Soft-Morality and Artificiality

by

Ritaprava Bandyopadhyay

Thesis submitted for the degree of Ph.D. in Cognitive Science

Under the School of Cognitive Science

Faculty of Interdisciplinary Studies, Law and Management

Jadavpur University

Kolkata

India

November 2022.

List of Publications

Two of my research articles as the first author were published/accepted for publication in blind-double refereed, UGC Care listed journal during the time of pursuing Ph.D. The details are attached:

1. Bandyopadhyay, Ritaprava, Maushumi Guha and Amita Chatterjee. 2022, 'The Correlation Between Fall In Profit and the Deployment of AI Labour: A Marxian Analysis'. *Antrocom Online Journal of Anthropology*, vol. 18. n. 1 (2022) – ISSN 1973 – 2880. http://www.antrocom.net/archives/2022/180122/Antrocom_18-1.pdf

2. Bandyopadhyay, Ritaprava, Maushumi Guha and Amita Chatterjee. Forthcoming, 'Tathya Ebong Samyogamadhyam Prayuktir Prabhabe Sambadmadhyame Naitik Kartar Badoler Ruparekha', accepted for publication in *Rabindra Bharati Journal of Philosophy*. ISSN NO. 09730087.

Ritaprava Bandyopadhyay

9-11-2022.

List of Presentations in National/International/Conferences/Workshops

During the time of pursuing Ph.D, I have attended, presented papers and joined as a resource person in different conferences and workshops, the photocopies of which are attached.

Ritaprava Bandyopadhyay

09.11.2022

যাদবপুর বিশ্ববিদ্যালয়

FACULTY OF ARTS, DEPARTMENT OF PHILOSOPHY

CENTRE OF ADVANCED STUDY, PHASE V (2016 - 2021)

TO WHOM IT MAY CONCERN

This is to certify that Sri Ritaprava Bandyopadhyay is pursuing his Ph.D. in the School of Cognitive Science under my supervision (jointly with Professor Amita Chatterjee). Prior to that, he was my student at the M.A. level where he took the Optional Philosophy of Cognitive Science Papers I and II. During that time and later when he was associated with the erstwhile Centre for Cognitive Science as a part-time Project Fellow UPE II scheme, he attended all seminars, conferences, workshops and other events and made poster and paper presentations individually and jointly. His presentations have always been very positively received by an academic audience.

Maushumi Guha
12.3.19

MAUSHUMI GUHA

Statement of Originality

I, Ritaprava Bandyopadhyay, registered on 06.04.2016 do hereby declare that this thesis entitled "Soft-Morality and Artificiality" contains literature survey and original research work done by me as part of Doctoral studies. All information in this thesis have been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work. I also declare that I have checked this thesis as per the "Policy on Anti Plagiarism, Jadavpur University, 2019", and the level of similarity as checked by iThenticate software is 7 · %.

Signature of Candidate:

Ritaprava Bandyopadhyay · Registration no: D-7/ISLM/24/16

Date : 9 · 11 · 2022

Certified by Supervisor(s):

(Signature with date, seal)

1. Manishita Guha 9.11.2022

Professor
Department of Philosophy
Jadavpur University
Kolkata-700 032

2. Amita Chatterjee 9·11·2022

Amita Chatterjee
Emeritus Professor
School of Cognitive Science
Jadavpur University
Kolkata -700 032, India

CERTIFICATE FROM THE SUPERVISORS

This is to certify that the thesis entitled "Soft-Morality and Artificiality" submitted by Shri Ritaprava Bandyopadhyay who got his name registered on 06.04.2016 for the award of Ph. D. (FISLM) degree of Jadavpur University is absolutely based upon his own work under the supervision of Professor Maushumi Guha and Professor Amita Chatterjee (under the School of Cognitive Science) and that neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

Signatures of the Supervisors and date with Office Seal.

1. Maushumi Guha

Professor
Department of Philosophy
Jadavpur University
Kolkata - 700 032

9.11.2022

2. Amita Chatterjee

9.11.2022

Amita Chatterjee
Emeritus Professor
School of Cognitive Science
Jadavpur University
Kolkata -700 032, India

Content

Preface

First things first, as they say. Let me confess that I never imagined that I could complete an entire Ph.D. thesis! I have very scanty experience in editing some popular books in Bengali, and writing features, post-editorials, obituaries, or reports for newspapers or websites, but writing a thesis was beyond my imagination. This would not have been possible without the help and support of my supervisors, Professor Amita Chatterjee and Professor Maushumi Guha.

Let me confess that, I have witnessed the circle of time in my life, it has sometimes compelled me to take on new challenges and made me run--- courtesy of working in the media. Hence, I ran without knowing where the race would end and in this marathon, many people came to encourage me, helped me, and showed me the right way, so that I could reach the destination.

The journey, indeed, is like a marathon as I told you earlier! I finished my MA in 2006. In the same year, I was absorbed as a part-time project fellow at the erstwhile Centre for Cognitive Science. But the penchant for knowing the self-other asymmetry was formally implanted in the year 2001 when, as a student of undergraduate (first year), I had the opportunity to attend the classes of Professor Amita Chatterjee. Professor Chatterjee, (*Amitadi* as I call her) used to teach us psychology, one of the participatory disciplines of Cognitive Science. During my earlier days, even though I could not entirely grasp her teachings in their total richness, yet like a new raga unfolding its myriad facets before an avid listener, her classes left a lasting impact on me.

Later, on, when I chose Philosophy of Mind and Cognition as my special paper for the M.A. in Philosophy degree, I had been privileged to have *Amitadi* as a teacher. This time I began to enjoy her classes wholeheartedly, achieving far greater comprehension and understanding than in my undergraduate days. I even gathered the courage to ask her questions that I considered being naïve. She always addressed my questions, however trivial they may be. After my post-graduation, I also got the opportunity to work under her able guidance on a project. At that time, she was the Coordinator of the Centre for Cognitive Science, Department of Philosophy, Jadavpur University.

My tryst with the media began as I got a job in a newly launched Bengali daily. Even amid my gruelling schedule --- night duties, work without a break for weeks, etc. --- I could never lose contact with *Amitadi*. We were connected over the phone and e-mail. A few years later, it suddenly came to my mind that I would pursue research for a Ph. D. *Amitadi* was then a Professor Emerita, at Jadavpur University. Taking her suggestion to prepare for the Research Aptitude Test and having been trained by her in the Research Methodology Class, I ended up fulfilling my dream of pursuing my Ph.D. under her guidance. Whenever she found any material related to my research area, she immediately brought it to my notice and discussed it with me and that eventually shaped my thesis. She read the thesis both in parts and in full several times, made corrections, gave invaluable feedback and added so much value to it became easier for me to reach my conclusion. There are no words to express my gratitude towards someone who will be my teacher for life. I pay my regards to her with every breath I take.

