Ethics Aspects of Embedded and Cyber-Physical Systems

Abhilash Thekkilakattil[1] and Gordana Dodig-Crnkovic[2]

[1]Mälardalen University, Västerås, Sweden
[2]Chalmers Technical University and University of Gothenburg, Gothenburg, Sweden

***Abstract*—The growing complexity of software employed in the cyber-physical domain is calling for a thorough study of both its functional and extra-functional properties. Ethical aspects are among important extra-functional properties, that cover the whole life cycle with different stages from design, development, deployment/production to use of cyber physical systems. One of the ethical challenges involved is the question of identifying the responsibilities of each stakeholder associated with the development and use of a cyber-physical system. This challenge is made even more pressing by the introduction of autonomous increasingly intelligent systems that can perform functionalities without human intervention, because of the lack of experience, best practices and policies for such technology.**

In this article, we provide a framework for responsibility attribution based on the amount of autonomy and automation involved in AI based cyber-physical systems. Our approach enables traceability of anomalous behaviors back to the responsible agents, be they human or software, allowing us to identify and separate the "responsibility" of the decision-making software from human responsibility. This provides us with a framework to accommodate the ethical "responsibility" of the software for AI based cyber-physical systems that will be deployed in the future, underscoring the role of ethics as an important extra-functional property. Finally, this systematic approach makes apparent the need for rigorous communication protocols between different actors associated with the development and operation of cyber-physical systems that further identifies the ethical challenges involved in the form of group responsibilities.

I. Introduction

The increasing use of computer-based systems in day-to-day applications has revolutionized human life. Today, the influence of computing extends to cyber-physical systems (CPS) that interact with the physical world, and therefore must operate dependably, safely, securely, and efficiently [1]. Some of the advantages offered by computing in cyber-physical systems include environmental sustainability through efficiency, high performance, cost reduction, and flexibility. Cyber-physical systems are found in transportation, aerospace, robotic systems, manufacturing, process control, environmental control and smart cities. CPS are considered to be the basis of the next computing revolution and are expected to develop rapidly in coming years. Embedded systems can be considered as precursors of CPS. An example of such system is drive-by-wire system that replaces large amounts of hardware, such as wires and cables, reducing cost while enabling flexibility in the design of the system. In addition to replacing traditional hardware components, computer systems enable improved user experience through driver assistance functionalities in automobiles, as well as, in recent years, performing more advanced tasks of autonomous driving. In the future, it is expected that highly intelligent systems independent from human control will be developed, that will co-exist with humans and perform traditionally human tasks like decision-making and control. Case of special interest for our analysis are safety critical systems, for which a failure of the system may lead to catastrophic consequences such as loss of human lives. Both embedded and cyber-physical systems can be safety critical. Up to now most lessons-learned regarding safety-criticality are related to embedded systems, and we can draw conclusions about cyber-physical systems by extrapolation. Safety-critical systems have to be highly dependable. Particularly important among the attributes of dependability are reliability, availability and security. In order to guarantee dependability of cyber-physical systems, the personnel involved in their research, development and use needs to conform to the highest ethical standards. Bowen [2] examines the ethical aspects of safety-critical systems. The article identifies and presents a detailed analysis of the practices that should be avoided while building traditional safety-critical systems. Bowen notes that many decisions in building safety-critical systems directly depend on economic considerations, rather than on safety criteria. Leveson [3] examines the role of software in spacecraft accidents and notes that the flaws in safety culture present an important cause of majority of these accidents. Moreover, the author discusses some methods to prevent the flaws in future development efforts. Leveson [4] also relates the situation of lack of safe engineering practices in developing software to a similar situation involving early steam engines in which safety features were introduced after gaining sufficient experience with their use.

