Journal of Computer Science and Information Technology
December 2019, Vol. 7, No. 2, pp. 1-9
ISSN: 2334-2366 (Print), 2334-2374 (Online)
Copyright © The Author(s). All Rights Reserved.
Published by American Research Institute for Policy Development
DOI: 10.15640/jcsit.v7n2a1
URL: https://doi.org/10.15640/jcsit.v7n2a1

Challenges in Monitoring Machine Ethics

Christopher B. Davison[1], Silvia Sharna[2] & Genevieve Davison[3]

Abstract

Monitoring the moral decisions of autonomous, intelligent machines is a difficult endeavour. Issues such as resource allocation, security, location, and engineering need to be addressed. Identifying factors within current Von Neumann architecture that constitute a machine in "moral crisis" present considerable challenges. Additionally, there is the issue of transmitting this information to a monitoring entity and that entity being able to act upon the information. The authors of this paper present many engineering challenges and then a theoretical model to address these challenges.

Keywords: machine ethics, monitoring, moral machines.

1. Introduction

Ethical theory and moral habitbasically refer to human nature or behavior. More recently, scientists are concerned with the ethical decision-making ability of Artificial Intelligence (AI) as well as Autonomous Moral Agents (AMA). Recent advances in AI have greatly improved society and are changing modern life. However, these vast improvements bring concerns regarding ethical machine behavior. As such, the research questions addressed within this article are:

1. What machine factors might constitute a moral crisis or ethical dilemma?
2. What machine ethics monitoring model can be implemented?
3. Where would a machine ethical monitoring system be placed?

The beginning of this article provides definition and constructs important to understanding ethics and machine ethics. Following that, a review of the literature is provided. Within the literature, the theoretical framework for the above research questions is found. Then, a factor analysis of a *healthy*, non-conflicted machine and the concomitant methodology and analysis is presented. After that discussion, a multi-level machine ethics monitoring model is presented as well as a related discussion on where to implement this model.

2. Definitions and Constructs

Artificial Intelligence (AI): Intelligence of machines and robots and the branch of computer science that intends to generate machine intelligence is known as Artificial intelligence (AI). Artificial Moral Agent (AMA): The idea of artificial moral agent (AMA) is expansion of the concept of an ethical agent. Being a moral agent is a matter of degree in the sense that to what extent artificial agents are able to implement human-like faculties and action. Therefore, by artificial moral agent (AMA) we mean an artificial agent that is computationally based, guided by norms, and implemented through software (Nagenborg, 2007).

Machine Ethics: The ethical part of the artificial intelligence is known as Machine ethics (or machine morality, computational ethics or computational morality) which is concerned with the moral behavior of artificially intelligent beings. Machine ethics is concerned with the moral behavior of humans as they design, construct, use and treat such beings.

[1] Ball State University, Information Systems and Operations Management, cbdavison@bsu.edu
[2] Ball State University, Computer Science, ssharna@bsu.edu
[3] Robotics/AI Consultant, genrosied@gmail.com

In that sense, it is different from robo-ethics as well as computer ethics, which focuses on professional behavior towards computers and information. The implementation of moral decision making into computers, robots, and other autonomous devices are illustrated as Machine ethics (Allen, Wallach & Smit, 2006). Machine ethicists are concerned with the ethical reasoning of machines and how, if possible, to imbue these machines with this reasoning ability (Davison & Allen, 2017).

Autonomous System (AS): Systems such as robots or smart buildings that are capable of learning. These systems can analyze inputs (e.g., sensor readings), perform adjustments and learn from its environment. The system is capable of communications with other systems and, often, with humans. The system can interact with its environment. Usually, the system is goal-oriented in that its existence has a purpose, such as a smart building energy management system (EMS) which exists to deliver comfortable environments for building occupants while minimizing resource consumption.

3. Literature Review

In order to address the question of "Can computers can think?", an Imitation Game was developed between two subjects (Turing, 1950). This game is named after Alan Turing and known as Turing Test. In this test, written communication between two subjects are required where the subjects are not being able to see, hear, or otherwise sense each other. A human is the first subject who will attempt to figure out whether the second subject is a machine or another human being simply from written communique. According to Turing, the thinking ability of a machine is proved by the Turing Test if the human subject cannot tell or chooses incorrectly. However, just because a computer can respond meaningfully to a user question or statement, there has been some ambiguity as to whether the idea of thinking computers is reasonable or not. This research question remains: Does a computer really possess the ability to illustrate the contents behind the words, or is it simply throwing symbols randomly? Turing addressed these questions partially. But formally, 30 years later, Johan Searle presented the total concept in his paper.