Along with *Amitadi,* I would like to acknowledge the contribution made by Professor Maushumi Guha, not only as my co-supervisor but also in my life as a whole. It was 2005. I

was a student of PG I. It was in that year that *Ma'am*, as I call her, joined the Department of Philosophy, at Jadavpur University. *Ma'am* was pursuing her Ph.D. under *Amitadi* and before that, she had received the Jawaharlal Nehru Memorial Cambridge Scholarship and completed her M.Phil from the University of Cambridge, UK. I was happy to have her guidance at the Masters's level. This is when she introduced a relatively new area of study, namely, Folk Psychology. Around this time, *Amitadi* provided me with an article, which claimed that robots can have ethical rights. Since *Ma'am* was interested in this topic, I immediately sought an appointment, and together we read the article.

Upon reading it, some questions came to our minds. We observed that as time flows, the need for the ethical dimension of machines themselves preoccupies ethical and technological theorists. *Ma'am* asked me to write an article on these concerns and we ended up discussing and writing the paper together! Our paper, 'Contextualizing Ethics in the Realm of Robotics', was accepted for presentation at the Wesleyan Philosophical Society Conference on 'Philosophy and Science: Contemporary Explorations', held on Thursday, March 13, 2008, at Duke University Divinity School, USA. Even though we could not manage the funds to present it physically, our paper was read out by the chair of that session.

Inspired by this success, I wrote my second paper on the interface of ethics and robotics, this time in Bengali. This paper was accepted and published in the *Jadavpur Journal of Philosophy*. When I was writing these articles, *Ma'am* had given me a book by Luciano Floridi. That was my first encounter with the term, 'Information Ethics', which is one of the key research areas in my thesis. Information ethics was a new concept at that time and it was hard to comprehend, and needed guided reading. *Ma'am,* amid her busy schedule, helped me immensely to understand those difficult areas. After a few years, when I was admitted to the

Ph.D., I first approached her to be my supervisor and she was kind enough to accept me as her scholar. She provided me with new books, and whenever I would seek an appointment, she accommodated me, read the thesis, and discussed the writing in depth. Ethics, artificial intelligence, and humans occupy major chunks of my thesis, the ideas and interests which I developed during her classes, over our informal chats at *Milanda*'s canteen while sipping a cuppa tea and also during our phone conversations and WhatsApp exchanges.

Now I want to mention the contribution made by Dr. Ritajyoti Bandyopadhyay, my older cousin. *Dadabhai,* as I call him, is my constant source of inspiration. It would not have been possible for me to get myself admitted to Jadavpur University had *dadabhai* not been there and had he not guided me as my pole star. From my undergraduate days till now, *Dadabhai* remains my true guide in every sense of the term. When I face any challenge in my professional or academic life, the first person I call or leave a message with, is *dadabhai*. He gave this thesis a new dimension when he first introduced me to the Actor-Network Theory formulated by Bruno Latour and asked me to find out why Artificially Intelligent Agents as labourers could not qualify as 'labourers' in Latour's original scheme of things. He also introduced me to Marx's value theory of labour. Marxian literature also requires guided reading. He taught me this theory from the various locations where he was stationed at different points in time and helped me devise an equation that proves Marx's hunch (to demonstrate the reason behind the slowing down of the profit rate with the implementation of artificially intelligent agents as labourers).

I have received immense help and support from Professor Amrita Basu, Director, School of Cognitive Science, Jadavpur University. Apart from her constant encouragement, she helped me in obtaining two Ph.D. extensions. Without her help, I could not have finished

my dissertation. I received valuable advice and encouragement from senior teachers like Professor Lopamudra Choudhury, Professor Mihir Chakraborty, Professor Sibaji Bandyopadhyay, Professor Prajit Basu and Professor Amit Konar which has shaped my thought. Professor Prajit Basu was a member of my RAC and as such, he gave a detailed analysis of the work that I presented before him at the RAC meetings. His final suggestions have helped immensely. Sibaji*da*, listened to the synopsis and gave some valuable suggestions, often enquired about my progress even when he was ill. I don't know how to thank him.

My father, Sri Kalyankumar Bandyopdadhyay, is my first teacher. I have been learning from him since childhood. Apart from continuously motivating me to finish my thesis, *baba* was kind enough to read the chapters, helping me paraphrase some of the sections, and pointing out errors that escaped my attention. Even during the course of writing this acknowledgement, he patiently listened to me and gave me some valuable suggestions. It is not easy to acknowledge one's parents but I wish to register my deep debt to my father here even about this intellectual work of mine.

My mother, who has undergone a series of operations in her life has always kept track of my progress and created a space where I can enter my cocoon and concentrate on my study. All of this despite her frail health. Even in the middle of all the household work, she found time to listen to some sections of my chapters. Nothing in the world can help me express my grateful thanks to my mother for whom my thesis has seen the light of day.

Dr. Malini Siddhanta, was also a big inspiration to me. She is currently teaching History at Lady Brabourne College and her thesis was itself a source of inspiration for me. She occasionally provided me with valuable advice, patiently listened to some of my chapters,

and made suggestions for changes. One thing I must acknowledge is that Malini made room for me by relieving me of many family obligations over the last seven or eight years, while I was occupied with my thesis. I am grateful for whatever she has done for me.

Dr Ujjwal Siddhanta and Mrs Indrani Siddhanta, my parents-in-law, have always been by my side. *Baba*, as I call Dr. Siddhanta, was a phenomenal student himself and a well-known sportsperson. H helped me a lot, not only by providing ample inspiration for my work but also by helping me stay fit during my covid-19 infection as well as other illnesses. *Ma*, Mrs. Siddhanta, a spiritual person at heart, prayed to her god that I could finish the dissertation while working in the corporate sector, knowing well that my professional life was demanding and tiring, occupying my entire day.

Dr Aloka Siddhanta, my aunt-in-law or *Pishimoni*, was a student of Philosophy and retired from College, often asked about my progress. She had provided me with some books from her collection. I am grateful for her support.

Mr Alok Bandyopadhyay (*Jimoni)* and Mrs. Shipra Bandyopadhyay (*Mamoni),* Dr. Anwesha Bandyopadhyay (*didi),* Sankar Bhattacharya (Sankar-*da),* Dr. Debarati Bagchi *(Debaratidi),* Sohini Siddhanta *(Rumpi),* were supportive throughout the journey and I know they would be very happy to see me receiving the Degree. I am sure *Anando, Oishi, and Abhijnan* would be delighted and one day they would recognize my effort.

