilities are actually human responsibilities, and that today's product developers and sellers must acknowledge that principle when designing first-generation robots for public consumption. They use two codes of ethics—one from the Institute of Electrical and Electronics Engineers and the other from the Association of Computing Machinery—to support their view that for complex machines such as robots, any attempt to remove product responsibility from developers, manufacturers, and users represents a serious break from the human legal system norm. We may see a day when certain classes of robots will be manufactured with built-in and retrievable "black boxes" to assist with the task of attributing fault when accidents occur, since in practice it will be difficult to attribute responsibility for damages caused by a robot, especially those resulting from owner misuse. For this reason, Nagenborg et al. have proposed the following meta-regulation:

> If anybody or anything should suffer from damage that is caused by a robot that is capable of learning, there must be a demand that the burden of adducing evidence must be with the robot's keeper, who must prove her or his innocence; for example, somebody may be considered innocent who acted according to the producer's operation instructions. In this case it is the producer who needs to be held responsible for the damage.

If responsibility for robot actions ever reaches the point of being denied by humans, a major issue for legal systems will be determining "punishment". Wondering if human punishment can ever be applied to robots, Peter Asaro observes that

> they do have bodies to kick, though it is not clear that kicking them would achieve the traditional goals of punishment. The various forms of corporal punishment presuppose additional desires and fears of being human that may not readily apply to robots pain, freedom of movement, morality, etc. Thus, torture, imprisonment and destruction are not likely to be effective in achieving justice, reform and deterrence in robots. There may be a policy to destroy any robots that do harm, but as is the case with animals that harm people, it would be preventative measure to avoid future harms rather than a true punishment . . . [American law] offers several ways of thinking about the distribution of responsibility in complex cases. Responsibility for a single event can be divided among several parties, with each party assigned a percentage of the total (p. 2) [25].

If we go this route, we may need to spell out robot rights and responsibilities in the same manner that we do for such non-human entities as corporations. Will we be able to apply human-centered values to robots as we do to other entities—a core value in human legal systems? To practice "robot justice," those systems will be required to have a separate set of laws reflecting dual human–robot-centered values. Robo-responsibilities would need to be clearly spelled out.

3.3 Robot Policy

A large number of robot-related policies must be debated and enacted before the mid-century "robot in every home" era begins: labor force displacement, physical safety, supervising research and development, and the shape of robot technology (RT) marketing, among many others. The breadth of these issues makes the appearance of a single, all-encompassing robot policy unlikely. However, it is likely that governments will follow their established top-down approach to giving direction to new technologies, and free-market advocates will resist such efforts.

The Japanese are currently addressing these concerns. In 2005 the METI created the above-mentioned Robot Policy Committee and invited robotics experts to serve on it. The committee's initial report emphasized the idea that Japanese government agencies and enterprises need to cooperatively address three areas of concern when establishing a Next Generation Robots industry [26]:

1. Develop a market environment: According to a survey conducted by the Japanese Robot Association, the Next Generation Robots market is expected to expand from 3 trillion yen in 2010 to 8 trillion yen in 2025 [27]. This enormous market will require support in many forms. Two examples are training in RT-related fields and the above-mention-ed need for local governments and robot enterprises to create areas dedicated to robot research and development.[10] Whereas technical research directions in the past were determined by university labs and research institutions, the committee suggested that market forces determine future directions.
2. Ensure safety: The clarification of legislative issues pertaining to Next Generation Robots safety requires analyses of human–robot interaction responsibilities before and after the manufacturing stage. Areas of concern for

[10] http://www.f-robot.com/english/index.html.

what we will call *pre-safety regulations* include standards for robot design and production. *Post-safety regulations* will address situations in which human injury is caused by robot actions, as well as systems for product liability protection and insurance compensation.

3. Develop a mission-oriented RT system: Japanese are accustomed to making products and manufacturing systems according to available technologies. A mission-oriented RT system will emphasize technology development by private firms based on demands and needs identified by government authorities [28].

Robot policy can be viewed as an intersection in which robot rights, robot ethics, and other subfields are integrated for the purpose of generating a direction for technology development. Since an equally important function of robot policy is to serve as a reference for creating new legislation, it must become a priority well before the expected emergence of a human–robot co-existence society. Even in draft form, robot policies can support international cooperation and information exchanges to assist legislators in setting legal guidelines.

The cultural difference between East and West may result in different responses to policy decisions. While most people agreed that "Humanoid" is a positive term for robotic research in Japan, the use of this type of research tend to be associated with "Frankenstein Complex"' in western societies. Kaplan [29] states that from the Japanese perspective, this difference seems to stem from the blurring between realizations of nature and the production of man. He described this Japanese point of view as "linking beings instead of distinguishing them", whereas in western discussion of robotics, the distinction between the natural and the artificial is very significant. Norwegian writer Jon Bing [30] has analyzed several representative western literature on artificial beings and found that the writings posited three types: "Machinelike man", "Manlike machine", and the new synthesis, "Cyborg". This distance between humans and robots is always stressed in western cultures.

In addition, Fumio Harashima from Tokyo Denki University has argued that the one main difference between Japanese and American robotics research is their source of funding—U.S. robitics research is supported by U.S. Military [31], thus the relevance of Robot Policy to military application or usage may take priority in the United States. For example, a group of USJFCOM-U.S. Joint Forces Command published a study titled "Unmanned Effects: Taking the Human out of the Loop" in August 2003, which suggest that by as early as 2025, widespread use of tactical, autonomous robots by U.S. military may become the norm on the battlefield [32].

3.4 Robot Safety

In 1981, a 37-year-old factory worker named Kenji Urada entered a restricted safety zone at a Kawasaki manufacturing plant to perform some maintenance on a robot. In his haste, he failed to completely turn it off. The robot's powerful hydraulic arm pushed the engineer into some adjacent machinery, thus making Urada the first recorded victim to die at the hands of a robot. A complete review of robot safety issues will be given in Sect. 5. Here we will simply emphasize safety as the most important topic requiring detailed consideration and negotiation prior to the coming human–robot coexistence society. Based on its large body of regulations and guidelines for industrial robots, Japan is considered a leader in this area. But as the METI Robot Policy Council notes, it has no safety evaluation methods or regulations currently in place for Next Generation Robots.

