

ALMA MATER STUDIORUM - UNIVERSITA' DI BOLOGNA

SCUOLA DI SCIENZE POLITICHE

Corso di laurea magistrale in

Comunicazione Pubblica e d'Impresa

TITOLO DELLA TESI

**Artificial Intelligence and Human Rights:
Case Study of Sophia the Robot**

Tesi di laurea in

Big Data and Cybersecurity

Relatore: Prof. Giacomello Giampiero

Correlatore: Prof. Massimiliano Fanni Canelles

Presentata da: Simin Liang

Appello

II

Anno accademico 2017-2018

Index

Introduction.....	4
Chapter 1. Theory of Artificial Intelligence.....	7
1.1 History of Artificial Intelligence.....	7
1.2 What is Artificial Intelligence?.....	13
1.2.1 Turing Test.....	14
1.2.2 The Loebner Prize.....	15
1.2.3 ELIZA Program.....	18
1.2.4 The Chinese Room.....	20
1.3 Artificial Agent and the Account of Agency.....	23
1.3.1 Causal Agency.....	25
1.3.2 Intentional Agency.....	25
1.3.3 Triadic Agency.....	26
1.3.4 Moral Agency.....	28
1.4 Artificial General Intelligence (AGI).....	30
Chapter 2. The Approach of AI.....	32
2.1 Machine Learning.....	32
2.1.1 What is Machine Learning (ML).....	33
2.1.2 Deep Learning (DL).....	41
2.1.3 The Neural Networks.....	43
2.2 The Applications of Deep Learning.....	44
2.2.1 Computer Vision.....	44
2.2.2 Speech Recognition.....	46
2.2.3 Natural Language Processing (NLP).....	47
Chapter 3. Case Study: Sophia the Robot.....	49
3.1 Background of Sophia the Robot.....	49
3.1.1 Who or What is Sophia the Robot?.....	49

3.1.2 What Does Sophia Want?.....	52
3.1.3 What Is Behind Sophia?.....	59
3.1.4 What Sophia Serves For?.....	60
3.1.5 Does Sophia Deserve A Citizenship?.....	63
3.2 Algorithm and Ethic.....	66
3.2.1 The Social Impact of Sophia.....	66
3.2.2 Algorithmic Error.....	75
3.2.3 Algorithmic Bias.....	77
Chapter 4. Ethical Principles of AI.....	79
4.1 Asilomar AI Principles.....	79
4.2 The Legal Status of Robots.....	82
Conclusion.....	86
Bibliography.....	89
Website.....	92
Appendix I.....	94
Appendix II.....	97

Introduction

On 25th October 2017, a robot named Sophia become a citizen in Saudi Arabia. The news shocks all over the world, because we don't know how to respond to this decision making by the Saudi Arabian government. Sophia suddenly became famous over night, having over 93,000 followers on her twitter, participating one of the famous talk show Jimmy's Tonight Show, taking the cover pages on some famous fashion magazines, and she even made a speech about humanity on the United Nations Development Program (UNDP). She's definitely a celebrity, and she is a legal robot citizen. But what is "robot citizen" anyway? There wasn't such a thing before. Her case becomes a hot debate topic among the AI researchers and it gets lots of public attention. Being a citizen is getting an honor from a country, it means you can have the equal rights and duties as the other citizens. In other word, Sophia has the same equal rights as other Saudi Arabian citizens, and officially has an identity in the world if she travels. The public starts to get worried if we're ready to have robot citizens or not. Will robots domain the human world in the future just like in the fiction movies? How smart these robots are? How intelligent Sophia is so that she can have that privilege to be a citizen? Should we worry? With all these questions and concerns, it motivates me to write this thesis and to open this Pandora's box.

I will start to talk about the history of artificial intelligence in Chapter 1 which will show us a brief development of artificial intelligence back to 1950s. It is essential to know some important breakthrough programs which help us to understand artificial intelligence, such as the famous Turing Test in Section 1.2.1, which is still the standard test in the academic point of view. In Section 1.2.2 I will briefly talk about the current Loebner Prize which is still on going to encourage the development of AI. The reason I mentioned about ELIZA Program in Section 1.2.3 is not only because it's a famous program in the AI history, but also it shows how the chatbot works and generates conversations. It's important to understand the basic knowledge of the functions of chatbot which also runs on Sophia. You will be easily find out that the conversation generators in the system is not possible analogous to human level. In Section 1.2.4 I will also talk about the famous debate and theory in the history, that is the

“The Chinese Room” against the Turing Test, it helps us to understand better if we should consider the AI system as a whole or not.