Mr. Ritwik Bhattacharya, and Mrs. Arunima Bhattacharya read some parts of the thesis and suggested some corrections which I will never forget. I hope to receive their help in coming days.

Thanks to Mr. Achintyarup Roy, the then News Editor of *Ei Samay Sangbadpatra*, for providing me with a 'no objection certificate', required for pursuing the course. I thank Mr. Rupayan Bhattacharyya, my former vertical head, to grant me a study leave just before the final semester of Course Work. I also thank Mr. Mukul Das, News Editor, and Mr. Anindya Jana, Editor, *Anandabazar Online,* for granting me leave so that I could finish writing my dissertation. I am thankful to some of my present and former colleagues like Roshni Mukherjee, Saubhik Ghosh, Jaydip Banerjee, and Shiladitya Saha.

During my undergraduate and post-graduate days, I learned a lot from my classmate Mr. Dipankar Roy. I know Dipankar would be very happy to know that I have completed my thesis. During the Course work, I befriended Ritu Bhattacharyya, Moumita Bhowmick, Bicky Mahata, and Kutubuddin Sheikh. I cherished every bit of their companionship during and even after the coursework. I want to thank Mr. Susovan Pramanik for being my special friend on this journey. Whenever we used to meet or talk over the phone, he would ask about my progress and whenever I felt slightly depressed for not being able to do justice to my thesis due to my professional commitments, he used to say, "You will surely get another extension and will be able to finish the dissertation.'' It healed. It helped immensely. It's true that some relationships remain forever.

I am indebted to my friends Dr. Rajat Subhra Chakraborty, an erstwhile faculty of IIT-Kharagpur, Mr. Anirban Dutta Choudhury, Senior Scientist at TCS, and Dr. Sunando Patra, Assistant Professor of Physics at Bangabashi College for the formal and informal chats I shared with them. Some of them provided research articles whenever I needed them. Since they are close friends of mine, I think mere formal thanks would not be enough for them. I'd like to express my gratitude to Anirban (Joy), who helped me keep the dissertation free of

academic dishonesty. It is because of him that I was able to keep iThenticate's similarity index within 7%. I must acknowledge Professor Prasanta Sahoo, Department of Mechanical Engineering, for his assistance in helping me understand the nitty gritty details of iThenticate towards the close of writing and editing my dissertation.

I was fortunate to have some illustrious teachers in School and at the University. Mr Arun Banerjee, was the headmaster of my school. I still can remember his board work in his bewitching handwriting. I thank all my teachers of my university days—Hiranmay Banerjee, Amita Chatterjee, Shefali Moitra, Tusharkanti Sarkar, Tapan Kumar Chakraborty, Indrani Sanyal, Chhanda Gupta, Proyash Sarkar, Maushumi Guha and Smita Sirker. I have learned a lot from them.

Nainadim (didima), passed away peacefully last year during the lockdown. I know, she would have been very happy to know that I have finished writing my dissertation.

Maam (thakuma) left this world for good in 1996. I was in class VIII at the time. I know how much she adored and wished me. Every day, I feel her blessings.

Last but not the least, there is a small person who has suffered a lot for the last five years for my preoccupation. She sometimes got cross with me and tried to distract me but all her attempts acted as an inspiration to finish my dissertation quickly. She is Riti Banerjee---my five-year-old daughter. Precisely, she is the reason I could finish my Ph.D.

In this research, a qualitative approach has been adopted. I have cited some examples from newspaper reporting. As a professional journalist, I believe that conceptual analysis and critical consideration of problems that touch upon our lives must be situated in actual happenings around us. Newspapers and other news media are good source of information on

these happenings. Literature survey of the theories and concepts in this domain has been the other pillar of this thesis. I have examined and critically considered the research already conducted in this area, tried to point out the research gaps, cited and formulated various thought experiments, and tried to explain the research problem taking a cue from Marxist ethics as a case study. After that, I have tried to engage briefly with the Husserlian notion of the lifeworld.

I am thankful to various libraries starting from the Departmental and Central Library, Jadavpur University, Library at the Schools of Cognitive Science, and many online Libraries from where I collected the research materials for this dissertation.

I have tried to be as true and sincere as possible to this academic work and I take responsibilities for any small errors or omissions that may remain.

Chapter 1

Introduction

The word ‘robot’ has just crossed its hundredth anniversary in the year 2021. Coincidentally, I began writing this dissertation that year! Eminent Czech writer Karel Capek, in his inaugural work *Roussum’s Universal Robot* (RUR), introduced this word for the first time. The play revolves around a bunch of artificially ‘intelligent’ labourers who would eventually displace their bosses. On the contrary, in Kazuo Ishiguro's *Klara and the Sun*, there is a story of another type of artificially intelligent agent. These are not like the irritated Robots portrayed in Karel’s play. In Ishiguro's story *Klara* isn't a hero. She is kind, easily scared, and mortal. She was created to provide friendship to a lonely child. Klara addresses us in her vulnerability. (Unudurti, Jaideep. 2021)

It was the summer of 2017. To me, this season in Gangetic West Bengal is very cruel. Being an employee in the service sector, I spend long hours in the office. Amidst this, I got admitted to a Ph.D. programme at the School of Cognitive Science, Jadavpur University.

One day while I was busy in my office, Professor Amita Chatterjee, one of my supervisors, called me up to talk about an editorial column published in a Bengali daily *Anandabazar Patrika* which discussed a draft report by the European Union parliament.

The Legal Affairs Committee of the European Union Parliament prepared a report which proposed to ascribe 'electronic personalities' to Artificially Intelligent agents (AIAs)[1] and self-learning robots (Committee on L.A. EU Parliament. 2016). This created a furore. The editorial column argued about whether personhood should/could/would be granted to robots.[2]

Meanwhile, experts from the various EU Member States sent an open letter to the European Commission. They expressed their disagreement from various legal and ethical perspectives.

During this time, I came across another piece of news, which stated that the European Union's Sovereign data security specialist, the data safety controller, set up the Ethics Advisory Group. Its goal was to investigate the issues faced by digital advancement and current regulation, particularly the GDPR (The General Data Protection Regulation). (EU 2016).