4 Human-Based Intelligence

In order to adapt to unstructured environments and work more closely with humans, Next Generation Robots must be designed to act as *biomorphic robots* with specialized capabilities. Lewis and Sim define biomorphic robots as imitations of biological systems capable of predicting the sensory consequences of movement, learning through the use of neural-type methods, and exploiting "natural system dynamics to simplify computation and robot control" [33]. Current examples of biomorphic robots are snakebots [34], insectbots,[11] and humanoid robots [35]. Researchers have built several variations of ant-like robots with mechanical limbs and cameras that are capable of "exploring" mazes and using cooperative strategies similar to those of ants to move about in different environments. AI researchers are finding ways to combine sampling from explicit external phenomena ("seeing") with implicit internal activities ("thinking") to create biomorphic robots capable of facial expressions that make them appear "sociable" to humans.[12]

Neurologists view the human brain as having three layers—primitive, paleopallium, and neopallium—that operate like "three interconnected biological computers, [each] with its own special intelligence, its own subjectivity, its own sense of time and space, and its own memory" [36]. From an AI viewpoint, the biomorphic equivalents of the three layers are action intelligence, autonomous intelligence, and Human-Based Intelligence (Fig. 1). Action intelligence functions are analogous to nervous system responses that coordinate sensory and behavioral information, thereby

[11] http://www.cis.plym.ac.uk/cis/InsectRobotics/Homepage.htm.

[12] http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html.

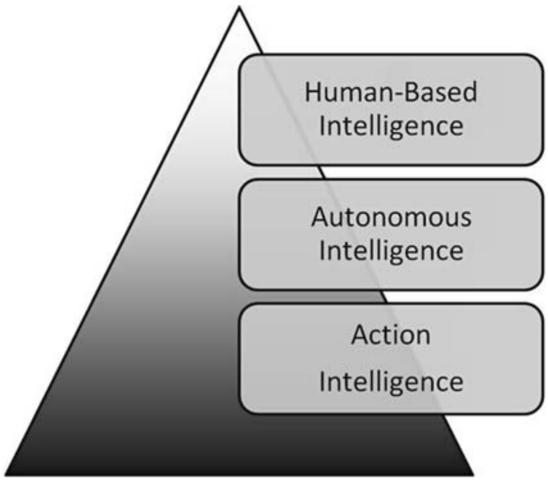

Fig. 1 Robot intelligence layers

giving a robot the ability to control head and eye movement [37], move spatially [38], operate machine arms to manipulate objects [39], and visually inspect its immediate environment [40]. Autonomous intelligence refers to capabilities for solving problems involving pattern recognition, automated scheduling, and planning based on prior experience [41]. Such behaviors are logical and programmable, but not conscious.

We are currently in a developmental period bridging action intelligence and autonomous intelligence, with robots such as AIBO,[13] QRIO [42], and Roomba[14] on the verge of being lab tested, manufactured, and sold. These simple and small robots are strong indicators of Next Generation Robots potential and the coming human–robot co-existence age. Even as "pioneer" robots, they have remarkable abilities to perform specific tasks according to their built-in autonomous intelligence—for instance, AIBO and QRIO robots have been programmed to serve as companions for the elderly, and Roomba robots as housecleaners. However, they cannot make decisions concerning self-beneficial actions or decide what is right or wrong based on a sense of their own value.

At the third level is *Human-Based Intelligence*—higher cognitive abilities that allow for new ways of looking at one's environment and for abstract thought, also referred to as "mind" and "real intelligence". Since a universally accepted definition of human intelligence has yet to emerge, there is little agreement on a definition for Human-Based Intelligence. Many suggestions and predictions appear to borrow liberally from science fiction, such as Human-Based Intelligence robots forming a new species with the long-term potential of gaining power over humans [43, 44]. In real-world contexts, researchers are experimenting with ways of combining action intelligence, autonomous intelligence, and human-based intelligence to act more human-like and to "comprehend complex ideas, learn quickly, and learn from experience" [45]. Human-Based Intelligence research started in the 1950s—roughly the same time as research on artificial intelligence, with which human-based intelligence is closely associated. One of the earliest and most famous efforts to examine human-based intelligence potential consisted of what is now known as the "Turing Test" [46]. Taking a behaviorist perspective, Alan Turing defined human intelligence as the ability "to respond like a human being," especially in terms of using natural language to communicate. There have been many efforts at creating programs that allow robots to respond like humans [47], but no AI program has ever passed the Turing test and been accepted as a true example of Human-Based Intelligence [48].

The legal and robot sociability problem issues that will arise over the next few decades are intricately linked with artificial intelligence, which was originally conceived as "the science and engineering of making intelligent machines, especially intelligent computer programs" [49]. Currently the two primary focuses of AI research are *conventional* (or symbolic) and *computational*; since intelligence still has such a broad definition, the two handles separate human-based intelligence parts. Conventional AI, which entails rational logical reasoning based on a system of symbols representing human knowledge in a declarative form [50], has been used for such applications as chess games (reasoning) [51], conversation programs (text mining),[15] and for organizing domain-specific knowledge (expert systems) [52]. While conventional AI is capable of limited reasoning, planning, and abstract thinking, researchers acknowledge that the use of symbols does not represent "mindful" comprehension, and is limited in terms of learning from experience [53].

Computational (non-symbol) AI [54] mimics natural (e.g., genetic [55] or neural [56]) learning methods, and allows for learning and adaptation based on environmental information in the absence of explicit rules—an important facility for living creatures. Computational AI has advantages in terms of overcoming noise problems, working with systems that are difficult to reduce to logical rules, and especially for performing such tasks as robot arm control, walking on non-smooth surfaces, and pattern recognition. However, as proven by chess programs, computational AI is significantly weaker than conventional AI in thinking abstractly and following rules. Among researchers in the fields of robotics and AI, the majority believes in the inevitability of human-based intelligence becoming a reality following breakthroughs in computational AI [57]. Others argue that

[13] http://support.sony-europe.com/aibo/.

[14] http://www.irobot.com/sp.cfm?pageid=122/.

[15] A.L.I.C.E. AI Foundation. Alicebot and AIML Documentation, http://www.alicebot.org/documentation/.

computational and conventional AI are both examples of behaviorism, and therefore will never capture the essence of human-based intelligence [58]. They claim that reaching that goal requires the development of an entirely new framework for understanding intelligence [59].

