Summer June 4, 2007

The Legal Crisis of Next Generation Robots: On Safety Intelligence

Yueh-Hsuan Weng, *National Chiao Tung University*
Chien-Hsun Chen, *National Chiao Tung University*
Chuen-Tsai Sun, *National Chiao Tung University*

The Legal Crisis of Next Generation Robots: On Safety Intelligence

Yueh-Hsuan Weng
National Chiao Tung University
College of Computer Science
1001 Ta Hsueh Rd.
Hsinchu 300, Taiwan
yhweng.cs94g@nctu.edu.tw

Chien-Hsun Chen
National Chiao Tung University
College of Computer Science
1001 Ta Hsueh Rd.
Hsinchu 300, Taiwan
Lucemia51@hotmail.com

Chuen-Tsai Sun
National Chiao Tung University
College of Computer Science
1001 Ta Hsueh Rd.
Hsinchu 300,Taiwan
ctsun@cs.nctu.edu.tw

ABSTRACT

Robot intelligence architecture has advanced from action intelligence to autonomous intelligence, whereby robots can adapt to complex environments and interact with humans. This technology, considered central to *next generation robots* (NGRs), will become increasingly visible in many human service scenarios in the next two decades. Accordingly, there is an emerging need to predict and address intertwined technological and legal issues that will arise once NGRs become more commonplace. Safety issues will be of particular interest from a legal viewpoint. As robots become more capable of autonomous behavior, regulations associated with industrial robots will no longer be effective. In this paper we will discuss issues associated with autonomous robot behavior regulations associated with the concept of *safety intelligence* (SI). We believe the SI concept (one of several *robot sociability problems*) is crucial to the development of "robot law" that will accompany the establishment of a society in which humans and robots co-exist.

Keywords

Safety Intelligence, Safety Engineering, Robot Intelligence, Human-Robot Co-Existence Society, Robot Law.

1. INTRODUCTION

The novelist and playwright Karel Capek created the word *robot* from the Czech term *robota*, meaning "forced labor" or "dreary work." He first used the word in his 1920 drama entitled *R.U.R.*—an acronym for *Rossum's Universal Robots*, the name of a fictitious factory that produced machines with human-like shapes but without self-awareness or the ability to think.[1]

Five decades passed before robots started to appear in industrial settings. The machines used *action intelligence*, an artificial intelligence concept whereby industrial robots perform simple, repetitive, and labor-intensive tasks. These robots lacked the ability to adapt to changes in their environments or to interact with people.

At the end of the twentieth century, the letters AI were more commonly used to represent *artificial* or *autonomous intelligence*, which enables robots to perform tasks in unstructured environments. Breazeal [2] classifies autonomous robot behaviors and social capabilities along a low-to-high sociability scale: as tools, cyborg extensions, avatars, and sociable partners. The final category is the focus of this paper, although we will use the term *next generation robot*, or NGR.[3] According to the Japanese Robot Policy Committee (RPC, established by the country's Ministry of Economy, Trade and Industry, or METI), NGRs consist of 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—in other words, functions through which they co-exist with humans in businesses and homes.[4]

The simple intelligence architecture of today's industrial robots consists of one control and one planning unit for locomotion and task arrangement, respectively. In comparison, the autonomous intelligence architecture of NGRs consists of five sub-intelligence units that control reactive behavior, deliberative behavior, adaptive behavior, cooperative behavior, and mutual understanding.[5] The autonomous intelligence function is meant to make intelligent interactions possible between robots and people and to allow robots to move about in complex environments (i.e., human society).[6]

The change from action to autonomous intelligence has raised many social and legal questions. Such leading robot research institutes as the Massachusetts Institute of Technology's Media Lab Robotic Life Group[1] and Waseda University's Wabot-House Lab[2] have been addressing these issues since the 1990s. In 2004,

[1] http://robotic.media.mit.edu.

[2] http://www.wabot-house.waseda.ac.jp/index.html.

ICAIL '07, June 4–8, 2007, Palo Alto, CA, USA.

Japan's METI sponsored a seminar to discuss difficulties that engineers and legal experts will face over the next two decades as they construct what the meeting organizers call a *human-robot co-existence society*.[3]

2. ROBOT SAFETY ISSUES

In its final seminar report, the METI Next Generation Robot Vision group noted potential legal issues to consider when making NGR policy, especially in the area of safety.[7] Robots are strong and require large amounts of electricity to perform their tasks, both of which are potential hazards when interacting with humans. The seminar participants agreed that new legislative and regulatory actions are required in order to mitigate the risk of human injuries. Industrial robots are already heavily regulated in Japan, but they cannot be used with NGRs that utilize autonomous behaviors to perform their tasks in unstructured environments.