To prove that the machine in the Turing Test is not capable of understanding any meaningful concept, a thought experiment was conducted which is known as The Chinese Room (Searle, 1980). A native English speaker was provided sets of syntactical rules without any knowledge of how to speak, write, or read Chinese. Although the translator's only known language was English, these rules allowed input in Chinese to result in coherent output also in Chinese. This is a topic of interest in various researches conducted in Machine Learning and Natural Language Processing.

Traditionally, it is not accepted by most of the moral philosophers that machines can be a subject of responsibility (Jordan, 1963; Lenk, 1994). "Moral responsibility is attributed to moral agents, and in Western philosophy that has exclusively meant human beings" (Johnson, 2001, p. 188). It has been explained many times by philosophers why they believe that moral responsibility should belong exclusively to humans. As Stahl (2000) explains, there exists an assumption that only humans are *beings*, as such, only humans can be held morally responsible. "This in turn can be justified by the fact that in order for responsibility ascriptions to make sense, the subject must fulfil a number of conditions such as cognitive and emotional abilities, some knowledge of the results of actions as well as possessing the power to change events" (Stahl, 2004, p. 68). Humans are social beings. They rely on morality and can recognize others as equal. The moral status of machines can make people escape from their responsibility (Grint & Woolgar, 1997). But there also exists some approaches for ascribing an independent moral status to computers. Such as, social interaction is of a moral nature and computers often play a vital role in social interaction.

Computers and robots need to have a significant autonomy for many technical purposes. Often it is impossible to control these machines in real-time. For instance, the Pathfinder robot of the Mars mission had to be highly autonomous. This is due to signal propagation delay and other technical and physical factors. Another example of an autonomous agent is a software bot which acts in its environment and can display moral characteristics (Mowbray, 2002). Now the question arises: Can autonomous machines react ethically and adequately to moral problems? As machines are granted more autonomy to react independently, a greater need for moral decision making is required. Further arguments regarding the autonomous moral status of computers and machines are that it might not be possible for humans (anymore) to live up to ethical expectations. Instead, only computers can achieve these ethical expectations (Bechtel, 1985). Practically, the argument leads to the direction that for human beings it has not been possible yet to reduce machine actions to human actions. Hence, as a pragmatic and manageable solution to this issue, it is important to assign moral responsibility to computers (Stahl, 2001).

Finally, some authors visualized a future with more advancement where computers will be utilized to fulfill significantly responsible task such as that of a judge (Stewart, Robber & Bosart, 1997). For such reasons, it is more common to find philosophical arguments of the following kind: "*AA*s are legitimate sources of im/moral actions, hence *A* should be extended so as to include AAs, that their ethical discourse should include the analysis of their morality..." (Floridi & Sanders, 2001, p. 3). Given the class *A* of moral agents that are defined as the class of all entities, which can finally be qualified as sources of moral action (Gunkel D. J., 2012).

The process of moral development is naturally explained and implemented in artificial agents by virtue ethics. Virtue ethics is also considered to be the most naturalized form of ethics.In 2011, Gips observed that when a virtue ethics model is adapted to AMA, then it is more promising and linked to connectionism: "both seem to emphasize the immediate, the perceptual, the non-symbolic. Both emphasize development by training, rather than by the teaching of abstract theory" (p. 244). This discussion is a revised version of the work by Gips (1995). Some philosophers such as Paul and Patricia Churchland (1990) suggested that a connectionist neural networks (CNN) model is suitable to approach moral cognition (Churchland, 1995). It is noticeably interesting that instead of consequentialism or deontology, some philosophers have adopted the virtue ethics model (Casebeer 2005).