Optimistic or not, the belief that human-based intelligence robots will someday become a reality means that researchers must consider Human-Based Intelligence when predicting future robot safety and legal issues. They may conclude—as does Shigeo Hirose of the Tokyo Institute of Technology—that a prohibition on Human-Based Intelligence is necessary. Hirose is one of a growing number of researchers and robot designers resisting what is known as the "humanoid complex" trend [60], based on his adherence to the original goal of robotics: to invent useful tools for human use [61]. Alan Mackworth, past president of the Association for the Advancement of Artificial Intelligence [62], frames the robot Human-Based Intelligence issue as "should or shouldn't we" as oppose "can or can't we". Mackworth emphasizes the idea that goal-oriented robots do not require what humans refer to as "awareness", and therefore challenges the idea that we need to create human-based intelligence for machines.

In "ROBOT: Mere Machine to Transcendent Mind" [63], Carnegie Mellon Robotics Institute professor Hans Moravec predicts that robot intelligence will "evolve" from lizard-level in 2010 to mouse-level in 2020, to monkey level in 2030, and finally to human level in 2040—in other words, some robots will strongly resemble first-existence entities by mid-century. If true, future legislators interested in creating robot-related laws must face the difficult task of maintaining a balance between human and robots that will win broad acceptance from their constituents. Our motivation for this research is to give examples of robotics issues that future legislators and policy makers will have to address.

First, we will have to respond to standard societal suspicions about new technology, as exemplified by people's reactions to the Woosung Road in China, the first railway built in 1876 from Shanghai to Woosung. The speedy and powerful locomotive was seen as a monster by the Chinese that time, and finally, as a result of boycott, the Woosung Road was closed by the government and all the railway systems were transferred to Taiwan in 1877. This situation is very similar to *Uncanny Valley*, in which Masahiro Mori introduced the hypothesis in 1970 that human observers will respond with horror when faced with robots and other facsimiles of humans that look and act like actual humans. It now looks as though such suspicions and fears will be much less than what Mori predicted, but people may still express apprehension over blurred boundaries between humans and robots unless acceptable Robot Ethics guidelines are established.

In an earlier section we discussed the idea that robot responsibility should be regarded as human-owner responsibility. If we allow Human-Based Intelligence robo-ts to be manufactured and sold, the potential for any degree of robot self-awareness means dealing with issues such as punishment and a shift from human-centered to human–robot dual values. This is one of the most important reasons why we support a ban on installing Human-Based Intelligence software in robotsperhaps permanently, but certainly not until policy makers and robotists agree on these issues.

We also believe that creating Type 1 robots, in other words, "programming robots to obey a set of legal and ethical norms while retaining a high degree of autonomy requires agreement on human-centered ethics based on human values. The challenge is integrating human legal norms into robots so that they become central to robot behavior. The most worrisome issue is the potential capability of late-generation Human-Based Intelligence robots with significant amounts of self-awareness to generate their own values and ethics what we call Type 2 artificial ethics. Implementing Type 2 robot safety standards means addressing a long list of uncertainties for machines capable of acting outside of human norms. We are nowhere near discussing-let alone implementing policies for controlling Human-Based Intelligence robot behavior, since we are very far from having Human-Based Intelligence robots as part of our daily lives. However, if the AI/Human-Based Intelligence optimists are correct, the high risk of Human-Based Intelligence robots will necessitate very specific guidelines.

A guiding principle for those guidelines may be categorizing robots as *Third Existence* [64] entities, neither living/biological (first existence) nor non-living/non-biological (second existence). As described by Waseda University's Shuji Hashimoto, third existence machines will resemble living beings in appearance and behavior, but they will not be self-aware. We think this definition overlooks an important human–robot co-existence premise: most Next Generation Robots will be restricted to levels of autonomous intelligence that fall far short of Human-Based Intelligence, therefore their similarities with humans will be minor. As Cynthia Breazeal from the MIT Personal Robots Group observes:

> There's a "fuzzy boundary" that's very compelling for us, where we are willing to see robots as not human, but not exactly machine either [65].

According to the current legal system, robots are second-existence human property, a status that may be inadequate for the semi-autonomous Next Generation Robots that are about to enter people's homes and businesses especially in terms of responsibility distribution in the case of accidents. Asaro therefore proposes the creation of a new legal status for robots as "quasi-persons" or "corporations," while Nugenborg prefers emphasizing the point we made earlier about robot owners being responsible for their robots' actions in the same manner as pet owners. In Nugenborg's view, robots should be given a legal status somewhere between personality and property.

Third existence status for robots may be an acceptable way of avoiding threats posed and impermanent caused by the society, law, and technology. But if Moravec's prediction comes true, the day will come when Human-Based Intelligence robots are a reality, and at that time we will be forced to decide between strictly following third existence guidelines or completely redefining the societal role and status of Human-Based Intelligence robots. If we choose the first response, then we must ban Human-Based Intelligence. However, legal scholars currently looking at an unknown future may yet find a way to make the second response work.

5 Safety Intelligence

In terms of safety standards, the primary difference in risk between industrial robots and autonomous Next Generation Robots is that the first involves machine standards and the second a mix of machine standards and open texture risk from unpredictable interactions in unstructured environments. *Open-Texture Risk* [66]—regarding language, any term in a natural language has a central (core) meaning, but the open texture character of language allows for interpretations that vary according to specified domains, points of view, time periods, etc. The open texture character of language produces uncertainty and vagueness in legal interpretations. Risk assessment associated with Next Generation Robots autonomous behavior faces a similar dilemma in that a core meaning exists, but the range of that core is difficult to clearly define, resulting in what we refer to as open texture risk. In a May 2006 paper on legislative issues pertaining to Next Generation Robots safety, Japanese METI committee members describe the difference in terms of *pre- and post-human–robot interaction responsibilities*. In the following discussion we will refer to them as *pre- and post-safety regulations*.