To address these concerns for the emerging NGR industry, in 2005 the Japanese METI created the above-mentioned RPC and invited robotics experts to serve on it.[4] That committee's initial report emphasized the idea that Japanese government agencies and enterprises need to cooperatively address three areas of concern when establishing a NGR industry:

1. Develop a NGR market environment. According to a survey conducted by the Japanese Robot Association, the NGR market is expected to expand from 3 trillion yen in 2010 to 8 trillion yen in 2025.[8] Whereas past technical research directions were decided by university labs and research institutions, the committee suggested that future research directions be determined by market forces. Furthermore, local governments and robot enterprises need to cooperate to establish a research area specifically dedicated to robot development research.[5] Another focus in this area of concern will be personnel training in robot technology (RT)-related fields.
2. Ensure NGR safety. The clarification of legislative issues pertaining to NGR safety requires analyses of pre- and post-human-robot interaction responsibilities. The former (hereafter referred to as pre-safety regulations) includes standards for robot design and production. Post-safety regulations address situations in which human injury is caused by robot actions; they prescribe product liability protection systems and compensation from insurance companies.
3. Develop a mission-oriented RT system. Whereas Japanese are accustomed to making products and manufacturing systems according to available technologies, a mission-oriented RT system emphasizes technology development by enterprises based on demands and needs identified by government authorities.[9]

[3] "Toward 2025 and the Human-Robot Co-Existence Society: The Next Generation Robot Vision Seminar".

[4] 平成１８年５月　ロボット政策研究会　報告書.

[5] ロボット開発実証実験特区.

2.1 Pre-Safety Regulations

Achieving the goals outlined in the Robot Policy Committee report requires cooperation between business firms and research institutions in terms of technological development. Regarding legal guidelines, it will be the government's responsibility to make laws establishing product safety standards to be built into NGRs. Current regulations for industrial robots are created and enforced by the Japanese Industrial Standards group, the American National Standards Institute, and Switzerland's International Standards Organization (ISO). NGR regulations may end up looking very similar to these. Using ISO regulation 12100 as an example,[10] safety standards for industrial robots currently consist of four parts:

- *risk assessment*, for identifying potential dangers and estimating risk.
- *inherently safe design*, to reduce risk.
- *safety protection equipment*, for risks that can't be fully addressed using inherently safe design features.
- *user information*, for example, publishing user manuals and printing warnings and instructions on stickers to be placed on robots.

According to the Robot Policy Committee report, there are no official or standardized safety evaluation methods or regulations for NGRs. Until they are established, manufacturers and users must rely on standards for other industrial products when conducting NGR experiments.[6]

2.2 "Third Existence"

Currently, objects regulated by legal systems are referred to as either "first existence" (living/biological) or "second existence" (non-living/non-biological). Hashimoto argues that there are two reasons why a "third existence" category should be created for NGRs.[11] First, the autonomous intelligence nature of the NGR architecture means that they will not be capable of generating the self-awareness associated with science fiction robots. However, since NGRs will be capable of a limited degree of autonomous behavior, they cannot be placed in the second existence category. The third existence Hashimoto refers to is somewhere in between.

We will use the ants of Herbert A. Simon as another example of the uniqueness of the NGR autonomous intelligence architecture.[12] In describing the way that ants walk across the surface of a sandy beach, Simon notes that the ants' paths between two points are both irregular and complex, yet as simple independent behaving systems they are capable of navigating those paths. According to Simon, the complexity factor is not found in the ants but in the beach, therefore the ants' complex behavior is best viewed as a reflection of the complex environment in which they exist. In the case of NGRs, their autonomous behaviors will reflect the complexity of the human-robot co-existence society.

[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.

2.3 Open Texture Risk

A characteristic of general industrial product mechanisms is predictability. Machines that are built according to specific standards can only perform tasks that can be reduced to their corresponding mechanisms—in other words, machines cannot alter their mechanisms to match changing environments. The purpose of performing a risk assessment is to measure machine risk in order to design mechanisms for achieving approved safety levels.