In Section 1.3 the artificial agent and the account of agency brings the basic theory of the ethical issue of AI system, we will see from the academic point of view, there are causal agency, intentional agency, moral agency and the new notion of triadic agency by Johnson and Verdicchio(2018). This section is one of the most important part in this thesis, because we will discuss in the next few chapters whether Sophia has the moral agency, the account of agency distinguishes between human and non-human. In section 1.4, I will bring up the concept of artificial general intelligence (AGI) which is the key in this thesis, many people never heard of AGI, but some experts and researchers who are working on the advanced AI field are aware of it. Because of that, many people assume the AI will conquer the world by taking over their jobs or destroying civilization, but they don't really know the scenario they are afraid of, is actually came from AGI. It is AGI that gets people panic, especially regarding to the human rights.

In Chapter 2, I will take a few paragraphs to talk about the basic knowledge of the artificial intelligence and its well known supervised learning algorithm and unsupervised learning algorithm which are all belong to machine learning. I will use some simple examples to illustrate the concepts of machine learning, deep learning and the neural networks, which show what artificial intelligence fundamental systems are, and finally I will show you the types of applications including computer vision, speech recognition, natural language processing which are widely used in all kinds of AI system.

In Chapter 3, I will focus on the case study of “Sophia the Robot”. The background of Sophia will be shown in section 3.1.1. Since we know nothing about Sophia at all and she is a legal citizen, I will treat her like a real person who has needs or intentions. I will check on what she's posting on her twitter by using KNIME, so that I can analyze her posts and the messages she's delivering to the public. In section 3.1.3, I will tell you what systems are running on Sophia so that she can appear to be a woman that can talk, and in section 3.1.4, I will find out why Sophia keeps appearing to the public and what's her job. In section 3.1.5, I will talk about the most controversial issue: does Sophia deserve a citizenship at all? Later on, I will discover more about the public opinion on the future legislation for AI and AGI. I originally made a survey, targeting on the objects who already understand AI and AGI, so that

they can have a better vision on this topic, the result is very interesting and quite surprising which gives me a broaden view on this issue. In section 3.2.2 & 3.2.3, I will show you more consequences of algorithmic ethical cases which arouse public attention on the ethic of AI.

In the last chapter, I will show you the worldwide AI researchers and EU government are doing their best to proactive on the fast developing AI technology. The Asilomar AI principles are mostly recognized by the scholars and researchers at the moment, EU committee is proposing the new law regarding to Robotic, there would be “electronic person” in the future if the law allows, which is still under discussion. We are creating a new specie in the world which still under human’s control, that’s why it’s so important for us to make it right at the beginning, and we all should be part of it.

Conclusion

Artificial intelligence is absolutely a hot topic these days, not only because it is widely applied in business, but it also inspires many Hollywood film makers. From the famous and successful American TV series “Westworld” to a newest scientific fiction movie “Upgrade⁶³” describing a paralyzed man is manipulated by a “chip” which allows him to control his body, artificial intelligence already arose the debates among researchers and the public. Sophia granting the citizenship in Saudi Arabia is a special case that should not happen at the beginning.

My agreement in this thesis is to disapprove the decision of Saudi Arabia granting a citizenship to Sophia. First of all, Sophia is not qualified to become a citizen according to Saudi Arabian Citizenship System; secondly, Sophia is not a social entity because she is not an moral agent; last but not least, it’s not a smart decision or I should say it’s too early to grant a citizenship to a robot.