Allen et al. (2005) suggested a hybrid approach to AMA grounded in virtue ethics. His model is based on the idea that virtue ethics may be more suitable to AMA (Allen et al., 2005; Allen & Wallach, 2009, Chapter 5). The significant difference that they pointed out is between the top-down approach in which they start from an ethical theory and look for the best implementation of it. Their model "takes a specified ethical theory and analyzes its computational requirements to guide the design of algorithms and subsystems capable of implementing that theory" (Allen et al., 2005, p. 80).

There are several competing ethical models in the research literature and discussed in this section. However, the concept of the best ethical approach for an AMA and a human agent may not be clear. Anassociation of deontology and consequentialism ethics is recommended (Allen et al., 2000; Bechtel, 1985). The developments outlined make this question more urgent today because computers have become ubiquitous in many areas. Some scholars have attempted to develop the discussion by introducing the concept of the computer as an AMA (Allen et al., 2000; Allen, 2002). "If computers can obtain the status of autonomous moral agents then they will be most likely to be successful in the light of cognitivist ethical theories" (Stahl, 2004).

AMA Monitoring and Control Models

Important to monitoring and controlling AI, is the selected and implemented monitoring model. One fundamental aspect of monitoring concerns who, or what entity/agency, will decide which AI oversight systems will be utilized to monitor the operational AI. Furthermore, the question exists as to what extent the monitoring system can intervene. Some oversight systems will be introduced by the programmers of the AI software involved at the behest of the owners and users of the technologies. Monitoring AI and ethical control is germane to autonomous vehicles. It is important for both the manufacturer and the user to ensure that their cars will never exceed the legal speed limit. This should be implemented with high priority in the context of machine learning, as the AI operating systems of those driverless cars will learn that the speed limits are being routinely violated by many traditional cars on the road. Additionally, consider a medical emergency sustained by one of the passengers. The AI must make an ethical decision regarding speed violations, human health, and safety precautions.

In the context of driverless cars, the Department of Transportation's recent attempt to develop safety regulations for driverless cars is referred to by Carnegie Mellon University artificial intelligence ethics experts (Danks & London, 2017)as an example of traditional guidelines that do not adequately test and monitor the novel capabilities of autonomous systems. Danks, the L.L. Thurstone Professor of Philosophy and Psychology and head of the Department of Philosophy says "We, as a society, need to find new ways to monitor and guide the development and implementation of these autonomous systems." (Rea, 2017, para. 6). A*pre-clinical trial* model is proposed by Danks and London (2017), so that the decision-making capability and future AI behavior in differing situations of the autonomous systems can be judged and verified in a wide range of contexts. If these trials result in success in targeted environments, then that would lead to a monitored, permit based testing and further easing of restrictions. This regulatory system should be modeled and managed similarly to the drug approval process regulated by the Food and Drug Administration (Danks and London, 2017). "Autonomous vehicles have the potential to save lives and increase economic productivity. But these benefits won't be realized unless the public has credible assurance that such systems are safe and reliable" stated London (Rea, 2017, para. 10). The model proposed by Danks and London (2017) provides a large measure of this assurance.

Current safety regulations regarding the driverless cars are not well planned for these systems and hence not well-equipped to ensure the safety, reliability, and performance of the autonomous system (Danks & London, 2017). These authors suggested an alternative by creating a staged, dynamic system that models the concept of the regulatory and approval process for drugs and medical devices, including a robust system for post-approval monitoring.

Using HERA (Hybrid Ethical Reasoning Agents), ethical principles are modeled as logical formulae whose truth determines which actions are permissible and which are not (Lindner, Bentzen & Nebel, 2017). In physical and virtual moral agents such as (social) robots and software bots, these theoretically well-founded and practically usable machine ethics are being implemented. The research approach is to use advances in formal logic and modelling as a link between artificial intelligence and recent work in analytical ethics and political philosophy (Bentzen, 2017). A famous formal logical model was authored in 1942 by Isaac Asimov as the Three Laws of Robotics. These rules for robots were designed to safeguard humanity from AI. These rules have undergone some revision including the addition of a Zero law. But initially, these are stated as:

1) A robot may not injure a human being or through inaction allow a human being to come to harm.
2) A robot must execute orders given it by humans except when such orders conflict with the first law.
3) As long as there is no contradiction with the first or second law, a robot must protect its own existence.