[1] A philosopher like Dennett would argue that an artificial system can have mental states and intentional agency. This is known as an instrumentalist stance. Realists, on the other hand, contradicts this notion. To them, it is far from obvious. Some scholars believe that even if artificially intelligent machines are incapable of acting on their own volition, as proposed by standard theory, they might be competent in other kinds of agency. Minimal agency, they think, cannot not necessitate the ownership of mental states. Rather, it requires responsive legislation of the agent's environmental coupling as well as biochemical self-maintenance. It is assumed that a moral agent should also be able to satisfy at least a few of morality's requirements. The notion of artificial agency is the primitive conception that I have used in my thesis. In this thesis, it is not my intention to establish the possibility of artificial agency. On the contrary, assuming such agency is possible as in cognitive science, artificial intelligence, and Robotics, I shall consider the possibility of applying notions that we generally apply to other ethical agents. Without going into the question of whether this notion is a comprehensible or cohesive or acceptable one, allow me now discuss about the notion of ethics we can conceive in terms of artificial agents. Minimally we can say that, wherever there is an agency, there is this notion of responsibility, action, and free decision-making and ethics. (Schlosser, Markus. 2019).

[2]. It was 1956. There held a conference at Dartmouth College. On that conference John McCarthy introduced the word artificial intelligence (AI). After that it flourished. In the domain of AI, the concepts of agency, autonomy, and intelligence are all hazy and difficult to define. Furthermore, agency is inextricably linked to qualities such as 'autonomy', 'situatedness', and 'embodiment'. Many scholars avoid providing minimal definitions because such definitions are invariably either too broad or too narrow. According to Russell and Norvig (1995), the concept of an agent is intended to be a device for inspecting system, rather than an utter classification that divides the entire world into agents and non-agents. Furthermore, Florian (2003) holds that the various definitions available in the literature are frequently inconsistent with one another. For our current purpose, we will take Artificially Intelligent Agents such as Robots are machines whose shape varies and whose decision-making abilities and actions are based on algorithms. In this dissertation Artificially Intelligent Agent, Robots, Automata, Intelligent Machines are used interchangeably. (Russell, S. J., & Norvig, P. 1995. And Florian, R˘azvan V. 2003).

At a glance, however, this may seem merely to be a piece of news, new information. However, to a researcher in Cognitive Science, it gives something beyond the 'information'. It entails that an institution has formally recognized the need for ethical thinking about machines. My dissertation, indeed, owes much to this kind of thinking which involves ethics, artificial intelligence and humans.

Luciano Floridi (2018), moreover, pointed out that the report published by the EU's Data Protection Supervisor's advisory group is important for the moral democratic accountability of the digital society in the European Union. It entails that, officially an institution recognizes the presence of ethics in the digital realm.

Accordingly, these two pieces of news point out the subtle change in the relationship between Humans, Artificially Intelligence Agents and Ethics.

After a few years, I came across another document. This tells us that in November 2021 the 193 member states of UNESCO's General Conference accepted the *Recommendation on the Ethics* (2022) of AI. It was the first worldwide standard-setting event on this subject. This is a ground-breaking agreement at UNESCO on how government and tech companies should design and use AI (Koukku-Rondem, Ritva and Ramos Gabriela. 2022). The principles took two years to develop. It aims to profoundly alter the balance of power between citizens, businesses, and governments through the development of artificially intelligent agents. Countries that are UNESCO members accepted to put this *Recommendation* into action by endorsing policies that govern the entire Artificially Intelligent Agents system's life cycle, from the study, strategy, and development to application and usage (Koukku-Rondem, Ritva and Ramos Gabriela. 2022). This means that affirmative action must be used to ensure

different minority groups and representatives of different sexes and genders are represented on the AI design team. This could take the form of quota systems that ensure the diversity of these teams, or it could take the form of dedicated funds from their public budgets to support such inclusion programmes. The report also emphasizes the significance of proper data management, privacy, and information access. It underscores that individuals must maintain control over data, enabling persons to admit and use it as necessary.

It also appeals to its member countries to develop proper safeguard arrangements for the dispensation of delicate data, as well as operative accountability and redressal system in the event of destruction. All of this raises the bar for enforcement. The principles of the *Recommendation* have already been used in AI regulation and policy in a number of countries, demonstrating their practical viability. Finland is an example of good practice in this regard[3]; its AI strategy was the first of its kind in any European country and demonstrated how government can effectively promote ethical AI use without jeopardizing the desire to be at the forefront of new technology (UNESCO. 2022).

Moreover, 'robots, work, and social impacts' were given special consideration in UNESCO's 'Recommendation'. It has been observed that there is debate about robotics, employment, and labour. It says, Robots, as artificially intelligent agents, have been linked to an increase in global productivity. It has been observed that increasing production through the use of AIAs results in a significant reduction in labour costs. As a result, the report concluded

[3] In 2017, Finland launched one of the world's first national artificially intelligent strategies and action plans in order to boost artificially intelligent research and education. The government is already incorporating artificially intelligent agents into its functionings to improve effectiveness and service delivery. For example, the Ministry of Finance launched the 'Aurora AI' programme, which assists individuals and companies by recommending services based on their requirements. Finland has the potential to more than double its economic growth rate by 2035 if AI is successfully applied (Accenture and Frontier Economics 2017). Finland provides top-tier AI education in universities and raises citizen awareness through open online courses like Elements of AI. (Keski-Äijö, Outi etal. 2021.)

that robotics has a global impact on employment and the nature of work. It has been mentioned in the report that it is now difficult to provide exact figures, but many jobs around the world will be converted or may evaporate as a result of the increasing use of Artificially Intelligent Agents such as robots. (UNESCO 2022.) The report points out that when robotics removes a specific type of work, the work that has become automatic becomes invisible and disappears from social value. (UNESCO and COMSET. 2015) In my thesis, I will contest this notion and will show that artificially intelligent agents cannot replace human labourers in certain situations.

Meanwhile, India is advancing towards developing accountable and moral AI governance. NITI Aayog's 'hashtag AI for all' campaign and several business plans have been implemented to guarantee that Artificial Intelligence is advanced with shared humanistic values in its heart (Roy, Anna. 2021). I will not delve deep into the detail as it falls outside our research area. I intend to show that recently the discussion regarding the ethical use of AI has got momentum globally. In the whole narrative, human beings, Artificially Intelligent Agents and their relation to ethics become the central theme. Thus, my research revolves around these three pillars; human beings, Artificially Intelligent Agents, and their relation with ethics.