After learning the systems behind Sophia, we know that she is a machine with inputted scripts when she’s delivering a speech in public, and she is far away from being artificial general intelligence (AGI). Sophia can see, hear and speak by operating machine learning, but she is not capable of working autonomously. I used KNIME to analyze Sophia’s twitter and I don’t find that she’s promoting humanity nor gender equality as what she speaks in the public, there are lots of jokes on her twitter though. From the result of the poll, we can see that her humor actually makes the public adore her. Therefore, she is a celebrity with no specific object, but being the tool of a marketing strategy for Hanson Robotics. Sophia is not a moral agent, and she does not deserve to have a citizenship by using the current systems. My assumption of her citizenship is that Saudi Arabia government wants to promote and develop the advanced technology without considering the ethical and liable issues, such as human rights. Once we grant human rights to a robot, we cannot deny the fact that they are not toys anymore, we would become the same equal social entity in the world. What they could do or not do is based on their own decision. We cannot stop them to make a new society as what they wish, and we cannot stop them especially when they are much smarter than us. Now you

⁶³ [https://en.wikipedia.org/wiki/Upgrade_\(film\)](https://en.wikipedia.org/wiki/Upgrade_(film))

may think Sophia is just a PR tool and her citizenship is a joke. Imagine one day AGI reaches the human level intelligence, and it's a software, which means it can run on Sophia, by then, she suddenly becomes superintelligence, and don't forget that she is a legal citizen. She can run for presidency, or go to the moon. It is not a joke, it's a serious issue and it may have unimaginable consequences in the future.

Fortunately from the result of the survey, we can see that the public disagree on the decision of granting a citizenship to Sophia, and they think AI and AGI should not have any legal rights. However, my objects of this survey are the people who understand AGI, they have clear mind of how robot works. Most people who haven't learned anything about AI and AGI can be easily confused by the intelligence of a robot. What's worse, the trend of anthropomorphizing robot is highly increased, Sophia is the example. There will be more and more humanoid robots with sophisticated skills in the coming years. Anthropomorphic robots would easily affect human's emotion. If it appears to be beautiful and passionate, people with weak sentiment could be easily obsessed with it and rely on it. This is not a healthy way of living, and it's obviously the problem that AI creates.

What's more, algorithm is the core of artificial intelligence, and it's not perfect, because sometimes there would be errors and bias which may cause vital accidents. AI researchers and regulators are trying to make principles to guarantee the safety of physical, mental and digital human lives. However, there is a communication gap between AI experts and the government, they should work together on all the issues related to AI from both technical and ethical point of views. Legislation is the basic and final guard to protect human.

We are creating a new species in the world which still under controlled by human, that's why it's so important for human to make this right, to embed with the diversity of values and appropriate ethical performance in the algorithm. Hereby, I urge the public to pay attention on the ethical and liable issues on the fast developing AI area:

- 1) We should never grant any legal rights to artificial intelligence. Human control should be always the priority. Let's keep in mind that artificial intelligence is a technology, it doesn't have any feelings, it cannot autonomously perform moral agency. Machine and human are not equal entities, we create machines to help human, not inverse.
- 2) We should make the ethical principles clear and solid by law. Since artificial intelligence develops extremely fast and we are not yet ready to take control of it. We

cannot slow down the development, but we should catch up on it, and take action on the legislation not only when there's an accident occurs. The ethical principles should be always put in the first place. Legislation is not only to help human physical protection (safety) , but also their mental and digital life (data protection).

3) We should cooperate in a global scale. Although there are AI researchers from all over the world holding conferences in different countries, there are rare researchers from Asia and the southern pacific oceans etc. There are different levels of artificial intelligence in the world, especially refer to the autonomous weapons, therefore, a global community such as United Nations has the obligation to make some agreements on the restrictions and protection for all the countries.

4) We should make good use of technology on helping the poor countries as priority. I understand artificial intelligence are monetized applied to the business. It's been used in manufacturing, health care sector, education, etc. But it only brings connivence to the people who can afford it. I believe technology should help the resource-poor counties, such as Africa, bringing food and water to the people who still suffering in the poor counties. This is the ultimate goal of using technology.

5) Anthropomorphizing robots should be avoided. Human is an emotional social entity, we can be easily affected by interacting with other entity. The relationship between human and human is complex, the relationship between human and their animal pets is simple and compassionate. But the relationship between human and humanoid robot could be twisted if human emotionally attach on it. It could distance the human from the social environment, this is not a healthy life and it's definably not helpful to human. Therefore, I think we should make a robot more robust and less human like, so that people can distinguish the human and machine clearly.