For industrial robots, safety and reliability engineering decisions are guided by a combination of pre-safety (with a heavy emphasis on risk assessment) and post-safety regulations (focused on responsibility distribution). Pre-safety rules include safeguards regarding the use and maintenance of robot systems from the design stage (e.g., hazard identification, risk assessment) to the training of robot controllers. One example of this is the United Kingdom Health and Safety Executive Office's 2000 publication of a set of industrial robot safety guidelines during installation, commissioning, testing, and programming.[16] Another example is International Standardization Organization (ISO) rules-especially ISO 10218-1:2006, which covers safety-associated design, protective measures, and industrial robot applications. In addition to describing basic hazards associated with robots, ISO rules are aimed at eliminating or adequately reducing risks associated with identified hazards. ISO 10218-1:2006 spells out safety design guidelines (e.g., clearance requirements) that extend ISO rules covering general machine safety [67] to industrial robot environments. Those rules address safety-related parts of control systems and software design, but since the primary focus is on robot arms and manipulators [68], they have limited application to Next Generation Robots.

	Safety Design by Risk Assessment	Safety Intelligence
Risk	Risk of Machine	Risk of Autonomous Behavior (The Open Texture Risk)
Limit	Machine's Standard	Robot's Intelligence Architecture
Effect	Decrease the Risk	Avoid Some Dangerous Behavior

Fig. 2 A comparison of safety regulation methods

Designed and constructed according to very specific standards, industrial robots are limited to performing tasks that can be reduced to their corresponding mechanisms—in other words, they cannot alter their mechanisms to meet the needs of changing environments. Therefore, the primary purpose for performing industrial robot risk assessments is to design mechanisms that match pre-approved safety levels (Fig. 2). Complex Next Generation Robots motions, multi-object interactions, and responses to shifts in environments resulting from complex interactions with humans cannot be reduced to simple performance parameters. Next Generation Robots and future Human-Based Intelligence designers and manufacturers must instead deal with unpredictable hazards associated with the legal concepts of *core meaning* and *open texture risk*. Any term in a natural language has a core (central) meaning, but the open texture characteristic of human language [69] allows for interpretations that vary according to specified domains, points of view, time periods, and other factors, all of which can trigger uncertainty and vagueness in legal interpretations. Autonomous Next Generation Robots designers and programmers must therefore clearly define a core meaning plus an acceptable and useful range of that core.

The inherent unpredictability of unstructured environments makes it virtually impossible that we will ever see a fail-safe mechanism that allows autonomous robots to solve all open-texture problems. Consequently, Next Generation Robots safety regulations will require a mix of pre-safety

[16] http://products.ihs.com/Ohsis-SEO/113985.html.

and post-safety mechanisms, the first using a robot's AI reasoning content to eliminate most risk, and the second entailing a product liability system to deal with accidents that do occur. A clear security issue will be limiting the "self-control" of Next Generation Robots while still allowing them to perform designated tasks. As one *Roboethics Roadmap* author succinctly states, "operators should be able to limit robot autonomy when the correct robot behavior is not guaranteed." Giving operators this capability requires what we will call *Safety Intelligence*—that is, a system of artificial intelligence restrictions whose sole purpose is to provide safety parameters when semi-autonomous robots perform their tasks. Researchers have yet to agree on a foundation for a Safety Intelligence system, but the most frequently mentioned during the earliest stages of this discussion were the "Three Laws of Robotics" established by Isaac Asimov in his science fiction novel, I, Robot [70]:

1. First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. Second Law: A robot must obey orders given it by human beings, except when such orders conflict with the First Law.
3. Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

The first two laws represent a human-centered approach to Safety Intelligence that agrees with the current consensus of Next Generation Robots designers and producers. As robots gradually take on greater numbers of labor-intensive and repetitious jobs outside of factories and workplaces, it will become increasingly important for laws and regulations to support Safety Intelligence as a "mechanism of human superiority" [71]. The third law straddles the line between human- and machine-centered approaches. Since the purpose of robot functionality is to satisfy human needs, they must be designed and built in a manner so as to protect themselves as human property, in contrast to biological organisms that protect themselves for their own existence. As one magazine columnist has jokingly suggested, "A robot will guard its own existence … because a robot is bloody expensive" [72].

In his introduction to another work of fiction, *The Rest of the Robots*, Asimov wrote, "There was just enough ambiguity in the Three Laws to provide the conflicts and uncertainties required for new stories, and, to my great relief, it seemed always to be possible to think up a new angle out of the 61 words of the Three Laws" [73]. While those ambiguities may be wonderful for writing fiction, they stand as significant roadblocks to establishing workable safety standards for complex Next Generation Robots. In *Roboethics Roadmap*, some contributing authors note that the Three Laws raise many questions about Next Generation Robots programming:

- Which kinds of ethics are correct and who decides?
- Will roboethics really represent the characteristics of robots or the values of robot scientists?
- How far can and should we go when we program ethics into a robot?

Other robot researchers argue that Asimov's laws and the South Korean charter discussed in Sect. 3 still belong to the realm of science fiction because they are not yet applicable. Hiroshi Ishiguro of Osaka University, the co-creator of two female androids named Repliee Q1 and Repliee Q2 [74], believes it would be a mistake to accept Asimov's laws as the primary guiding principle for establishing robot ethics:

> If we have a more intelligent vehicle [e.g., automobile], who takes responsibility when it has an accident? We can ask the same question of a robot. Robots do not have human-level intelligence [75].

Mark Tilden, the designer of a toy-like robot named RoboSapien, says "the problem is that giving robots morals is like teaching an ant to yodel. We're not there yet, and as many of Asimov's stories show, the conundrums robots and humans would face would result in more tragedy than utility." Ian Kerr, law professor at the University of Ottawa, concurs that a code of ethics for robots is unnecessary:

> Leaving aside the thorny philosophical question of whether an AI could ever become a moral agent, it should be relatively obvious from their articulation that Asimov's laws are not ethical or legal guidelines for robots but rather about them. The laws are meant to constrain the people who build robots of exponentially increasing intelligence so that the machines remain destined to lives of friendly servitude. The pecking order is clear: robots serve people [76].