The literature on robot ethics has emerged since the 2000s, with both enthusiastic and critical reflections since 2010. Philosophically, the ethics of robots are linked to the post-humanism debate (UNESCO and COMSET. 2015). A new epistemology that is not anthropocentric and not based on Cartesian dualism is purported to be offered by post-humanist theory. It aims to dissolve the conventional distinctions between technology, animal, and human. Many different kinds of ethical questions emerge from this. Following UNESCO's report I can classify these questions under the following heads:

a. **Responsibility**

Under this head, the main questions are: Can an Artificially Intelligent Agent be held accountable for its deeds? If some untoward incident occurs, then who will take the liability—AI Agents, its makers or those who write the codes? Who is responsible, the robot, the manufacturer, the software developer, the designer, the user, or the one who controls the robot? The logical consequence is that it is the absence of a sense of accountability among the concerned persons who ascribe different activities to the Artificially Intelligent Agent.

b. **Autonomy**

Can we speak about the autonomy of an Artificially Intelligent Agent? Is it meaningful to mean that an Artificially Intelligent Agent is autonomous? If yes, can the robot be considered to be a moral agent?

c. **Emotions**

The main questions are: should human emotions be mimicked by humanoid robots? Are the animal emotions for robots be mimicked by the company? Are we creating new types of relationships? What new behaviours or attachments can this induce socially?

d. Shield of privacy

The shield of confidentiality is a huge challenge for ICT (Information and Communication Technology). When we use robots for surveillance, then also this challenge remains. In modern times it can be asked, how to protect privacy? Protection of privacy is a big challenge in the present time. It is not only needed for humans but also for artificially intelligent agents.

e. Deskilling of sentient

The excellence of artificially intelligent agents renders an individual incompetent. The problem is that, if a professional is replaced by an Artificially Intelligent Agent, the professional gradually drops her skill.

f. Living beings' current societal reliance on an automaton atmosphere

The rising difficulty of Sentient-Artificially Intelligent Agents interaction makes individuals more susceptible when disasters or interruptions (bugs, power outages, etc.) occur.

g. Production of robots tells upon the environment

In a consumer economy, the problems of heavy metal contamination, and recycling are exacerbated (UNESCO and COMSET. 2015).

Another ethical issue got a special mention in the report (UNESCO and COMSET. 2015) and that is the relationship between Robotics, employment, and labour. It has been mentioned that Artificially intelligent agents like Robots are linked to an increase in global productivity. It has also been observed that an increase in production through the deployment of Artificially Intelligent Agents involves a drastic reduction in labour costs. Hence, the report has pointed out that Robotics has a worldwide impact on employment and the nature of work. It has been stated that it is presently problematic to provide statistics, but several employment opportunities around the world will be transformed or may disappear as a result of the proliferation of robots. When a specific kind of job is eliminated by new technology, the job that has become automatic becomes unnoticeable and loses social value. (UNESCO and COMSET. 2015).

Hence, in certain situations, Artificially Intelligent Agents are posing a serious threat to humanity. This threat may be perceived in different ways but human labour is one of the important areas where this threat is real. I can term the workers of Amazon's firm as 'labour', whose jobs are taken away by the robots. Likewise, I can also call the drivers as 'labour', whose jobs are at stake because of the growing popularity of automatic driverless cars.

Now, what is labour and who identifies as labour? The term labour can refer to both a noun (a person) and a verb (referring to the act). By labour, I imply the capacity to understand an issue and implement a solution to a problem. It combines the sentient's muscular strength, intellect, and creative thinking. To define labour, I will rely on what Karl Marx said. Marx calls labour a 'process'. According to him, in this 'process' both humans and nature participate. Not only that, human of their own volition begins, adjusts, and controls the relationship between the nature and the human beings. Using human beings' body and strengths an individual opposes to nature and in this way utilises nature in accordance with an individual's wants. (Marx, 1845).

Marx, however, identified the fundamental aspects of the labour process. These are

a) work (activity of a human being)

b) the theme of that work (subjects)

c) instruments.

According to Marx, in the labor process, human action, with the assistance of the tools of labour, causes a change in the substance that was planned from the start. According to him, the process is lost in the product. To Marx, the second is the use-value. The environment's material is modified to human's needs through a change in form. (Marx, K.1845)

A question may arise at this juncture: How to measure the wage of a labourer? Marx's labour theory of value is a key tenet of traditional Marxian economics, as demonstrated by Marx's masterpiece, *Capital* (1867). The theory states that the product's value is objectively measured by the 'socially necessary labour time' required to make any use-value under routine production conditions with the ordinary level of skill and density prevailing in a society. If conditions of production change, then socially necessary labour time changes.

'Labor-power' is identical to human's labour in nonconcrete way which is common in all forms of use-value producing concrete labour. Marx explained, the value of labour-power must be determined through the number of labour hours it takes society on average to feed, clothe, and house a worker (social reproduction) so that he or she has the capacity to work again. To put it in different words, the long-term wage workers receive can be resolute by the number of labour hours required to produce a worker. Assume it takes five hours of labour each day to feed, clothe, and protect a worker so that the worker is ready to work the next morning, if one labour hour is worth one rupee, the correct daily wage is five rupees. Thus Marx says, "Wages are that part of already existing commodities with which the capitalist buys a certain amount of productive labor-power...What, then, is the cost of production of labor-power? It is the cost required for the maintenance of the laborer as a laborer, and for his education and training as a laborer." (Marx, Karl. 1867 [1981]).

According to Marx, the social distinction between simple and complex labour is extremely important in the determination of the price of labour-power (wage). Simple average labour in a particular society at a particular time is a given. Complex labour is intensification of simple labour and is equivalent to multiples of simple average labour. So,

if a society is able to fix the minimum living wage of a worker dispensing the simplest labour, one can ascertain the wage against complex labour as a multiple of simple labour.

In my thesis, I will show that this notion of ‘labour’ comes into conflict with Artificially Intelligent Labour. I will show that one of the main reasons behind this conflict is that I am talking about a machine that claims to replace one of the basic species-specific features of human labour in some situations. According to Marx, one of the basic species-specific behaviour of human labour is adding ‘value’ to what he/she produces. Though there may be some critical differences in defining value, and some scholars may disagree with how Marx defined it, I will be following the Marxist explanation of 'value' in order to answer the research question. Many AI scientists are talking about a device a very special mechanical standard that claims to substitute human autonomy, intellect, creativity — a machine that devices problems and innovates solutions in some situations (Dewhurst, Martin and Willmott, Paul. 2014).

History tells us that human labour is being empowered with the aid of instruments. This is one thing. But in some situations, technology (AI) itself claims to replace human labour, that is a different issue. This confronts us with an ethical dilemma. Our research question stems from this dilemma. Hence my research question is: if we conceive Artificially Intelligent Agents as ‘agents’ and these ‘agents’ as ‘labour’, then can these ‘agents’ replace human labour in some situations? In this dissertation, I will try to answer this.