To sum up, AGI will come one day, we cannot and should not stop it. To embrace the advanced technology, we must get ready to cooperate and live with it. The scenes in scientific fictions may come true, and we should not let our guard down in any circumstances. At the end, there's an augment "Guns don't kill people, people kill people", which I transform into my augment: artificial intelligence don't destroy people, people destroy people.

Bibliography

- Allen, C., & Wallach, W. (2012). Moral machines: contradiction in terms or abdication of human responsibility. *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, Cambridge (MA), 55-68.
- Andrew Ng .(2013). Machine Learning. *OpenStax CNX*. Available on: <http://cnx.org/contents/4c74010a-38d0-4a6a-9a11-4dfe8a7cbf96@4.2>.
- Ballard, D. H., Hinton, G. E., & Sejnowski, T. J. (1983). Parallel visual computation. *Nature*, 306(5938), 21.
- Bostrom N .(2014). Superintelligence: paths, dangers, strategies. Oxford University Press, Oxford.
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, 86(1), 117-128.
- Chikhale, S.N; Gohad, D.V. (2018). “Multidimensional Construct About The Robot Citizenship Law’s In Saudi Arabia”, *International Journal of Innovative Research and Advanced Studies (IJIRAS)* Volume 5 Issue 1, January 2018.
- Coeckelbergh, M. (2009). Personal robots, appearance, and human good: a methodological reflection on roboethics. *International Journal of Social Robotics*, 1(3), 217-221.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). Cambridge: MIT press.
- Himma, K. E. (2007). Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? 2007 Ethicomp Proceedings. *Global e-SCM Research Center & Meiji University*, 236-245.
- Johnson, D. G., & Verdicchio, M. (2018). AI, agency and responsibility: the VW fraud case and beyond. *AI & SOCIETY*, 1-9.

- Karpathy, A., Johnson, J., & Fei-Fei, L. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*.
- Kotler, P., Kartajaya, H., & Setiawan, I. (2016). *Marketing 4.0: Moving from traditional to digital*. John Wiley & Sons.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
- Malle, B. F., & Scheutz, M. (2015). When will people regard robots as morally competent social partners?. In *RO-MAN* (pp. 486-491).
- Mauldin, M. L. (1994). Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In *AAAI* (Vol. 94, pp. 16-21).
- McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*, ak peters. Natick, Mass.
- Mitchell, T. M., & Michell, T. (1997). *Machine Learning* McGraw-Hill Series in Computer Science. WCB/McGraw-Hill: Boston.
- Omohundro, S. (2016). Autonomous technology and the greater human good. In *Müller V (ed) Risks of artificial intelligence*. CRC Press, Boca Raton, pp 9-27.
- Poole, D. L., Mackworth, A. K., & Goebel, R. (1998). *Computational intelligence: a logical approach* (Vol. 1). New York: Oxford University Press.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.
- Samuel, A.L. (1959). Some Studies in Machine Learning Using the Game of Checkers, *IBM Journal of Research and Development*, vol.3, 1959, pp.210-229. Available on: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf>.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(3), 417-424.

- Schlosser, M .(2015). “Agency”, *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), Available on: <https://plato.stanford.edu/archives/fall2015/entries/agency/>.
- Shi,C.Y., Huang, C.N., & Wang. J.(1993). Principle of Artificial Intelligence. Beijing, China: Tsinghua University Press. Available on:<ftp://218.31.79.211/%B5%E7%D7%D3%CD%BC%CA%E9/035/T/A2037930.pdf>
- Shieber, S.(1992). *Lessons from a Restricted Turing Test*. Technical Report TR-19-92, Harvard University, Sept., 1992. Revision 4.
- Shillingford, B., Assael, Y., Hoffman, M. W., Paine, T., Hughes, C., Prabhu, U., ... & Mulville, M. (2018). Large-scale visual speech recognition. *arXiv preprint arXiv:1807.05162*. Online available on :<https://arxiv.org/pdf/1807.05162.pdf>.
- Turing, A.M.(1950). Computing Machinery and Intelligence. *Mind* 54(236):433-460, October, 1950.
- Veruggio, G. (2006). The euron roboethics roadmap. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on* (pp. 612-617). IEEE.
- Weaver, J.F.(2017). What Exactly Does It Mean to Give a Robot Citizenship?, *Slate*, NOV. 6 2017. available on: http://www.slate.com/articles/technology/future_tense/2017/11/what_rights_does_a_robot_get_with_citizenship.html.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation.
- Yampolskiy RV .(2016). Utility function security in artificially intelligent agents. In: Müller V (ed) *Risks of artificial intelligence*. CRC Press, Boca Raton, pp 115–140.
- Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.