Unfortunately, this regulation model does not fit well with the safety requirements of NGRs according to the legal concepts of "core" meaning and "open texture." Regarding language, any term in a natural language has a central (core) meaning, but the open texture character of language [13] 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 NGR 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*. NGR safety problems can be divided into risks from machine standards and risks from autonomous behavior. Machine standard risks can be regulated via a process of assessment and design, but the autonomous behavior of NGRs makes their risks complex, changeable, and unpredictable, thus requiring a different approach to risk assessment.

3. SAFETY INTELLIGENCE

A clear security issue for NGRs is how to limit robot "self-control." The current security policy contains no information on appropriate methods to analyze or test this feature, thus creating a need for new methods.

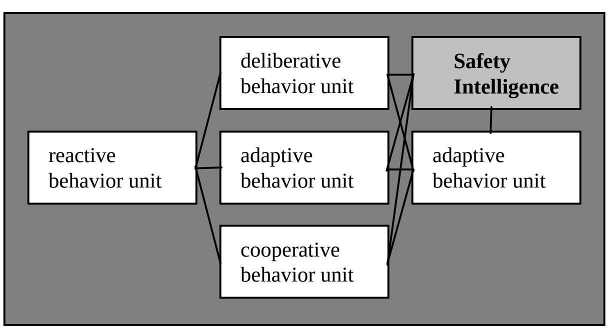

Figure 1. Safety intelligence architecture

We suggest that the *safety intelligence (SI)* concept—that is, a system of regulations restricting artificial intelligence—is a solution for this security issue (Fig. 1). This method for limiting robot behavior differs from the approach of safety design via risk assessment, which limits robot standards during the production phase.

3.1 Three Laws of Robotics and SI

In a human-robot co-existence society, NGRs are situated in complex and highly interactive environments that demand response functions to guarantee human safety. According to the Three Laws of Robotics[7] that Isaac Asimov described in his novel *I, Robot*, [14] a robot must guarantee the safety of the people it is interacting with before performing tasks, fulfilling its programmed roles, or obeying human commands. Ensuring its own functionality is of secondary importance. Although first described in a work of fiction, these laws have received considerable attention in AI and robotics research. As a result, the concept of designing robotic behaviors in a manner that ensures human security is generally agreed upon by all parties involved in NGR design and production. Furthermore, as robots gradually take on labor-intensive and repetitious jobs outside of factories and workplaces, The Three Laws of Robotics will serve as a "mechanism of human superiority."[15] Still, there will be an important distinction when it comes to addressing the legal ramifications of Asimov's three laws. Since the purpose of robot functionality will be to satisfy various human needs, they will be built in a manner so as to protect themselves *as the property of humans*, whereas biological organisms protect themselves *for their own existence*. Robots will be "born" for human purposes, therefore they will need to act independently and in accordance with the human ethics that underlie the Three Laws of Robotics.

Of course, a major challenge for NGR designers will be deciding how to apply the Three Laws of Robotics to the realities of a human-robot co-existence society. In his novel, Asimov addressed contradictions between the Three Laws and the concept of textualism—a formalist theory of statutory interpretation.[16] For example, according to the first law, in order for Asimov's robots to serve as a police force, they had to distinguish between blood resulting from a helpful surgical operation and blood resulting from acts of violence. Such decisions in real life require human-based intelligence (HBI)—a mix of common sense, understanding current situations, and making appropriate comprehension judgments.

A minority group of robotists argue that Asimov's Three Laws are not the only possible foundation for implementing the Safety Intelligence concept. In noting that circumstances such as the surgery/attack example above require "morality engineering," Shigeo Hirose argues that if doctrinal reasoning conflicts with morality in robotics, the Three Laws may become contradictory and unnecessary.[17] Accordingly, in extreme circumstances robots might be allowed to commit homicide based on the wishes of a human majority. This extreme example touches on the fears that many people have when they consider autonomous robots—

---

[7] First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.

Second Law: A robot must obey orders given it by human beings except where such orders would conflict with the First Law.

Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

that is, they are troubled by the idea of letting robots obey rules that are impossible to express legislatively, as well as letting them defend laws established by imperfect humans.

The majority of robotists consider the Three Laws to be a reasonable norm for representing SI. They generally believe that SI implementation difficulties are more technical—that is, they don't know how to build robots with sufficient intelligence (or "consciousness") to obey the Three Laws.[18]

3.2 Robot Sociability Problems

As robots are increasingly integrated into human society, associated problems will resemble or merge with those in other fields. Examples include safety intelligence (a merging of engineering and legal issues concerning NGR autonomous behavior and resulting risks); third existence theory (philosophical and legal issues tied to defining NGR legal status); and robot-human environments (architectural, urban planning, and engineering issues regarding human-robot co-existence). These interdisciplinary issues—all associated with the movement of NGRs from laboratories to the human world—can be referred to as *robot sociability problems* in the same manner that safety intelligence is often referred to as a *robot technical problem*.