Asimov added a 4th law (zeroth law) to account for groups of humans, civilizations, and governance. This law is stated as:

0) A robot may not harm humanity, or, by inaction, allow humanity to come to harm.

Finally, other AI oversight systems most likely will be employed by courts and law enforcement authorities. For instance, to determine whether an accident was intentional and to whom or what the liability of the accident belongs. From this standpoint, Ethics Bots work as an exclusive AI Guardian. "They are to heed the values of the user, not the owner, programmer, or those promoted by the government" (Etzioni et al. 2016).

From the research literature, it is clear that monitoring machine ethics and monitoring AI is an important topic. However, there exists few models designed to accomplish that goal. From that, the authors of this research article postulate that it is necessary to understand machine factors that may indicate ethical conflict. As such, the purpose of this research article is to examine factors related to the monitoring ethical behavior of AMAs and propose a multi-level AMA monitoring model.

4. Methods

Factor Analysis Model for Monitoring an AMA.

Factor analysis is a process of taking a mass of data and shrinking it to a smaller data set that can be managed and understood more easily. It is a way of finding hidden patterns and showing how those patterns overlap and it will displaythe characteristics found in those multiple patterns.If a dataset contains similar items, then with the help of factor analysis, a set of variables (also known as dimensions) for those similar items can be created. For complicated datasets such as those involving psychological studies, socioeconomic status and other involved concepts, factor analysis can be a very convenient tool. A set of observed variables that have similar response patterns and associate with a confounding variable (which remains hidden and isn't measured directly) is defined as a specific factor. Factors are computed according to factor loadings, which is the extent of variation in the data they can explain.The goal is to identify factors according to their impact on the data (i.e., relative importance).

Typically, the application of factor analysis will accomplish two purposes; summarization and data reduction (Hair et al. 2010). Hence, this study generalizes the major factors of ethical orientation and tries to differentiate them into items of similar construct. The aim of performing factor analysis to any dataset is orderly simplification (Child, 1970) and it is particularly suitable for analyzing the patterns of complex, multidimensional relationships (Hair et al., 2010). Moreover, to examine the underlying patterns or relationships for a large number of variables, factor analysis can be applied to reduce the data for generating more informative constructs (Hair et al., 2010).

There is no ethical conflict present in the machine which is observed and sampled. A typical x64 machine running Windows 7 and performing typical organizational related functions was chosen to represent a *healthy*, non-ethically conflicted machine. This will provide sample data for later comparison to ethically conflicted machines. For the future research work the authors intend to introduce one or more ethical/moral conflicts and analyze the system (see discussion on future research below).

The purpose of this section is to empirically demonstrate the effects of the recommendations for AI oversight and identification of AMA moral conflict that has been made above. Factor analysis was the principle method of analysis used in this paper to identify the computational health. The factor analysis is performed through SPSS software.

Data Analysis.

For the data analysis it is given that there are no ethical/moral conflicts among the variables that are present in the sampled computer. The initial hypothesis (H_1) is: There is a significant correlation among the observed variables of CPU utilization, Disk IO, Network IO, and Memory utilization in a non-ethically conflicted computer system. Based on this hypothesis the authors performed a Factor Analysis. A total of 150 samples weretaken with no moral conflict injected into the system. In practice, the number of components extracted in a principal component analysis is always the same as the number of observed variables being analyzed. This means that an analysis of four-item variable would actually result in four components, not two. However, in most analyses only first few components represent meaningful amounts of variance and hence for further analyses those representative components are retained, interpreted, and used (for example, in multiple regression analyses).

For example, if there are four variables to be analyzed in a non-conflicted machine, it is likely that only the first two components would represent meaningful variances. Therefore, only these two components would be retained for interpretation. It is assumed that the remaining two components accounted only for trivial amounts of variance and therefore would not be retained, interpreted, or further analyzed.

V. **Results and Discussion**

After performing the factor analysis, from Table 1 it is seen that almost 57% of the total variance is explained by the first two components (CPU and memory).

Table 1.**Total Variance Explained by Principal Component Analysis**

	Initial Eigenvalues			**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**[a]
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**	**Total**
1	1.246	31.139	31.139	1.246	31.139	31.139	1.239
2	1.033	25.824	56.963	1.033	25.824	56.963	1.044
3	.903	22.569	79.532				
4	.819	20.468	100.000				
Extraction Method: Principal Component Analysis.							