Website Reference:

https://en.wikipedia.org/wiki/Fifth_generation_computer
https://en.wikipedia.org/wiki/Logic_Theorist
[https://en.wikipedia.org/wiki/Lisp_\(programming_language\)](https://en.wikipedia.org/wiki/Lisp_(programming_language))
https://en.wikipedia.org/wiki/Semantic_network
[https://en.wikipedia.org/wiki/Commonsense_knowledge_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Commonsense_knowledge_(artificial_intelligence))
https://en.wikipedia.org/wiki/Commonsense_reasoning
<https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>
<https://www.documentcloud.org/documents/4483190-NTSBuber.html>
<https://sites.google.com/site/narswang/home/agi-introduction#TOC-What-is-AGI>
<https://www.cnbc.com/2016/03/16/could-you-fall-in-love-with-this-robot.html>
<https://www.instagram.com/p/BNhvUfTgNjY/>
https://www.instagram.com/p/BeSNYFmgoz1/?utm_source=ig_embed
<https://www.instagram.com/p/BgMB-ylBt2W/?tagged=cosmoindiaexclusive>
<http://www.bbc.com/future/story/20170906-how-it-feels-to-meet-sophia-a-machine-with-a-human-face>
www.sophiabot.com
facebook.com/real sophiarobot
twitter.com/real sophiarobot
www.instagram.com/real sophiarobot
www.knime.com
<http://hplusmagazine.com/2017/11/05/sophia-singularitynet-qa/>
<http://www.hansonrobotics.com/how-sophia-the-robot-works-goertzel/>
<http://www.abc.net.au/news/2017-10-25/sophia-the-robot-the-future-of-artificial-intelligence-and-ai/9082852>
<https://www.spindox.it/en/blog/sophia-artificial-citizenship>
<https://openai.com>
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
<https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
<https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>
<https://futureoflife.org/principles-signatories/>
<https://futureoflife.org/ai-principles/>
<http://www.europarl.europa.eu/committees/fr/supporting-analyses-search.html>
<http://www.europarl.europa.eu/sides/getDoc.do?type=COMPARL&mode=XML&language=EN&reference=PE582.443>
[https://en.wikipedia.org/wiki/Upgrade_\(film\)](https://en.wikipedia.org/wiki/Upgrade_(film))

Photo:

https://it.wikipedia.org/wiki/Test_di_Turing#/media/File:Turing_Test_Version_3.svg
<http://www.lazytd.com/lti/pub/aaai94.html>
<https://techcrunch.com/2018/05/24/uber-in-fatal-crash-detected-pedestrian-but-had-emergency-braking-disabled/>
<https://www.instagram.com/p/BeSNYFmgoz1/?taken-by=stylistmagazine>

Video:

<https://www.youtube.com/watch?v=TlO2gcs1YvM>
<https://www.youtube.com/watch?v=h0962biiZa4>
https://www.youtube.com/watch?v=PJ_KCTsOCrs
https://www.youtube.com/watch?v=Bg_tJvCA8zw
<https://www.youtube.com/watch?v=kWIL4KjIP4M>
https://www.youtube.com/watch?time_continue=10&v=BwFEFQUDNTs
<https://www.youtube.com/watch?v=dMrX08PxUNY>
https://www.youtube.com/watch?time_continue=10&v=BwFEFQUDNTs
<https://www.youtube.com/watch?v=KVp33Dwe7qA&list=PLpxRpA6hBNrwA8DlvNyIOO9B97wADE1tr&index=6>
https://www.youtube.com/watch?v=Elx3BbwPl_g