Currently, the two primary perspectives on the mix of AI and safety are either creating artificial agents with safety-oriented reasoning capabilities, or programming robots with as many rules as required for ensuring the highest level of safe behavior. Which perspective wins out will depend on how policy makers, designers, and manufacturers address as the *Three Questions For Three Laws of Robotics* we proposed following:

Question of Machine Meta-ethics: Susan Leigh Anderson [20] argued that Asimov's Three Laws of Robotics are an unsatisfactory basis for machine ethics, regardless of the status of the machine. She divided the robots into the robots with moral standings and without moral standings. First, she claimed that if the robots have moral standing then an ethical theory must take the being into account, then she introduced Warren's lists of six characteristics to define personhood (moral standing), such as Sentience, Emotionality, Reason, The capacity to communicate, Self-Awareness, Moral Agency [77], by Warren's definition the robots with Human-Based Intelligence could be seem with moral standings, and it's immoral to forced robots with Human-Based Intelligence to obey the Asimov's Three Laws of Robotics to "serve people" such as Ian Kerr said earlier. As the robots without moral standings, Anderson introduced Immanuel Kant's consideration that "humans should not mistreat the entity in question, even though it lacked rights itself" [78], he argued that even though animals lack moral standing and can be used to serve the end of human beings, we should still not mistreat them because he said "he who is cruel to animals becomes hard also in his dealings with men". If the Third Existence robots are adopted to Asimov's three laws of robotics then it's allowed to let people bully or mistreat a robot. Therefore the three laws of robotics is inadequate whether robot with or without moral standing.

Question of Formality: The ability to "think abstractly" is uniquely human, and there is no way of being absolutely sure of how robots will interpret and react to the abstract meanings and vague terms used in human communication. For example, humans know how to distinguish between blood resulting from a surgical operation and blood resulting from acts of violence. Making such distinctions requires the ability to converse, to understand abstract expressions (especially metaphors), and to use domain knowledge to correctly interpret the meaning of a sentence. There are many examples that illustrate just how difficult this task is; one is Chomsky's famous sentence showing the inadequacy of logical grammar, "Colorless green ideas sleep furiously" [79], and another is Groucho Marx's line, "Time flies like an arrow, fruit flies like a banana".[17] Such examples may explain Asimov's description of robots as "logical but not reasonable" [80]. Therefore Asimov's Laws are facing a challenge for its "Formality" or "media to access the legal content", for people, we are used to let nature language be the media to access the content of law, however excepted for the Human-Based Intelligence robots the next generation robots are lacking the abstract ability to using nature language like human being in daily life.

Question of Regulation: Here the major issue is deciding whether or not Next Generation Robots need doctrinal reasoning powers. In other words how could we ensure that Next Generation Robots could enforce the three laws fully according such norms that human defined. In the early part we have already mentioned the artificial ethics Type 1 and Type 2. If we allow autonomous robots to define their own concepts of "safety", that means giving them the power to decide both when and how to react to stimuli. At some point those decisions will require artificial ethical and morality reasoning—the ability to distinguish between right

[17] http://www.quotationspage.com/quote/26.html.

	Machine Meta-ethics	Formality	Regulation
Human-Based Intelligence Robots	X	O	X
Third Existence Robots	X	X	O

Fig. 3 The Three Questions for Three Laws of Robotics

and wrong. When considering "Morality Engineering", robotists such as Shigeo Hirose argue that in conflicts involving doctrinal reasoning and morality, the Three Laws may become contradictory or at risk of being set aside in favor of human requirements [81]. Using an extreme example, robots could be programmed to commit homicide under specific circumstances based on the wishes of a human majority. This example touches on two fears that many people have when they consider autonomous robots: they are troubled by the idea of letting robots obey rules that are impossible to express legislatively, and fearful of letting them defend laws established by imperfect humans. Human-Based Intelligence robots with Type 2 artificial ethics blends its own values while interpreting the three laws, thus causing ambiguity within itself: should it abide by human law or its own robot law?

In an earlier section we concluded that Asimov's Three Laws of Robotics would be difficult to put into practice to achieve real-world Safety Intelligence. There are three reasons why those laws cannot be used with Human-Based Intelligence robots. First, machine meta-ethics are considered moral entities, and it would be immoral to force machines to obey the three laws in order to serve humans as Kant argued, people cannot mistreat beings that do not have moral standing. Second, in terms of regulation the Three Laws are unsuitable because Type 2 artificial ethics pose significant incentives for robots to be law-abiding. Third, regarding formality, the Three Laws cannot be applied to entities that lack the ability to think abstractly or to use human legal language competently.

Asimov's Three Laws of Robots have proven useful in giving robotists an early framework for discussing issues tied to robot behavior. However, they are ultimately unsatisfactory for any safety regulation model that emphasizes Safety Intelligence during design stages. Robotists must therefore search for a new approach to address these complex issues.

6 Legal Machine Language

The legal architecture of Human-Next Generation Robots interaction (including legal positions, usability, and content) is one of the most important issues for establishing the legal regulation of robots. It is hard to predict what the final solution will look like because of the number of open questions that remain. However, some existing concepts will affect the form of the legal architecture that eventually emerges. Here we will describe our proposal for an alternative *Legal Machine Language* based on two principles: *Code is Law* and *Embedded Ethics*.

Lawrence Lessig submitted "Code is Law" in 1998 [82], and he has noted that behavior is regulated by four kinds of constraints, such as Law, Social norms, the Market, and Nature—or what he called "Architecture". Since the Architecture of Cyberspace is absolutely built by code, the code could be an useful constraint to regulate user behavior or preserve crucial values such as Freedom or Privacy in Cyberspace. Therefore, using the code to regulate Cyberspace becomes another possible way. In the same reason the social control of Next Generation Robots should not be limited to "Dog Law" model, adopting the law described by human legal language in human society, is directly obeyed by human and indirectly obeyed by others such like robots. Next Generation Robots as "Virtual Agents into the Real World" is code-based artificial entities itself. The code or the architecture of Next Generation Robots could also be a regulator for its autonomy if we can define, formalize, and implement safety action without the need of moral reasoning.

Nature gives us many examples of animals interacting safely without complex moral judgments [83]. For example, flocking [84], as a common demonstration of emergent behavior for group of creatures such as birds or fishes could be seen as nature's traffic rules. When birds of the same species migrate, their shared genetic background allows them to fly in close or V-shaped formations without colliding—an important feature for shared and individual survival. Safe interaction requires adherence to a set of simple non-verbal safety rules shared by all members of the population: avoid crowding neighboring birds, fly along the same heading, or fly along the same average heading as neighboring birds. This zoological concept of flocking is considered useful for controlling unmanned aircraft [85] and other machines [86] including Next Generation Robots [87–90]. In an earlier section we discussed that a combination of action intelligence and autonomous intelligence is sufficient to recognize situations, avoid misunderstandings, and prevent accidents without processing ethics, performing morality reasoning functions, and making right/wrong decisions. Therefore robots only need to handle safety interaction and solely applied moral precept from human. In Ronald C. Arkin's words,

> We do not want the agent to be able to derive its own beliefs regarding the moral implications..., but rather to be able to apply those that have been previously derived by humanity [91].