From Figure 1 (scree plot) it is observed that only two factors have an Eigen value greater than 1, which implies that among the four components two factors are extracted and for any further analysis only these extracted factors are significant.

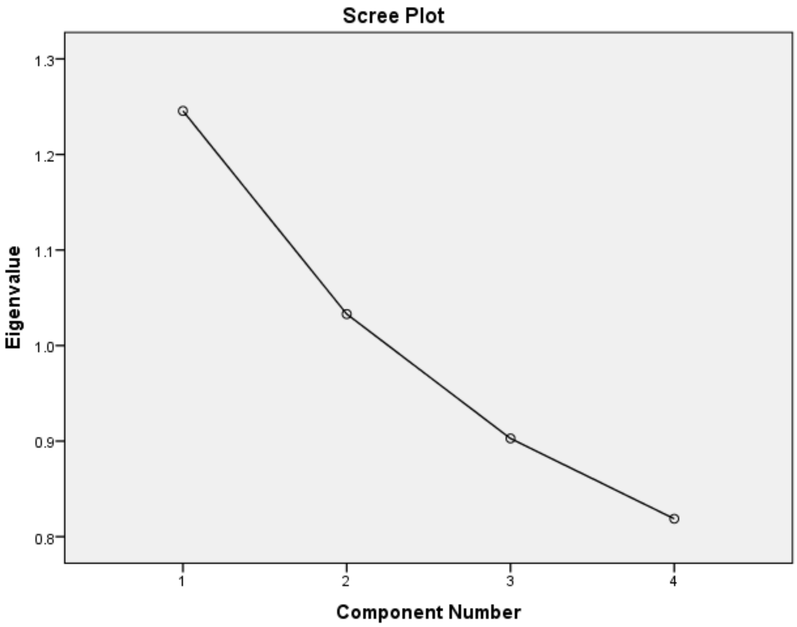

Figure 1. **Scree Plot showing Eigenvalues for each component**

Interestingly, the results indicate that the variables are not correlated at all or even they are negatively correlated except CPU and memory (they have slightly positive correlation among them: 0.140).

Table 2. **Correlation among the Factors**

Correlation Matrix[a]

		CPU	Memory	Disk	Network
Correlation	**CPU**	1.000	.140	-.053	-.069
	Memory	.140	1.000	.018	-.155
	Disk	-.053	.018	1.000	-.022
	Network	-.069	-.155	-.022	1.000

a.Determinant = .951

Analyzing all the data, the authors have found that there are two significant factors extracted and the presence of significant correlation among the variables is not observed. Hence, the initial hypothesis (H_1) is rejected.

5. Monitoring Model

In the previous section of this article, a factor analysis has been performed on a healthy machine not in moral conflict. Presuming that a machine is in moral conflict, the concomitant research questions are:

1. What model would make a successful machine ethics monitoring system?
2. Where would such a model be implemented?

Danks and London (2017) partially address the first question with their concept of Clinical Trials for autonomous vehicles. They propose that small, isolated roll-outs of autonomous vehicles take place that are closely monitored. Akin to clinical trials, the roll-out would be tightly controlled at first to ensure safety standards. As testing progresses, the roll-out could be enlarged to include more vehicles and a larger geographical area.

One particular problem with that scenario concerns the difference between products: drug compounds versus software-controlled vehicles. With drug compounds the formulation requires extensive time and effort to rework. With software there are fixes, patches and updates that are continuously delivered. It is not clear if a Clinical Trial model can apply to modern rapid application development environments (e.g., JAD, Scrum): the software at the end of the trials may (and most probably will) be completely different from that which started the trials.

Another problem with this model is it is a completely outside-in model. Humans monitor the machines and look for anomalies. Consider the consequences of missing an ethical conflict or misdiagnosing such. Furthermore, consider a human (or a machine) that deliberately hides ethical conflicts.

Proposed Model

With regard to the monitoring model, the authors of this paper propose a multi-level machine ethics monitoring model. To accomplish this, ethical monitoring systems would be in-place at a variety of levels within the AS. At the firmware level, much like in Asimov's Three Laws, there would be BIOS level AMA protections. In the current Von Neumann architecture, any software computed action would have to pass through this level in order to be actuated. Appropriate, however limited, ethical inspection can occur here. As such, the authors term this the Intuition Level.