One of the challenges involved in making informed decisions while building cyber-physical systems (and embedded system as a special case) is to be able to trace decisions taken by various stakeholders into specific (anomalous) behaviors exhibited by the system. In other words, engineers building a safety-critical system must be able to predetermine the impact of their design decisions on the safety of the system. This enables the engineer (or any actor involved) to make better ethical judgments. In particular, it is necessary for engineers

This work was supported by the Swedish Research Council project CONTESSE (2010-4276)

to assess the impact of their design choices taking ethical aspects as a part of non-functional requirements. Similarly, other stakeholders such as users operating cyber-physical systems must be aware of the impact of their own decisions. For example, police forces deploying autonomous drones to destroy drug plantations must assess the safety risk to the environment surrounding the plantation, since the potential damage caused by the drone can be unpredictable. Finally, in the case of systems that employ Artificial In- telligence (AI) in their design, manufacturing or operation, it must be possible to separate inherent design errors from the decisions taken by the AI software. In order to be able to trace back unsafe behaviors of such system to its builders, users or to the system itself, a clear demarcation of the responsibility must be made, especially with the projected use of AI in cyber-physical systems. Such a classification can serve as a template for structuring the ethical responsibility of the different stakeholders associated with cyber-physical systems, as well as for accommodating the ethical responsibility of the software when highly intelligent systems are used in the future. The resistance to attribute functional responsibility to software can be detrimental [5] since the most likely the industry will be reluctant to rely on the advances of AI, fearing litigation resulting from potential accidents due to decisions taken by the AI software. In this paper, we classify cyber-physical systems based on the amount of autonomy of the system and automation involved. This allows us to isolate responsibilities of the different stakeholders (both humans and software) involved with the development and use of cyber-physical systems. Our classification framework enables us to consider the ethical responsibilities of software alongside humans, when truly intelligent software systems will be deployed that are independent of human control. It also highlights the critical importance of rigorous communication infrastructure between various stakeholders, requiring the study of the ethical aspects of such a communication.

II. Ethics

The field of ethics concerns the study of what constitutes an ideal conduct in various situations, while taking decisions that may affect other people or the environment. The main goal is to examine the moral aspects of conduct, and to determine actions that are deemed 'morally acceptable'. The question of what is morally acceptable needs a continuous evaluation, and so is the study of ethical aspects of major changes introduced into the society. In particular, it is important to understand and rigorously study the ethical aspects of new technologies, and their implications on the society. Cyber-physical systems are such a radically new and complex emerging technology that deserves careful ethical analysis. In particular, for those working in the areas of science and technology, it is important to know ethical implications of the knowledge and information/data they produce and products that they design and build. It is therefore necessary that people working in technology develop ethical autonomy to be aware about the ethical aspects of the decisions that they make. Professional ethics is one of important constitutive elements of professionalism that aims at providing necessary basis for dealing with ethical challenges.

One of the first steps towards addressing ethical aspects is to identify a moral problem, and then examine the various alternative actions at hand after collecting relevant facts. Then specific actions are then performed based on the decision taken. The effect of the decision are then evaluated and fed back into the process of ethical assessment. In other words, the process of addressing ethical issues need to be a continuous activity, typically based on professional group's codes of ethics. Moor [6] notes that the contemporary rapid development of computing technology has resulted in policy vacuums (that is lack of policies), due to lack of experience with radically new artifacts and situations related to them so we need to learn how to deal with the ethical issues that have emerged. Moor argues that the analysis methods used to deal with the new ethical challenges arising out of the use of technology must unify the classical approaches towards solving ethical problems, e.g., the deontological and utilitarian approaches. He further points out that any ethical policy of action must be based on justice (just consequentialism) — it is unjust to act in a certain way that one would not accept from others. Therefore, policies and actions must be based on impartiality, even if it may bring harm to some, and must be such that everyone could be allowed to follow it. Moor also points out that policies and actions for computing technology should change as the technology advances.