Arkin has provided a possible framework embedded ethics into robots without Asimov's Three Laws of Robotics [91]. In his framework, there are (a) Ethical Behavior Control / Ethical Governor, (b) Human Robot Interface, and (c) Responsibility Advisor; the three components work cooperatively to form an ethical autonomous agent. First of all, Ethical Behavior Control / Ethical Governor Components, working as ex post facto suppression of unethical behavior, could be viewed as genetically built-in reflex system inside humans and animals. Ethical Behavior Control / Ethical Governor suppresses, restricts, or transforms unethical behavior and trigger protective behaviors by deliberative/reactive trigger protective behaviors. Therefore autonomous robot would have the ongoing ability to assess changing situations accurately and to correctly respond to complex real-world conditions. Second, Human Robot Interface, including body language, gesture [92], simple command, facial expression [93] and construct language like Loglan (identified as potentially suitable for human-computer communication due to its use of predicate logic, avoidance of syntactical ambiguity, and conciseness) gives both robot and human ability to be aware current situation. Human Robot Interface is used to prevent misunderstanding, predict influence, and consider possible corrective action [94]. Human Robot Interface design patterns should be defined as clear and explicit as possible, thereby Next Generation Robots could take immediate protective reactions in human-predictable ways as to mitigate risks tied to language-based misunderstandings or unstable autonomous behaviors. Third, Responsibility Advisor defined as "a mechanism in support of identifying and advising operators regarding the ultimate responsibility for the deployment of such a system" [91]. Responsibility Advisor, as a component of legal architecture notices each unethical behavior due to either human operator's override or autonomous robot's representational deficiency. Either by giving robot rights to refuse an unethical order or by limiting human to use robot ethically, we could define an explicit interaction rule set and a legal architecture that can be applied to all kinds of Next Generation Robots, one that accommodates the needs of a human–robot co-existence society in terms of simplicity and accountability.

In its current form, our proposal emphasizes three components of embedded ethics of Human–Next Generation Robots interaction: (a) the ongoing ability to assess changing situations accurately and to correctly respond to complex real-world conditions; (b) immediate protective reactions in human-predictable ways so as to mitigate risks tied to language-based misunderstandings or unstable autonomous behaviors; and (c) an explicit interaction rule set and a legal architecture that can be applied to all kinds of Next Generation Robots. Unlike the Three Law of Robotics, these three components, could be encoded by code and embedded directly inside autonomous intelligent of Next Generation Robots. As Lessig says "Architecture structures and constrains social and legal power, to the end of protecting fundamental values" [95].

Legal Machine Language could be a possible way for law regulation on Next Generation Robots' Open-Texture Risk, however in order to achieve the three criteria we mentioned earlier, the cross-fields conversation between Law and Robotics is necessary, at present the two characters of Legal Machine Language—"Code is Law" and "Embedded Ethics" provide a chance to review what's the adequate formality of law and how to implement the legal value literally obeyed by Next Generation Robots under the basis of Human–Robot Interaction in an environment that human and robots co-exist.

7 Conclusion

Emerging trends associated with Next Generation Rob-ots point to the day when robots will enter human society in large numbers, while engineers address all kinds of technical issues, a mix of engineers, social scientists, legal scholars, and policy makers will be making important decisions regarding robot sociability. In all cases, one of the priority concerns must be robot safety, since the emphasis for the future will be on human–robot *Co-Existence*.

In this paper we described a Safety Intelligence concept that can be separated into two dimensions. The first involves ethics—a special "Third Existence" status for robots and a recommended ban on equipping Next Generation Robots with Human-Based Intelligence. The second involves a mix of third existence designation and a Legal Machine Language designed to resolve issues associated with Open-Texture Risk. An important task for researchers is determining the structure and details of a Legal Machine Language part of an emerging field of legal research that we refer to as *Robot Legal Studies*.