3.3 SI is not a panacea

Implementing an SI function requires an understanding of its objectives. It is unrealistic to assume that SI will prevent all risks; instead, it is better to view SI as one part of a legal regulatory system that also includes a technological aspect. In this manner, the SI objective will vary according to definition and purpose.

	Safety Design by Risk Assessment	**Safety Intelligence**
Risk	Machine risk	Autonomous behavior (open texture) risk
Limit	Machine standards	Robot's intelligence architecture
Effect	Risk reduction	Prevent some dangerous behaviors

Table 1. A Comparison of Safety Regulation Methods

Since SI only affects autonomous behavior risk and not machine-related risk, reliable standards are still needed for inherently safe machine design. Some machine-caused accidents result from "machine fatigue"—a factor that SI cannot prevent. The pre-safety risk assessment process can reduce the odds of a machine fatigue accident occurring, but cannot promise complete protection. In other words, a completely safe NGR cannot be promised even when pre-safety issues are properly identified and addressed. Insurance and product liability systems are therefore needed in response to post-safety issues. We are currently working on a NGR safety regulation model that emphasizes the role of SI during the pre-safety stage.

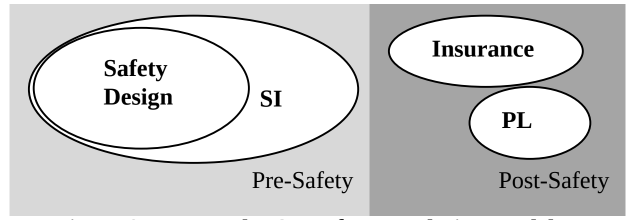

Figure 2. Proposed NGR safety regulation model

3.4 Do NGRs need doctrinal reasoning?

In Asimov's novel, multiple contradictions arise between \the Three Laws and doctrinal reasoning. Natural language is the medium used by human beings to access legal content. Even though human legal language always presents problems regarding vagueness and abstraction, laws and rules can be obeyed by doctrinal reasoning. The same process can be used to accomplish SI. This raises an important question: do NGRs need doctrinal reasoning to accomplish SI, and is it possible for them to acquire it?

The field of human-robot interaction (HRI) concerns communication and collaboration between people and robots. HRI researchers view human-robot communication as consisting of verbal communication (using natural language) and non-verbal communication. "Non-verbal information" consists of actions, gestures, non-linguistic sounds, and environment. Humans communicate using a mix of verbal and non-verbal information, with non-verbal information representing up to 60-70% of total communication[19]. Although verbal and non-verbal communication have potential for robot-human interactions, we believe that non-verbal communication is a more reliable means for implementing the SI function and avoiding contradictions arising from doctrinal reasoning.

3.5 Who decides the meaning of safety?

The term "safety" is unclear and open to contextual interpretation. We will someday need to address the question of whether NGRs should be allowed to make their own interpretations of the world. *Biomorphism* [21] refers to artificialities created by human designers. In robot technology, biomorphism is usually used to describe humanoid or semi-humanoid (i.e., cyborg) representations. The notion of a "biomorphic robot" is associated with different human-centered standpoints and ethics; it has no meaning to an NGR.[8]

But allowing an NGR to interpret the idea of safety is not a problem of can or can't, but one of should or shouldn't. It not only needs a human-centered viewpoint but also evaluations of the effects of safety. Legal scholars are more suited to this job than NGRs. Thus, definitions of ambiguous natural language-based terms such as *safety*, *injury*, and *protect* need to be pre-defined by lawmakers and transferred to a nonverbal–based legal

[8] For example, look at US animal protection laws. In creating a stable relationship between humans and dogs, dogs have certain rights, but those rights have no meaning to the dogs themselves.

machine language (i.e., automata) in order to give SI access to legal content.