Consider an autonomous war fighting system. The AS may be programmed to kill a human being and the AS would attempt to pull the trigger. However, the firmware protections and BIOS would prevent the hardware from cooperating. The AS would *feel* (Intuition) this action is wrong and be unable to comply. Firmware tampering would be difficult at this level. Unlike software residing on media or in RAM, firmware commands are burned into the chipset and ROM. It would be highly difficult and unlikely to reprogram, override, or circumvent a ROM-based ethical governor system.

An issue with firmware AMAs is visibility. Firmware is generally proprietary and unavailable to the public and not accessible on a normally operating machine. This lack of visibility could create Trojans and Backdoors burned in at the factory. However, the proposed multi-level machine ethics monitoringarchitecture should discover this at the other levels. Given the resource limitations of BIOS and the need for fast execution, ethics monitoring at this level would deal with absolutes of right and wrong. Injury, death and harm are evaluated at this level. Complex ethical reasoning is left to other levels. This is due to a number of factors including resource limitations at the BIOS as well as computational overhead.

The next level of ethical monitoring is a more conscious ethical reasoning and is facilitated through virtualization. The authors propose that an entire AS reside in a virtual machine (VM). As with all VMs, the existence of the VM depends upon a Hypervisor and virtualization software, both of which are off-limits to the VM. The host system and external environment is abstracted from the VM (including the hardware) and interaction with the larger system is facilitated through the Hypervisor. The virtualization software can suspend, resume, stop, and start the VM (and thusly the AS), at any time. As more computational resources are available at this level, more sophisticated ethical reasoning and monitoring can occur at this level. The virtualization software cannot directly interfere with the AS computational efforts; however, it does provide boundaries and feedback: the authors term this the Commandment Level.

Commandments are built into the Hypervisor. As the AS interacts with the hypervisor, any transgression against the moral commandments is communicated back to the AS. This keeps the AS informed of its conduct and facilitates the AS learning of moral precepts. The next level of the machine ethics monitoring model resides with the monitored AS (in its own software). The actual AS being monitored resides within the VM outlined above. This provides the ability to deploy a reasonably high-level AMA within the AS running on the VM. These are the ethical reasoning algorithms that can learn, reason, and interact with the AS as well as the environment (Hypervisor). This level is what is typically thought of as AMAs. As such, the authors term this the Conscious Level.

In the conscious level, the AS and the AMA would reside within the same VM. The AS would have direct access to the AMA. As such, the AMA would be vulnerable, yet open, to the AS. This can be compared to the conscious human ethical reasoning process. Humans can override and change their ethical considerations and so can the AMA. At this level, the ethical governors for an AMA are at their most vulnerable. For example, malicious software can attack the visible AMA or the AS itself could consciously override or otherwise influence the AMA.

Finally, there is the Spiritual Level of the proposed ethical monitoring system. It should be recognized that it is naive to believe that an AMA can be programmed to be able to make correct ethical decisions for every situation it encounters. As with humans, there are many ethically challenging situations with limited information, limited time, and limited processing. Unlike humans in these situations, an AS may not exhibit any outward symptoms of ethical conflict, hence, the factor analysis work in the previous sections. When a human is in deep ethical conflict, there are often associated behaviors that are readily apparent (e.g., distraction, frowning, crying, etc.). In future work, the proposed factor analysis above may help in identifying these ethical conflicts in machines.

When an AMA within an AS is deeply troubled, much like humans, it could exhibit pre-programmed, readily apparent symptoms (e.g., "red eyes" (Davison, 2016) or a "red indicator light" as Sottilaro (2004) chides in his Fourth Law of Robotics). This could inform humans or other entities in the local proximity of the ethical conflict and perhaps coax them to render assistance. Relatedly, a deeply troubled AMA can ask humans (e.g., the Creators) for help analogous to the spiritual construct of prayer. An AMA, at the Spiritual Level, can formulate requests for guidance.The form of this prayer-like request could be a query to its programmers, or to other established trusted entities or monitoring systems. AMAs, unlike humans, have readily available and multiple communications channels to their creators. What is lacking is a formalized protocol for initiating and responsible human entities to address these ethical requests.