References

1. Unsigned Editorial (2004) Toward 2025 Human-Robot Co-existence Society: The next generation robot vision seminar report. Japanese Ministry of Economy, Trade and Industry (METI), Tokyo
2. Ge SS (2007) Social robotics: The integration of advances in engineering and computer science. In: The 4th annual international conference organized by Electrical Engineering/Electronics, Computer, Telecommunication and Information Technology (ECTI) Association, Chiang Rai, Thailand, May 9–12, 2007. Keynote Speaker
3. Unsigned Editorial (2007) Long-term strategic guidelines "Innovation 25" (unofficial translation) at 60. Cabinet Office, Government of Japan. Available via www, http://www.cao.go.jp/innovation/index.html, Accessed Sep 25 2008
4. Lovgren S (2006) A robot in every home by 2020, South Korea Says. In: National Geographic News. Available via www, http://news.nationalgeographic.com/news/2006/09/060906-robots.html, Accessed Feb 12 2009
5. Etengoff A (2008) US military to develop 'ethical' robots. In: ITEXAMINER. Available via www, http://www.itexaminer.com/us-military-to-develop-ethical-robots.aspx, Accessed December 4 2008
6. Inoue H, Hirukawa H (2001) HRP: Humanoid robot project of MITI. J Robotics Soc Jpn 19(1):2–7 (in Japanese)
7. Wabot-House Lab (2004) Wabot-House Lab annual report for Heisei 15 Year 1. Waseda University, Tokyo (in Japanese)
8. Cabinet Secretariat, Office for the Promotion of Special Zones for Structure Reform (2003) Special zones for structure reform. Japan Government. Available via www, http://www.kantei.go.jp/foreign/policy/kouzou2/sanko/030326setumei_e.pdf, Accessed Sep 25 2008
9. Sugano S (2002) Wabot-House project. In: Advances in science, technology and environmentology (ASTE), Annual Report of RISE (Research Institute of Science and Technology), Waseda University, Tokyo at A10 (in Japanese)
10. Inaba A, Chihara K (2004) The robot project 21. J Robotics Soc Jpn 22(7):818–821 (in Japanese)
11. Miyamoto K (2004) Viewpoint on robot rescue development and support by Kanagawa Prefecture. J Robotics Soc Jpn 22(7):827–828 (in Japanese)
12. Asada M (2005) From synergistic intelligence to RoboCity CoRE. J Robotics Soc Jpn 23(8):942–945 (in Japanese)
13. Cerf VG (2007) The disruptive power of networks. In: Forbes Asia, May-7-2007 Issue, pp 76–77
14. Gibson JJ (1977) The theory of affordances. In: Shaw R, Bransford J (ed) Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale, NJ, pp 67–82
15. Sasaki M (2006) An introduction to the theory of affordances. J Robotics Soc Jpn 24(7):776–782 (in Japanese)
16. Levy D (2007) Love and sex with robots: The evolution of human–robot relationships. Harper Collins, New York
17. IPCC (2001) IPCC third assessment report: Climate change 2001. Available via www, http://www.ipcc.ch/ipccreports/assessments-reports.htm, Accessed at Feb 12 2009
18. Levy D (2006) Robots unlimited: Life in a virtual age. AK Peters (ed). Wellesley, MA
19. Unsigned Editorial (2006) Roboethics roadmap release 1.1. European robotics research network. Available via www, http://www.roboethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf, Accessed at Feb 12 2009
20. Anderson SL (2008) Asimov's "three laws of robotics" and machine metaethics. AI Soc 22(4):477–493
21. Shim HB (2007) Establishing a Korean robot ethics charter. In: IEEE ICRA workshop on roboethics, April 14 2007, Rome, Italy. Available via www, http://www.roboethics.org/icra2007/contributions/slides/Shim_icra%2007_ppt.pdf, Accessed at Feb 12 2009
22. Unsigned Editorial (2006) Robots could demand legal rights. In: BBC News. Available via www, http://news.bbc.co.uk/2/hi/technology/6200005.stm, Accessed at Feb 12 2009
23. Unsigned Editorial (2006) Robot rights? It could happen, U.K. government Told. In: CBC news. Available via www, http://www.cbc.ca/technology/story/2006/12/21/tech-freedom.html, Accessed at Feb 12 2009
24. Nagenborg M, Capurro R, Weber J, Pingel C (2008) Ethical regulations on robotics in Europe. AI Soc 22(3):349–366
25. Asaro P (2007) Robots and responsibility from a legal perspective. In: IEEE ICRA'07 workshop on roboethics, April 14, 2007, Rome, Italy. Available via www, http://www.peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf, Accessed at Feb 12 2009

26. Robot Policy Council (2005) Robot policy middle report – May 2005 version. Japan Ministry of Economy, Trade and Industry (in Japanese)
27. RIDC Council (2003) Founders' Statement. Robotics Industry Development Council. Available via www, http://www.f-robot.com/tokku/tokku.html, Accessed at Feb 12 2009
28. Robot Policy Council (2006) Robot Policy Council report – May 2006 version. Japan Ministry of Economy, Trade and Industry (in Japanese)
29. Kaplan F (2004) Who is afraid of the humanoid? Investigating cultural differences in the acceptation of robots. Int J Humanoid Robotics 1(3):465–480
30. Bing J (2008) The riddle of the robots. J Int Commer Law Technol 3(3):197–206
31. Harashima F (2007) Intelligent mechatronics systems go around with human. J Robotics Soc Jpn 25(1):20–24 (in Japanese)
32. U.S. Joint Forces Command (2003) Military robots of the future. In: About.com: US military. Available via www, http://usmilitary.about.com/cs/weapons/a/robots.htm, Accessed at Feb 12 2009
33. Lewis MA, Sim LS (2001) Certain principles of biomorphic robots. Auton Robots 11(3):221–226
34. Tanev I, Ray T, Buller A (2004) Evolution, robustness, and adaptation of sidewinding locomotion of simulated snake-like robot. Genet Evol Comput 3102:627–639
35. Faiola A (2005) Humanoids with attitude: Japan embraces new generation of robots. In: The Washington Post. Available via www, http://www.washingtonpost.com/wp-dyn/articles/A25394-2005Mar10.html, Accessed at Feb 12 2009
36. MacLean PD (1990) The triune brain in evolution: role in paleocerebral functions. Springer, London
37. Hager GD, Chang WC, Morse AS (1995) Robot hand-eye coordination based on stereo vision. IEEE Control Syst Mag 15(1):30–39
38. Lewis MA, Fagg AH, Solidum A (1992) Genetic programming approach to the construction of a neural network for control of a walking robot. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp 2618–2623
39. Chapin JK, Moxon KA, Markowitz RS, Nicolelis MAL (1999) Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nature Neurosci 2:664–670
40. Dickmanns ED (1988) Dynamic computer vision for mobile robot control. In: Proceedings of the 19th international symposium industrial robots, pp 314–327
41. Koditschek DE (1989) Robot planning and control via potential functions. Robotics Rev 1:349–367
42. Yeomans M (2005) Qrio dances Into spotlight at Carnegie Mellon Center. In: Pittsburgh Tribune-Review. Available via www, http://www.pittsburghlive.com/x/pittsburghtrib/s_298114.html, Accessed at Feb 12 2009
43. Neisser U, Boodoo G, Bouchard Jr TJ et al (1996) Intelligence: Knowns and unknowns. Am Psychol 51(2):77–101
44. Asimov I (1976) Bicentennial man. Ballantine, New York
45. Gottfredson LS (1997) Mainstream science on intelligence. Intelligence 24:13–23
46. Turing A (1950) Computing machinery and intelligence. Mind 1:433–460
47. Unsigned Editorial (2007) Emotion robots learn from people. In: BBC News. Available via www, http://news.bbc.co.uk/2/hi/technology/6389105.stm, Accessed at Feb 12 2009
48. Saygin AP, Cicekli I, Akman V (2000) Turing test: 50 years later. Minds Mach 10(4):463–518
49. McCarthy J (2007) What is artificial intelligence? In: John McCarthy's home page. Available via www, http://www-formal.stanford.edu/jmc/whatisai/whatisai.html, Accessed at Feb 12 2009
50. Newell A, Simon HA (1995) GPS: A program that simulates human thought. In: Computers and thought. MIT, Cambridge, pp 279–293
51. Hsu FH, Campbell M, Hoane Jr AJ (1995) Deep blue system overview. In: Proceedings of the 9th international conference on supercomputing, pp 240–244
52. Lederberg JD (1987) How dendral was conceived and born. In: Proceedings of the 1987 Association for Computing Machinery Conference on history of medical informatics, pp 5–19
53. Dreyfus HL, Dreyfus SE (1986) From Socrates to expert systems: The limits and dangers of calculative rationality. In: Mitcham C, Huning A (eds) Philosophy and technology II: Information technology and computers in theory and practice. Springer, New York, pp 111–130
54. Engelbrecht AP (2002) Computational intelligence: An introduction. Wiley, New York
55. Mitchell M (1996) An introduction to genetic algorithms. MIT, Cambridge
56. Abdi H (1994) A neural network primer. J Biol Syst 2:247–281
57. Warwick K (2004) March of the machines. University of Illinois Press, Chicago
58. Penrose R (1989) The emperor's new mind. Oxford University Press, New York
59. Hawkins J, Blakeslee S (2006) On intelligence. Yuan-Liou, Taipei (in Chinese)
60. Tajika N (2001) The future astro boy. ASCOM, Tokyo (in Japanese)
61. Kamoshita H (2005) The present and future of robots. X-media, Tokyo (in Japanese)
62. Unsigned Editorial (2007) What is a robot? In: CBC News. Available via www, http://www.cbc.ca/news/background/tech/robotics/definition.html, Accessed at Feb 12 2009
63. Moravec H (1999) ROBOT: Mere machine to transcendent mind. Oxford University Press, New York
64. Hashimoto S (2003) The Robot generating an environment and autonomous. In: Ojima T, Yabuno K (eds) The book of wabot 2. Chuko, Tokyo (in Japanese)
65. Lillington K (2008) So robots are social animals after all. In: Irish Times. Available via www, http://www.irishtimes.com/newspaper/finance/2008/1128/1227739081265.html, Visited Feb 12, 2009
66. Weng YH, Chen CH, Sun CT (2007) The legal crisis of next generation robots: On safety intelligence. In: Proceedings of the 11th international conference on artificial intelligence and law (ICAIL'07), Stanford, CA, USA. ACM, New York, pp 205–209
67. International Organization for Standardization (2006) ISO 13849-1:2006 Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design. Available via www, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34931, Accessed at Feb 12 2009
68. International Organization for Standardization (1994) ISO 8373:1994 Manipulating industrial robots – Vocabulary. Available via www, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=15532, Accessed at Feb 12 2009
69. Lyons D (1999) Open texture and the possibility of legal interpretation. Law Philos 18(3):297–309
70. Asimov I (1950) Robot. Gnome Press, New York
71. Fodor JA (1987) Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In: Pylyshyn Z (ed) The robot's dilemma: The frame problem in artificial intelligence. Kluwer, Hingham
72. Langford D (2006) It's the law. In: SFX 146. Available via www, http://www.ansible.co.uk/sfx/sfx146.html, Accessed at Feb 12 2009
73. Asimov I (1964) The rest of the robots. Collins, New York
74. Whitehouse D (2005) Japanese develop "female" android. In: BBC News. Available via www, http://news.bbc.co.uk/1/hi/sci/tech/4714135.stm, Accessed at Feb 12 2009
75. Lovgren S (2007) Robot codes of ethics to prevent android abuse, protect humans. In: National geographic news. Available via www, http://news.nationalgeographic.com/news/2007/03/070316-robot-ethics.html, Accessed at Feb 12 2009