A possible solution to this dilemma is to let legal scholars create definitions of safety responsibilities for robot companies to follow when designing the safety functions of their robots. However, the number of NGRs in a human-robot co-existence society will someday equal the number of cars or PCs in use today, making NGR regulation a considerable challenge. If different robot manufacturers use different intelligence architectures, solving this dilemma will be even more difficult, therefore robotics manufacturers need two standard platforms, one for programming and one for safety. A safety platform in the form of a SI unit could interact with legal machine language to reduce risks associated with a machine's autonomous behaviors. For this reason, we describe SI as an independent unit in the robot intelligence architecture.

4. CONCLUSION

The Nihon Keizai Shimbun is one of many periodicals predicting that the twenty-first century will witness a "robot revolution," with robot products being found in people's homes to the degree that they may affect the personal and spiritual lives of many.[22] If a human-robot co-existence society does emerge, it must be accompanied by well-reasoned laws that regulate robot-human relationships. The importance of the NGR safety issue is emerging during a period in which autonomous intelligence robots are in the design and testing stage.[23] Now is the time to address the safety and legal issues, otherwise the human-robot co-existence society will be delayed by unforeseeable setbacks.

5. ACKNOWLEDGMENTS

The authors wish to thank the following for their helpful suggestions: Mr. Ji-Long Hsieh, Mr. Jon Lindemann, Prof. Ken Yabuno, Ms. Nozomi Hoshino, Ms. Akane Hara, Ms. Judy Hsieh, Ms. Jenny Chen, and three anonymous reviewers.

6. REFERENCES

[1] N. Tajika. The future astro boy. ASCOM, 2001 (in Japanese).

[2] C. Breazeal (2004). Social interactions in HRI: The robot view. IEEE Transactions on Man, Cybernetics and Systems—Part C, 34(2):181, 2003.

[3] Y. Anzai. Next generation robots as real world-oriented media. Journal of Robotics Society of Japan, 16(1):15-18, 1998 (in Japanese).

[4] Japan Ministry of Economy, Trade and Industry Robot Policy Council. Robot policy middle report, May 2005 (in Japanese).

[5] M. Asada and Y. Kuniyoshi. Robot intelligence. Iwanami, 2006 (in Japanese).

[6] Robotics Society of Japan Research Committee on Human Friendly Robots. Technical targets of human friendly robots. Journal of Robotics Society of Japan, 16(3):288-294, 1998 (in Japanese).

[7] Japan Ministry of Economy, Trade and Industry Next Generation Robot Vision Seminar. Toward 2025 Human-Robot Co-Existence Society Report, Apr. 2004 (in Japanese).

[8] Robotics Industry Development Council. Founders' statement. http://www.f-robot.com/tokku/tokku.html.

[9] Japan Ministry of Economy, Trade and Industry Robot Policy Council. Robot Policy Council Report, May 2006 (in Japanese).

[10] Y. Ishida, A. Neudorfer, K. Futsuhara, and T. Fujita. Importance of inherent safety in human machine interface environment. Report of the Human Interface Society (Japan) Interface Symposium, 2002 (in Japanese).

[11] T. Ojima, S. Hashimoto, and K. Yabuno. The book of Wabot 2. Chuokoron-Shinsha, 2003 (in Japanese).

[12] H. A. Simon. The sciences of the artificial (3rd ed.). MIT Press, 1996.

[13] D. Lyons. Open texture and the possibility of legal interpretation. Law and Philosophy, 18(3):297-309, 1999.

[14] I. Asimov. I, Robot. Doubleday, 1950.

[15] J. A. Fodor. (1987). Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In Z. Pylyshyn (ed.), The robot's dilemma: The frame problem in artificial intelligence. Ablex Publishers, 1987.

[16] http://en.wikipedia.org/wiki/Textualist

[17] S. Hirose. A robot dialog. Journal of Japan Robotics Society, 7(4):121-126, 1989 (in Japanese).

[18] R.A. Brooks. Flesh and machines: How robots will change us. Eurasian Publishing, 2003.

[19] T. Kurogawa. Nonverbal Interface. Ohmsha, 1994 (in Japanese).

[20] S. Sugano and K. Shibuya. Anthropomorphic robots for nonverbal communication. Journal of the Robotics Society of Japan, 15(7): 975-978, 1997 (in Japanese).

[21] H. Aldsersey-Williams. Zoomorphic: New animal architecture. Laurence King Press, 2003.

[22] "Japanese robots will change the world" (unsigned editorial). Yazhou Zhoukan: 14-23, March 25, 2002 (in Chinese).

[23] S. Tachi. How to construct future society: Tele-remote controlled robots. Meme No. 4, Fuji Research Institute Corporation, Apr. 2003 (in Japanese).