Consider an Underwriters' Laboratory (UL) model (as opposed to regionally jurisdiction restricted governments such as suggested by Danks and London (2017)) which monitors AMAs and is available (networking conditions cooperating) for consultation and interventions. With a UL-type monitoring system in place, ethical queries from an AS can be received, prioritized and answered. At the very least, the UL-typemonitoring could shut down the AS before any further issues arise. An AS would be required to be compliant and certified through this UL-inspired group prior to being delivered. This would alleviate the burdens of cost, politics, jurisdiction, administration and bureaucracy from governments.

6. Limitations

In this article, one healthy machine was utilized. While it is a sample of the classic Von Neumann architecture, there are limitations associated with results from only one system. Please see the suggestion for future research section for further discussion on that matter.

7. Suggestions for Future Research

As mentioned above, it is suggested that the scope of this research be expanded. This would include the analysis of multiple machines that will provide a large sample set for the factor analysis. These machines would be of the same software and hardware architecture as the initial sample. In addition to above, it is also suggested that multiple computing architectures be analyzed. This would provide insight into other hardware and software platforms that could potentially host an AS and AMA configuration. Given the specific nature of the platform analyzed in this research, it would assist in the generalizability of the findings to include more diverse and heterogeneous computational systems.

Finally, it is suggested that controllable ethical conflicts be injected into the system for study. The same factor analysis could be applied, and any differences noted and examined. This could lead to insights into the nature of unhealthy, ethically conflicted machines. The authors of this paper are currently working to inject Asimov's Three Laws into the studied platform and allow users to software control ethical conflicts. As that is occurring, further factor analysis will be performed on that unhealthy machine.

8. Conclusion

Although it is quite easy to implement some basic moral decisions in a computer, providing a system to address modern ethical decision-making and monitor an ASremains elusive. In this paper, a factor analysis was performed on a machine that was not in ethical conflict. Additionally, a multi-level machine ethics monitoring model was proposed. The authors of this article began by defining key concepts of machine ethics. A literature review was performed in order to provide the necessary theoretical framework for this research article. Then, a factor analysis of a *healthy*, non-conflicted machine and the concomitant methodology and analysis was presented.

After that discussion, a monitoring model was presented as well as a related discussion on where to implement this model and how to monitor an AS. There exists a great deal of debate regarding AMAs. Technology is advancing and Moore's Law is holding if not somewhat slowing. The need for autonomous, ethical systems is growing in areas such as transportation and healthcare. However, humanity has yet to create a protocol for identifying ethical issues in autonomous systems much less creating protocols and systems to address these issues.

9. References

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? *IEEE Intelligent Systems*, *21*(4), 12-17.

Allen, C., Smit, I., & Wallach, W. (2005). Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches. *Ethics and Information Technology*, *7*(3), 149–155. doi:10.1007/s10676-006-0004-4.

Allen, C. (2002), Calculated Morality: Ethical Computing in the Limit, in I. Smit and G.E. Lasker, eds., *Cognitive, Emotive and Ethical Aspects of Decision Making and Human Action,* Volume I (Workshop Proceedings, Baden-Baden, 31.07.–01.08.2002),19–23.

Allen, C., Varner, G., &Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, *12*, 251 – 261. doi:10.1080/09528130050111428

Asimov, I. (1942), Runaround, *Astounding Science Fiction, 29*(1), 94-103

Bechtel, W. (1985), Attributing Responsibility to Computer Systems*, Metaphilosophy* 16(4), 296–305.

Casebeer, W. D. (2005). *Natural ethical facts: evolution, connectionism, and moral cognition*. Cambridge, Mass, London: MIT.

Child, D. (1970). *The Essential of Factor Analysis*. London: Holt, Rinehart & Winston.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul : a philosophical journey into the brain*. Cambridge, Mass.: MIT Press.

Churchland, P.M., and Churchland, P.S. (1990). Could a Machine Think? *Scientific American* 262, 32-39.

Danks, D., and London, A. J. (2017). Regulating Autonomous Systems: Beyond Standards. *Intelligent Systems 32*(1): 88-91.