Critics of computer ethics sometimes argue that it deals with problems that are unsolved or perhaps even unsolvable. Nevertheless, even though technology is constantly developing and one cannot expect a definite set of rules to guide ethical behavior, we can develop strategies and policies as a basis of an adaptive, learning, intelligent framework for ethical assessment. Within the domain of professional ethics of computing Dodig-Crnkovic [7] identifies the need for ethical education in computing curricula, in order to assure good ethical judgment in the computing profession. Georgiadou and Oriogun [8] present case studies that underline the need for teaching professional ethics as part of software engineering courses. Professional ethics aspects are part of yet a bigger picture which includes methods and tools of particular ethical analysis. While some ethicists claim that classical ethical approaches (consequentialism, virtue ethics, deontological ethics, justice ethics, etc.) can be applied to deal with contemporary fields like computing technology [9], others argue that radically new ways of ethical analysis are needed since the problems presented by computers are unique [6]. Floridi [10] supports uniqueness of computer ethics and shows that Information and Communication Technologies (ICT) bring about fundamentally new dimensions to old ethical problems and force us to think about the very foundations of our ethical policies.

III. Terminology

In order to address the ethics aspects of embedded and cyber-physical systems, we first define some basic terminology

that we will use consistently in the rest of the paper. First, for clarity, we present the formal definition of a stakeholder.

Definition 1. *A stakeholder associated with a cyber-physical system is defined as an agent who may interact with the cyber-physical system at various stages of its design, development, deployment and operation.*

A stakeholder can be a developer, a user or a software-agent that interacts with the cyber-physical system at some level.

Definition 2. *A developer is defined as an individual or organization associated with the development of a cyber-physical system, whose decisions affect the design of the system.*

As mentioned before, Leveson observed that most software related accidents are due to a lack of safety culture in the associated software development [3]. Therefore, according to our definition, a designer/developer is responsible for safety related accidents that are manifested as a result of ignorance of safety in the design and development. For example, engineers are responsible for evaluating the safety of their design, project managers are responsible for making sure that sufficient consideration is given to safety, company executives must ensure that sufficient time and money is spend in establishing a safety culture in the company.

Definition 3. *A user is defined as any stakeholder (person or software) that uses or operates a cyber-physical system, whose decisions affect how the cyber-physical system is used.*

For example, a pilot of an aircraft is referred to as a user. The user is responsible for following the safety practices during the use of a cyber-physical system. The safety practices are typically established by the developer's own best practice, guided by safety department in the software development organization and by a government regulatory body (such as the Federal Aviation Authority in the USA that decides the safety practices for flying aircrafts in the American airspace).

Definition 4. *A software agent is defined as any software associated with the development and use of a cyber-physical system, capable of making decisions that affect the system itself and its behavior.*

In the future, systems with increasingly advanced AI will be used in the development and use of cyber-physical systems. Any software, that is highly intelligent and involved in the decision making process is referred to as a software agent. Note that current software development tools do not fall into this category. We assume that they are highly intelligent and capable of taking autonomous decisions independent of human control.

IV. Classification of Safety Critical Systems

In order to facilitate traceability of decisions, we classify cyber-physical systems based on the amount of autonomy involved and based on whether or not a human is involved in the decision making. We give simple examples on what we mean by each, and prepare foundations on which the rest of the paper is based.

A. Automatic Systems

Automation has been one of the key driving forces behind the widespread adoption of computing in safety and mission critical systems. Although automation has its modest beginning in data processing, it was soon applied in the context of process automation and control in industrial, vehicular, avionics and aerospace systems. These types of systems replace a specific hardware, performing a particular task, and do not implement any amount of autonomy apart from *imitating* the specific hardware that it replaces. We refer to such systems as *automatic systems*. An automatic system can be a subsystem of a larger system. In automatic systems, there is no decision-making involved. Once the system is built, it is expected to perform the specified task repeatedly and no decisions, other than that specified at design time, are made by the system itself. Noorman and Johnson [11] refer to automatic systems as 'autonomous systems'. We however want to differentiate systems that are truly AI-based from systems that merely 'automate' the required process.

An example of an automatic system is the adaptive cruise controller that automatically adjusts the vehicle speed in order to maintain a safe distance from the vehicle ahead.