In the research question, I have mentioned that *in some situations* human labourers are facing threats from their Artificial counterparts. Let me explain that what I mean when I say the phrase 'in some situations' in the research question.

According to current research, artificial intelligence and robotics will largely replace human labour, primarily in the service, manufacturing, office, and administration sectors. Webster, and Ivanov (2020) have discussed in detail in which cases corporations from different segments of the economy have adopted/will adopt AIAs, because they argue, companies are on hunt of lesser costs, quicker manufacture time, a constant improvement in quality product and good supervision of supply chain processes, etc.

Initially, the producing sectors used industrial robots (Colestock, H. 2005). Currently, AIAs are widely used in different spheres of the economy and society. It is being used in areas like managing supply chain (Min, H. 2010), farming (Driessen, C. & Heutinck, L. F. M. 2015), autonomous vehicles (Maurer, M. Gerdes, J. C. Lenz, B. & Winner, H. (Eds.). 2016), warfare (Crootof, R. 2015), travel, tourism (Ivanov, S. & Webster, C. 2018), education (Ivanov, S. 2016.), journalism (Remus, D. & Levy, F. 2015) and additional services, to trading on the financial markets (Dunis, C. L. Middleton, P. W., Karathanasopolous, A., & Theofilatos, K. A. (Eds.). 2017), and implementing medical operations (Satwant, Kaur. 2012). Chatbots are being used by businesses to communicate with and maintain connections with their clients. (Hill, J. Ford. W. R. & Farreras, I. G. 2015). These examples demonstrate AIAs' pervasive integration into society, resulting in massive shifts in how people live, work, and conduct business (Makridakis, S. 2017).

In this context a question arises: will AIAs have curtailed more jobs than they have created? The study 'AI, Robotics, and the Future of Jobs' (Webster, C. & Ivanov, S. 2019) finds an answer to this question. According to the report, half of these specialists (48%) imagine imminent days when robots and digital agents would displace significant numbers of both

blue-collar and white-collar labours, while others are concerned that this may impact to yawning pay inequality, effectively rendering an enormous amount of people unemployed.

However, a large percentage of other experts (52%) presume that by 2025, innovation will not remove more employment than it creates. According to this group, by 2025, many works that are presently being performed by individuals would be captured by automatons or Artificially Intelligent Agents. They assume, however, that human fantasy, as it has done since the beginning of the Industrial Revolution, will spawn new employment opportunities and will show new ways to make a living. (Webster, C. & Ivanov, S. 2019)

However, researchers argue that in the near future industries will necessitate a skilled workforce. The World Economic Forum mentions some of the job skills that the workforce needs. The abilities listed appear to be mostly intellectual and affective in essence. According to the World Economic Forum, the key competencies include problem-solving, critical reasoning, and cooperation with others. According to the World Economic Forum top ten job skills needed are 1. Multifaceted Problem Resolving 2. Critical Reasoning 3. Originality 4. Man Management skill 5. Harmonizing with others 6. Emotional Intelligence 7. Judgment and Decision Making 8. Service Orientation 9. Negotiation 10. Cognitive Flexibility.

These important services identified by the World Economic Forum assume that humans have a reasonable benefit over robots. In my thesis, I will answer why human labour will remain *sui generis* in certain situations.

Now let us situate our research question in the broader context of ‘soft-morality and artificiality,’ which, actually, is the title of the dissertation. We have seen previously that in the report of UNESCO and in the EU discussion regarding the ethical claims of data,

information, Artificially Intelligent Agents, and their interaction with humans and society becomes prominent. This whole gamut comes under the periphery of Digital Ethics.

I will briefly discuss why traditional normative ethics is not sufficient or adequate to deal with actions performed by Artificially Intelligent Agents. Normative ethics in general put humans at the centre. It tries to explain the conflict of duties faced by humans as 'moral agents' in different situations. These ethical theories are either agent-oriented or action-oriented and essentially anthropocentric in nature. With the advent of digital technology, this notion of morality changes, as in the new situation humans alone cannot be considered as 'agents' and is not at the centre of moral actions. Moreover, different situations have emerged with the introduction of digital technology in which traditional normative theories fail to account for moral problems. According to Luciano Floridi, the AI boom has changed our viewpoints on values and priorities, good conduct, and the type of advancement that is not only viable but also socially recommended. Now the core problem of digital ethics is how to govern all of these. (Floridi, Luciano. 2018).

Floridi unfolds the nature of the ethical problems with examples of some questions: What is the next disturbance? What is the latest game-changing app? Will this be the year when virtual and augmented reality finally clash? Or will the internet of things, perhaps in conjunction with smart cities, represent the new frontier? Is the finish of television, as we know it, on the horizon? Will machine learning render healthcare unrecognizable, or should we instead focus on logistics and transportation automation? What will the new smart assistants in the home do besides tell us the information regarding climate and play our favourite song? How will military strategy evolve in response to cyber conflicts? (Floridi, Luciano. 2018).

Likewise, I can ask, whether the artificially intelligent agent as labour replaces human labour in some situations? Will Artificially Intelligent Agents replace the entire human race? Whether the autonomous driverless car is held responsible for the accident it meets with?

Hence we need different kinds of ethics, like Digital Ethics which will address these problems. Digital Ethics, as Floridi thinks, explores and examines ethical problems concerning data and information. It includes phenomena like the generation, recording, curating, processing, disseminating, sharing and usage of data. Moreover, an algorithm that incorporates Artificial Intelligence, Artificially Intelligence Agents, Machine Learning**,** and robots falls under the purview of digital ethics.

Further, accountable novelty, writing programmes, make some programmes hacked fall under its discussion. This is particularly needed "to formulate and support a morally good solution (e.g. good conduct or good values)." (Floridi, Luciano. 2018).

Additionally, we can divide digital Ethics into hard ethics and soft ethics. To form new regulations or to challenge the existing ones, we follow "what is morally right or wrong and what ought and ought not to be done". Hard ethics, for instance, acts as the prime mover in making or shaping the laws. (Floridi, Luciano. 2018).

Soft ethics, on the other hand, finds its root in hard ethics. It discusses what ought and ought not to be done in that of the existing regulations. In addition, it does not challenge the existing rules and regulations either. (Floridi, Luciano. 2018)

Furthermore, taking a cue from Kantian Ethics, Floridi holds that notion of feasibility is important in both hard and soft ethics. He, however, gives credit to Kant in formulating his

position. He maintains that both hard and soft ethics assume that ‘ought implies can’. (Floridi, Luciano. 2018).