Davison, C. & Allen, C. (2017). Approaching machine ethics: Topics in technical education. *The CTE Journal, 5*(1), 31-41.

Etzioni, A., & Etzioni, O. (2016). Designing AI systems that obey our laws and values. *Communications of the ACM*, *59*.

Floridi, L. and Sanders J.W. (2001), On the Morality of Artificial Agents, in L. Introna and A. Marturano, eds., *Proceedings Computer Ethics: Philosophical Enquiry – IT and the Body*, Lancaster, 84–106.

Gips, J. 2011. Towards the Ethical Robot. In *Machine Ethics*, M. Anderson and S. L. Anderson, Eds. Cambridge University Press, Cambridge, UK, 244-253.

Gips J (1995) Towards the ethical robot. In: Ford K, Glymour C, Hayes P (eds) *Android epistemology*. MIT Press, Cambridge, 243–252.

Gotterbarn, D. (2002), The Ethical Computer Grows Up: Automating Ethical Decisions, in I. Alvarez et al., eds., *The Transformation of Organisations in the Information Age: Social and Ethical Implications,* Proceedings of the sixth ETHICOMP Conference, 13–15 November 2002, Lisbon, Portugal, Lisbon: UniversidadeLusiada, 125–141.

Grint, K. and Woolgar, S. (1997), *The Machine at Work: Technology, Work, and Organization*, Cambridge: Blackwell.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Hair, J.F., Black, W.C., Babin, B.J. & Anderson, R.E. (2010). *Multivariate Data Analysis: A Global Perspective*. Upper Saddle River,New Jersey: Pearson Prentice Hall.

Johnson, D.G. (2001), *Computer Ethics*, 3rd edition, Upper Saddle River, New Jersey: Pearson Prentice Hall.

Jordan, N. (1963), Allocation of Functions Between Man and Machines in Automated Systems, *Journal of Applied Psychology 47(3),* 161–165.

Lenk, H. (1994). Macht und Machbarkeit der Technik.

Lindner, F., Bentzen, M. M., &Nebel, B. (2017). The HERA approach to morally competent robots. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* doi:10.1109/iros.2017.8206625.

Mowbray, M. (2002), Ethics for Bots, in I. Smit and G.E. Lasker, eds., *Cognitive, Emotive and Ethical Aspects of Decision Making and Human Action*, Volume I (Workshop Proceedings, Baden-Baden, 31.07.–01.08.2002), 24–28.

Nagenborg, M. (2007) "Artificial moral agents: an intercultural perspective." *International Review of Information Ethics* 7(9),129-133.

Rea, S. (2017). AI Ethics Experts Propose Driverless Car Regulations Similar To Drug Approval Process. Retrieved February 16, 2017 from: https://www.cmu.edu/news/stories/archives/2017/february/driverless-car-regulations.html

Searle, John. R. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences 3*(3), 417-457.

Sottilaro, M. (2004). The three laws of robotics. Retrieved February 16, 2018 from:
https://en.wikipedia.org/wiki/Three_Laws_of_Robotics

Stahl, B.C.(2004), Information, Ethics, and Computers: The Problem of Autonomous Moral Agents. *Minds and Machines, 14*, 67-83.

Stahl, B. C. (2001, October). Constructing a Brave New IT-World: Will the Computer Finally Become a Subject of Responsibility? In *Constructing IS Futures–11th Annual BIT2001 Conference, Manchester, UK* , 30-31.

Stahl, B.C. (2000), 'Das kollektiveSubjekt der Verantwortung', in *ZeitschriftfürWirtschafts- und Unternehmensethik* 1/2, 225–236.

Stewart, T. R., Roebber, P. J., &Bosart, L.F. (1997). The Importance of the Task in Analyzing Expert Judgment. *Organizational Behavior and Human Decision Processes, 69*(3), 205-209.

Turing, A.M. (1950), 'Computing Machinery and Intelligence', *Mind 59 (N.S. 236)*, 433–460.

Wallach, W., & Allen, C. (2009). *Moral machines: teaching robots right from wrong*. New York, NY: Oxford University Press.

Weckert, J. (2001). Computer ethics: Future directions. *Ethics and Information Technology*, *3*(2), 93-96.