B. Semi-automatic Systems

Although computing has extensively been used in industrial automation and control, many systems need human feedback thus requiring a human in the loop. These types of systems can be typically seen as a set of *automatic* subsystems coordinated by a human being. The key decision maker is the human, as (s)he specifies what task the system should perform, as well as how the task is to be performed.

Modern cars are an example of such systems. In modern cars many automatic subsystems such as the CAN, adaptive cruise controller, automatic gear shifter and anti-lock brakes are coordinated by the driver. Here, the driver is the decision maker; while the subsystems merely perform the preprogrammed tasks. According to the definitions presented in Section III, the driver is a user.

C. Semi-autonomous Systems

The increasing automation enables various *automatic* subsystems to be controlled and coordinated by computer software. Semi-autonomous systems are systems that are capable of autonomously performing tasks specified by humans. This kind of systems have *limited autonomy* in the sense that the system needs to be instructed with the specific task, and the task is carried out by the system based on its own decisions.

For example, consider an unmanned aerial vehicle tasked with taking pictures of a territory. These vehicles are usually provided with the mission, e.g., the coordinates of the target territory. The vehicle then autonomously computes the route and navigates to the target, takes pictures and returns without human intervention.

D. Autonomous Systems

Advances in artificial intelligence are expected to enable building of fully autonomous systems, such as humanoids, that reason and act like humans. They are expected to be highly intelligent learning systems that will be able to take autonomous decisions without direct human intervention, and will eventually be outside human control. Another example is autonomous cars that are currently being developed. We refer to such systems, that are capable of deciding what task the system should perform and how the task is to be performed by themselves, as autonomous systems. In this kind of systems, the software can be referred to as a software-agent according to the definitions presented in Section III.

V. Assigning Responsibility and Accountability

In this section, based on the classification in the previous one, we assign the responsibility of failure to the different stakeholders in the life-cycle of a cyber-physical system.

A. Automatic System Failures

Automatic systems are designed to perform a specific task, free from human control. Failures of automatic systems can be attributed to the designers/developers of the system, since they specify what task the system will perform and how the system will accomplish that. Therefore, the designers/developers involved must conform to the highest ethical standards. They must adopt safety culture as a part of their design/development process and must consult with a safety engineering team throughout the development and plan for the use of a cyber-physical system, even under budget and deadline constraints.

For example, if an engineer feels that a sub-system has not been thoroughly tested or a proper safety study has not been undertaken, (s)he should discuss with the project manager. Similarly, a project manager must listen to his/her engineering team since they are in a better position to identify flaws in the system.

B. Semi-automatic System Failures

Semi-automatic systems are typically composed of several automatic subsystems and one or more humans in the loop (users). Failures in semi-automatic systems can be attributed to either the designer/developer (that includes training and documentation crew, as well as testers), product maintenance function or the user. Production is the next link in the chain of responsibilities. The continuous maintenance of the product might be the responsibility of either producer or the user. Additionally, the users must adequately use the system, e.g., follow the safety procedures described by the producer. This is quite important because, for example, incorrect use of cyber-physical systems can also result in undesirable consequences. The Bhopal gas tragedy is a classic example of how ignorance of safety procedures led to a major catastrophe [12]. Given that large parts of process industry are controlled and operated as cyber-physical systems this should be kept in mind.

In order to understand the attribution of responsibility, let us take as a next example user responsibility of a pilot flying a plane. The pilot must follow the safety instructions and protocols laid down by the developers and producers of the system, incorporated in safety practices defined and monitored by a safety department of the company operating the aircraft. In that case, the failure of sub-systems, such as fuel meters, in spite of the human-in-the-loop following all safety instructions, are attributable to the previous links in the responsibility chain - either maintenance or producers and designers/developers.

C. Semi-autonomous System Failures

The advances in the field of artificial intelligence have enabled building of systems that are capable of performing intelligent decision-making. Many autonomous drones exist today that are able to perform tasks without human intervention. However, their intelligence is limited to specific tasks, such as surveillance, and cannot be used for other purposes.