Besides, Floridi has pointed out, soft ethics is ‘post-compliance’ ethics. According to him, in the cases of soft ethics follows the ‘ought implies may’ principle. At this juncture, one may ask, what is the difference between ‘ought implies can’ and ‘ought implies may’?

It should be noted that, in ethics, ‘ought implies can’ is a proposition which postulates that an agent has a moral obligation to perform certain actions only if it is possible for her to perform them. It is an ethical formula conceived by Immanuel Kant that claims ‘an actor if morally obliged to perform a certain action, must logically be able to perform it.’ (Floridi, Luciano. 2018).

If I put it differently, if a certain act is impossible for an actor to perform, the actor cannot have a moral obligation to do so. Kant holds that ‘ought implies can’ is considered as a minimal condition on the plausibility of any ethical theory. We can say after Kant that no such theory is justifiable if it implies that actors have duties to perform actions that they are unable to perform.

Furthermore, some critics opined that for Kant ‘ought implies can’ is the necessary and sufficient condition of morality. (Stern, R. 2004) That is to say, if any action is possible in a given situation, then the actor has a moral obligation to perform it.

Consequently, Floridi thinks that, in soft ethics, an actor can take the help of ‘opportunity strategy’. By ‘opportunity strategy’ Floridi means that it takes into account ‘social values of digital technology’. (Floridi, Luciano. 2018) This implies, in soft ethics,

socially acceptable or socially preferable new opportunities are given due importance. In my thesis generation and transfer of value occupies special importance. I will come into this later.

We know that, in English, 'may' is a verb that refers to the 'possibility' while 'can' is used to the 'ability to do something'. In other words, 'can' is used when someone can do something, or when you are allowed to do something. On the other hand, 'may' in this context, is used to discuss possibilities or happenings in the future. (Hornby, A, S. 2010.) In my case, however, the notion of possibility is important, as I am exploring new possibilities of digital ethics that considers Artificially Intelligent Agents as 'agents'.

We know that 'may' has diverse meanings in different situations. For example, in some situations 'may' refers to 'must'. But in my thesis, I want to emphasize its meaning as 'possibility'.

In this context, I must clarify some ideas about what Floridi means by 'post-compliance' ethics.

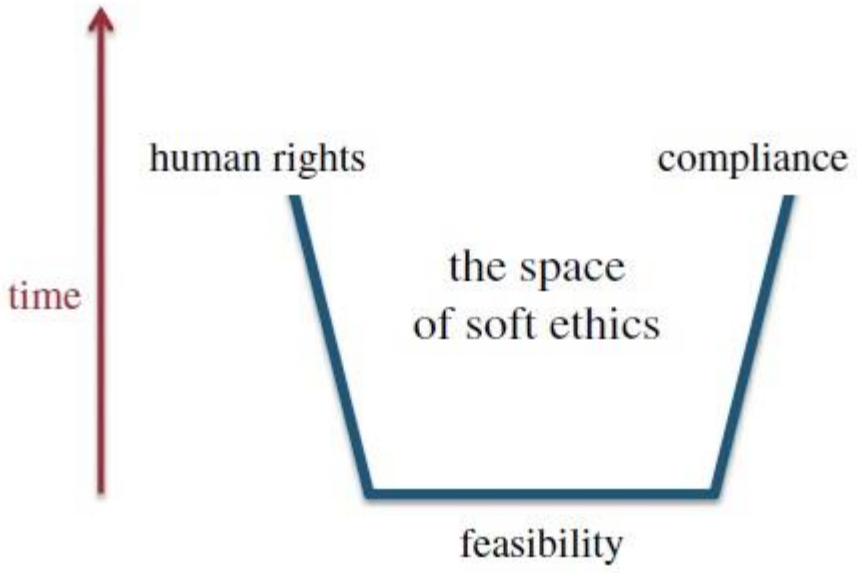
Floridi holds that hard ethics help formulate legislation and the Universal Declaration of Human Rights. Soft ethics, on the other hand, comply with legislation first. Moreover, he says that soft ethics first complies with the General Data Protection Regulation of the European Union. After that, it creates its own space. That's why Floridi uses the term 'post-compliance ethics'.

In this context, one should keep in mind, that currently, India does not have any personal data protection act. It was 11 December 2019 when 'The Personal Data Protection Bill 2019 or PDP Bill 2019' was placed in the Indian Parliament by the Ministry of Electronics

and Information Technology. As per reports published on 7 December 2021, a joint committee of parliament is examining the draft bill. (Roy, Priyanka. 2021) So in India, it is not possible to apply post-compliance ethics (soft ethics).

On the contrary, the EU accepts the application of the Universal Declaration of Human Rights, the European Convention on Human Rights and the Charter of Fundamental Rights of the European Union which evolved using the framework of hard ethics. Complying with this soft ethics is being developed.

Furthermore, Floridi, holds that "the space of soft ethics is both partially bounded, and yet unlimited." (Floridi, Luciano. 2018.) He uses a figure to illustrate the space of soft ethics (Roy, Priyanka. 2021).



Floridi uses this diagram to make us understand the scope of soft ethics (Floridi, Luciano. 2018.).

As you can see, the lower side of the figure represents a feasibility base. With time this can be expanded. It shows that one can do many things with the help of technological innovation. On the other hand, the other two sides represent legal compliance and human rights.

Floridi (2018) tells us, that the open-top side represents the space for Soft ethics. According to him, it helps to shape and guide the ethical development of our mature information societies.

One may find that post-compliance soft ethics follows a similar normative foundation as hard ethics. However, it accomplishes this by seeing what ought and ought not to be done in addition to or instead of existing regulations.

For example, Soft-ethics helps in risk management. The example of the Facebook–Cambridge Analytica data scandal shows that soft ethics help apprehend errors. (Confessore, Nicholas. 2018.)

I derive the concept of soft-morality from soft ethics. After Floridi I can say that soft-morality is 'post-compliance morality'. Nonetheless, it complies with the available notion of morality first. After that, it creates its own space.

Stanford Encyclopedia of Philosophy says that 'morality' is applied in two separate wide-ranging senses: a descriptive sense and a normative sense. To quote, "Morality can be used either descriptively to pertain to specific codes of conduct proposed by a group or group of people (a religion) or recognized by a person for her behaviour, or normatively refer to a code

of conduct that, under certain conditions, all rational individuals would follow." (Gert, Bernard and Joshua, Gert. 2020)

As I have constructed, soft-morality complies with the available notion of morality first. That is, soft-morality complies with both the notion of morality. It accepts that there are some codes of conduct offered by a society or given a specified condition that would be put forward by people which may be followed by both human and artificial agents. After complying with this, soft-morality creates its own space.