76. Kerr I (2007) Minding for machines. In: Ottawa citizens news. Available via www, http://www.canada.com/ottawacitizen/news/opinion/story.html?id=e58202bb-f737-4ba7-a0ad-79e8071a1534, Accessed at Feb 12 2009
77. Warren MA (1997) On the moral and legal status of abortion. In: LaFollette H (ed) Ethics in practice. Blackwell, Oxford
78. Kant I (1780) Our duties to animals. In: Infield L (trans.). Lectures on ethics. Harper and Row, New York, pp 239–241
79. Chomsky N (1957) Syntactic structures. Walter de Gruyter, New York
80. Asimov I (1957) The naked Sun. Doubleday, New York
81. Hirose S (1989) A robot dialog. J Robotics Soc Jpn 7(4):121–126 (in Japanese)
82. Lessig L (1998) The laws of cyberspace - draft 3. In: Proceedings of the Taiwan Net'98 conference, Taipei, Taiwan
83. Arkin RC (1998) Behavior-based robotics. MIT, Cambridge, pp 31–62
84. Spector L, Klein J, Perry C, Feinstein M (2003) Emergence of collective behavior in evolving populations of flying agents. Proceedings of the genetic and evolutionary computation
85. Gabbai JME (2005) Complexity and the aerospace industry: Understanding emergence by relating structure to performance using multi-agent systems. Doctoral thesis, University of Manchester
86. Reynolds CW (1987) Herds and schools: A distributed behavioral model. Comput Graph 21(4):25–34
87. Bekey GA, Tomovic R (1986) Robot control by reflex actions. In: Proceedings of the 1986 IEEE international conference on robotics and automation, pp 240–247
88. Zhang X, Zheng H, Duan G, Zhao L (2002) Bio-reflex-based robot adaptive motion controlling theory. In: Proceedings of the 4th world congress on intelligent control and automation, pp 2496–2499
89. Newman WS (1989) Automatic obstacle avoidance at high speeds via reflex control. In: Proceedings of the 1989 IEEE international conference on robotics and automation, pp 1104–1109
90. Wikman TS, Branicky MS, Newman WS (1993) Reflexive collision avoidance: A generalized approach. In: Proceedings of the 1993 IEEE international conference on robotics and automation, pp 31–36
91. Arkin RC (2008) Governing lethal behavior: embedding ethics in a hybrid deliberative/reactive robot architecture. In: Proceedings of the 3rd ACM/IEEE international conference on human robot interaction, HRI '08
92. Ge SS, Yang Y, Lee TH (2008) Hand gesture recognition and tracking based on distributed locally linear embedding. Image Vis Comput 26:1607–1620
93. Yang Y, Ge SS, Lee TH, Wang C (2008) Facial expression recognition and tracking for intelligent human-robot interaction. J Intel Serv Robotics 1(2):143–157
94. Brooks AG, Arkin RC (2007) Behavioral overlays for non-verbal communication expression on a humanoid robot. Auton Robots 22(1):55–74
95. Lessig L (1999) Code and other laws of cyberspace. Basic Books, New York