In these types of systems a user decides what the system will perform. Therefore, the user is responsible for the anomalous behaviors caused by its deployment for non-specified tasks. Nonetheless, if the artificial intelligence in the system takes a decision that causes an anomalous behavior, e.g., an autonomous drone ”decides” to crash into a building, the accountability can be traced back to the AI software (i.e., the software-agent). When the system consists of automatic subsystems its failures are attributable to the designer/developer, if the software-agent, such as an assisting robot, is unable to detect failures in the sub-systems.

D. Autonomous System Failures

In fully autonomous systems, the anomalous behavior of the system is the primary ”responsibility” of the software-agent. In this case, there must exist strong evidence that the autonomous system is truly autonomous. However, as developer other than a software-agent is involved in the development of the system, the developer can be deemed responsible only for the failure of the automatic subsystems, when it is established that the failure did not depend on the decisions made by the autonomous system. For example, if the braking system of an autonomous car fails, the autonomous software will be unable to prevent a catastrophe. This holds particularly if the software-agent is unable to automatically detect failures in the subsystems. Therefore the developers involved must be cautious while designing the sub-systems and their mutual communications. Nevertheless, the software-agent has the overarching ”responsibility” in this case and should be given particular care. Our argument analysing the character of ”machine responsibility” or accountability of autonomous agents is based on [5][13] and [7]. It is argued in [5] and [13] that increasingly autonomous and intelligent agents must have built-in ethical properties, otherwise they can have severe unwanted consequences.

For example, in the future, if drone strikes are mandated by autonomous (highly intelligent) software based on some threat perception, the ”responsibility” (which in case of a machine amounts to accountability) of an incorrect decision should be with the software as it is the agent making decision. While

many may argue against this assignment of accountability ("machine responsibility") to a machine, we believe that a policy vacuum in this regard is even more dangerous if we allow autonomous agents be without any moral guidelines.

VI. The Ethical Challenge of Embedded and Cyber-Physical Systems

In this section, we structure the ethical challenges faced by designers/developers/producers, users and software actors. This enables us to take the very first steps towards formulating guidelines and policies in order to fill the "policy vacuum" as defined by Moor [6]. The idea of ethical "responsibility" ("accountability") of software agents is strongly criticized and refuted by many [11]. Critics argue that the designers and developers can still influence the decisions of the software, as they develop and deploy its first version, which then is allowed to "learn" from its interactions with the environment. Underlying is the idea that we can predict possible behaviors of an autonomous intelligent system. That idea, however is unjustified. For a future autonomous intelligent cyber-physical system we can be sure that we will not be able to predict its behavior — that lies in the nature of autonomy. We point out that the class of systems whose behavior can be predicted lie in the context of what we refer to as automatic systems. An example of such a system is an automatic missile defense system. In this paper, we clearly differentiate automatic systems from autonomous systems that are highly intelligent and outside of immediate human control. While developers can, to some extent, influence the behavior of software agents, the negligence of "responsibility"/accountability of software can leave a hazardous policy vacuum in the projected use of AI in cyber-physical systems. Therefore, in order to formulate effective policies regarding the ethical aspects of cyber-physical systems, the "responsibility" (that is accountability) of software agents cannot be ignored as they act in the domain where natural intelligent agents have moral responsibility and technological artifactual agents should have corresponding "artifactual moral responsibility" designed/built in as long as possible [5]. Deploying highly autonomous intelligent cyber-physical systems without any moral/ethical considerations by design can have unforeseeable negative consequences [5] [13].

For the sake of discussion, we structure the *ethical boundaries* of the different stakeholders based on the responsibility associated with the safe functioning of the cyber-physical system for each of them. Each stakeholder, be it a designer/developer/producer, user or a software agent is bounded by the responsibilities described in Section V that make them *committed* to all the other stakeholders (e.g., passengers in an aircraft or the environment). This is important since each stakeholder acts assuming that the other stakeholder have done their job properly e.g., a pilot flies an aircraft assuming the designer/developer/producer have designed, developed and manufactured properly; that the maintenance crew have done their job and that flight control functions flawlessly and so on. It is the moral duty of all stakeholders to discharge their duties with the highest integrity and rigor. Since in this framework it is in principle possible to map anomalous behavior to a particular stakeholder or a group of stakeholders, each stakeholder must stick to the highest ethical standards to prevent the anomalous behaviors within his/her responsibility.

A. Communicating Information Responsibly

When there are multiple stakeholders in the picture, there is a *collective responsibility* that requires the designers/developers/producers, users and software actors to interact responsibly, especially without ambiguity, to ensure safe operation of the cyber-physical system [13]. There should be well defined interfaces so that relevant information is made available completely, and without ambiguity, especially by manufacturers who may not want to disclose certain information for various reasons. Communication protocols must be developed for this purpose, and communication must be maintained at least until the operation of the cyber-physical system is stable e.g., the users are adequately trained. Additionally, there must be sufficient infrastructure in place to communicate back diagnostic information to the developers to improve safety.

This highlights another important area for which ethical policies need to be formulated— type and amount of information that different stakeholders must disclose. For example, manufacturers may not want to disclose certain details of the cyber-physical system e.g., so that they don't lose their competitive edge. Also, many users may not want to spend too much time and money in getting trained at using the system because of financial constraints. Similarly, there could be inconsistencies in the communication interfaces, especially between AI based software-agents and humans, due to which information may get lost, remain unattended or is misinterpreted. For example, in the case of classical safety-critical systems, Leveson cites poor information flow as one of the reasons for software related accidents [3]. Therefore, the need for communication among different stakeholders requires us to investigate the ethical aspects of communication— with emphasis on what information needs to be 'disclosed' for ethical reasons, while protecting business secrets. To summarize, the main challenge involved in such a communication infrastructure is that there should be an adequate transfer of relevant information pertaining to the safe operation of systems– especially between AI based software agents and humans.

B. Attributing Responsibility to Software

In this context, the responsibility of software-agents can become a contentious issue [11] [5] [13]. The main claim of our article is that new autonomous and intelligent cyber-physical systems must be accompanied by suitable (self) regulatory mechanism that will assure their ethical behavior. It is still a widely debated topic and some ethicists refuse any possibility of machine ethics/ethical machines [11] while some others defend the same line of argument that we present [14][15][16]. We however, reiterate our argument that ignoring

responsibility of software can only leave the process of addressing the policy vacuum incomplete. Software-agents that are highly intelligent must be programmed to "understand" and apply ethics in order to guarantee that they learn and evolve just like humans do. Such an open-minded approach towards the ethical responsibility of software-agents enables social sustainability, also in the context of legal procedures arising out of accidents. On the contrary, if the software-agents associated with cyber-physical systems are provided immunity from responsibility, accidents that occur due to wrong "decisions" taken by software-agents will remain a controversial issue. Also, other stakeholders who are in no way responsible for such accidents risk being penalized. The risk of litigation may also make the stakeholders, such as developers, reluctant to adopt the advanced capabilities provided by the field of artificial intelligence in building safer and more efficient systems (e.g., the developers may be discouraged from using AI in systems to achieve energy efficiency by improving operational efficiency which may require complete autonomy for the software-agents). This reluctance may prove to be detrimental since it may further undermine the goal of social sustainability. Aircraft accidents suspected to be caused by pilots actualize discussions if not completely automatic systems would be safer.

In this regard, it is important that attributing the responsibility of failures to fully autonomous software gain widespread acceptance among diverse communities such as governmental, legal, user-communities etc. This can be quite challenging because of two main concerns fear of the unknown and inertia to change. While the first of the concerns can be addressed to a great extent by proper research and education, the second problem is more difficult. For example, there are huge legal obstacles for autonomous cars in some countries. In order to adopt truly revolutionary technologies, the society must be prepared to take risks associated with attributing "responsibility" of failures to software while continuously learning from experience and improving technology.

VII. Conclusions

In this paper, we presented a framework for assessment and attribution of responsibility based on classification of cyber-physical systems with respect to the amount of autonomy and automation involved. Our framework allows the traceability of anomalous behaviors to the responsible agents, particularly AI based software-agents, making it possible to structure and demarcate ethical responsibilities. Besides specific stakeholders' responsibility we also highlight the importance of collective responsibility [14][15] that addresses the central problem of information communication between different stakeholders. It requires developers, users and software-actors to interact responsibly, using unambiguous communication protocols to prevent incomplete or incorrect information exchange. The central role of communication as a basis of distributed responsibility presents a topic for future research. We advocate a case-by-case study to delineate ethical aspects of such a communication protocol. Finally, we elaborate the risks associated with the resistance to attribute "responsibility" (machine responsibility or accountability) of failure to the AI software. The main obstacle is the reluctance to adopt advances in AI for emerging autonomous intelligent cyber-physical systems as moral agents.

We conclude by emphasizing that our framework is the first and novel attempt to base the reasoning about attribution of responsibility (including artifactual responsibility of software agents) on the systematization of types of cyber-physical systems with respect to the type of decision making into automatic, semi-automatic, semi-autonomous and autonomous. Future work is expected to develop a more detailed scheme which involves identifying the application-specific ethical responsibilities of each entity for some sample of typical cyber-physical applications.

Acknowledgments

We would like to thank an anonymous reviewer who suggested us connection to the topic of Intelligent regulatory compliance and I* modeling language which could be used to intelligently assure compliance of a complex system to safety requirements. This will add one more layer of (regulatory compliance) AI on the top of the decision-making AI as a part of cyber-physical system.

References

[1] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: The next computing revolution," in *The 47th ACM/IEEE Design Automation Conference*, June 2010.

[2] J. Bowen, "The ethics of safety-critical systems," *Communications of the ACM*, 2000.

[3] N. G. Leveson, "The role of software in spacecraft accidents," *AIAA Journal of Spacecraft and Rockets*, 2004.

[4] ——, "High-pressure steam engines and computer software," in *Proceedings of the 14th International Conference on Software Engineering*. ACM, 1992.

[5] G. Dodig-Crnkovic and D. Persson, "Sharing moral responsibility with robots: A pragmatic approach," *In Proceedings of the Tenth Scandinavian Conference on Artificial Intelligence (Vol. 173)*, 2008.

[6] J. Moor, "Just consequentialism and computing," *Ethics and Information Technology*, 1999.

[7] G. Dodig-Crnkovic, "Computing curricula: Social, ethical and professional issues," *In proceedings of the Conference for the Promotion of Research in IT in Sweden*, 2003.

[8] E. Georgiadou and P. Oriogun, "Professional issues in software engineering curricula: case studies on ethical decision making," in *Proceedings of the International Symposium on Technology and Society*, 2001.

[9] D. G. Johnson, "Ethics online," *Communnications of ACM*, 1997.

[10] L. Floridi, "Information ethics: On the philosophical foundation of computer ethics," *Ethics and Information Technology*, 1999.

[11] M. Noorman and D. Johnson, "Negotiating autonomy and responsibility in military robots," *Ethics and Information Technology*, 2014.

[12] N. Leveson, "A new accident model for engineering safer systems," *Safety Science*, 2004.

[13] G. Dodig-Crnkovic and B. Çürüklü, "Robots - ethical by design," *Springer Special issue on Requirements Engineering Ethics and Information Technology*, August 2011.

[14] A. Adam, "Delegating and distributing morality: Can we inscribe privacy protection in a machine?" *Ethics and Information Technology*, 2005.

[15] C. Allen, I. Smit, and W. Wallach, "Artificial morality: Top-down, bottom-up, and hybrid approaches," *Ethics and Information Technology*, 2005.

[16] M. Anderson and S. L. Anderson, "Artificial morality: Top-down, bottom-up, and hybrid approaches," *AI Magazine*, 2007.