However, the need to develop soft-ethics and soft-morality emerges from society. Our lived experience shows that the domain of digital (both online and offline) and non-digital, indeed, are getting blurred day by day. Can we not say that in these days our very existence involves both digital (online and offline) and non-digital interactions? Undoubtedly, the pandemic situation unfolds this emphatically before us. We are experiencing that our society too, is digital (online as well as offline) and non-digital. It is a fact that we cannot sharply distinguish, when our offline life ends and when our online life begins.

Floridi (2018) terms this 'onlife'. Furthermore, according to him, we live in the 'infosphere', which is analogue and digital, offline and online. In the infosphere, we interact with humans as well as with artificial agents. Since we include artificial agents in our ethical framework, we arrive at a different situation.

One of the differences is, earlier we dealt with machines or instruments, which were not Artificially Intelligent. We can take the example of the film *Modern Times*. It is an American silent comedy movie. It was inscribed and directed by Charlie Chaplin (1926). The machines or instruments depicted in the films were controlled by humans and not Artificially Intelligent Agents. We saw what happens when these machines malfunction.

On the contrary, these days, we come across machines that can learn by themselves. These machines can evolve themselves. Even some intelligent machines control other machines (i.e computer controls a bot). This happens within the scope of the 'infosphere' and gives rise to a different situation where a 'conflict of duties' arises involving artificial agents and humans. (Frankena, William, K. 1973) This paves the way for exploring the scope of 'ontocentric[4]' ethics which focuses on mindless morality and sees artificially intelligent machines as 'agents'. Luciano Floridi, a pioneer of information ethics, sees it as the 'fourth revolution,' following the Copernican Revolution, Darwinism, and Freudianism. Human beings have been pushed from the centre to the periphery in each previous 'revolution.' (Floridi, Luciano. 2014)

In this context I want to mention that, I begin our introduction with a passage from *The Hindu,* where we find that a hundred years ago Czech writer Karl Capek conceived Robots in his debut science-fiction play *Rossum's Universal Robots* (Capek, Karel. Paul Selver and Nigel Playfair, 1923). In his play, he conceives a race of artificial agents who would eventually overpower their creators.

Capek, however, tells us that there is a factory that produces lots of robots. These Robots, according to this play, would change the world. They would make labour cheap and which eventually would eliminate all work and poverty someday.

What Capek had fancied a hundred years ago in his work, however, proved to be partially true in today's world. Though human labourers have not been replaced or perished by robots entirely, in some situations human labourers are facing threats from their artificially

[4] Floridi defines information ethics as: "Information Ethics is an ontocentric, patient-oriented, ecological macroethics." According to information ethics, each entity does have dignity as an expression of being, which is defined by its mode of existence and essence. This, Floridi term, as ontological equality principle. It holds different form of reality "has a minimal, initial, equal right to exist and develop in a way suitable to its nature simply by being what it is." Floridi proposes that ontocentrism replace biocentrism. According to ontocentrism, there is something more fundamental than life, namely being, and something more foundational than suffering, namely entropy. We have discussed this in detail in the second chapter.

intelligent counterparts. For example, a recent report shows that Amazon has deployed more than ten thousand robots in its warehouses. (Distefano, 2021). These robots efficiently move things around. In Amazon's warehouses, humans pick and pack belongings; while robots transfer orders around the big storerooms.

Another example would be the emergence of automated automobiles which give rise to many ethical issues (Kröger F. 2016).

This, indeed, is one scenario. On the other hand, we have many examples of robot-friend like Klara from the science fiction of Nobel laureate British writer Kazuo Ishiguro's *Klara and the Sun.* (Ishiguro, 2021). We get to know that Klara is a narrator and she is an artificial agent. Her job, however, is to befriend and care for a sick young teenager, Josie. In reality, we find many artificial agents which are used for caregiving activities. I will not be discussing this in my thesis.

Moreover, we have seen that experts have manufactured the living robots that breed using a completely new method of reproduction. (Kriegman et al. 2021). Scientists claim that they have not seen this phenomenon before. In this thesis, however, I am not discussing whether the biological life-world would be replaced by robots either.

However, we have seen that the concept of a robot is evolving. Its functionality is also changing. We are experiencing that technological advancements are going at breakneck speed. Moreover, onlife interaction, machine-learning, cloud computing, new forms of smart agency, organic and inorganic robot, artificial life etc. are constantly evolving and in return shaping our society and resulting a kind of technological determinism.

I have experienced that in a specific situation like that of Amazon's firm or usage of the automatic driverless cars, artificially intelligent machines have replaced/are replacing human labour and creating conflict with human world. Opponents may argue that deployments of Arificially Intelligent Agents may involve in decreasing the number of human labour in some cases, but it is also true that more jobs will be created in Robotics industries. I want to clear from the onset that there is a difference between what Artificially Intelligent Agents cannot do and what they will never be able to do. For the present purpose, I just mention this point. However, in my thesis, I will address this in detail.

One of the reasons behind this concern may stem from the pace of evolution of that artificial agent. Both Floridi and Stephen Hawking raised this concern in their work too.

Floridi thinks that the speed in which they evolve is amazing. To him this is the reason of concern. (Floridi, 2018). He has used the word 'apprehension'. By this word, he might have meant this 'replacement' thesis that I am talking about.

Similarly, a few years back, Stephen Hawking also echoed the same. He said that we cannot ignore artificial intelligence. But there is a threat in it. So when we will handle this, we will have to be extra cautious of its dangers. He has a fear that Artificially Intelligent Agents may substitute human beings. Development of full Artificial Intelligent, he fears, would destroy humans. (Cellan-Jones, Rory. 2014).

It has been mentioned earlier that, I am not discussing whether the entire human race would be replaced by Artificially Intelligent Agents as Hawking had apprehended. Instead what I want to say is the possibility that artificially intelligent machines would be able to successfully replace human labour in some situations or not. It is this possibility that makes me ask whether this can actually happen. Opponents may say that the industry might flourish

and the scope of employment in these industries might increase because of the advent of AI agents. However, this falls outside the scope of my dissertation. I want to differentiate between what AI agents still cannot do and what they will never be able to do. According to my understanding, in order to qualify as natural agents and in order to replace human labourers AI agents need to generate value and at the same time need to participate in the lifeworld where empathy, understanding others, cooperation with others plays a predominant role.

To address my research question the thesis has been divided into six chapters. First, is the introduction.

The name of the second chapter is 'Socio-Technical Issues in AI: Can an Artificial Agent differentiate between doing and allowing?' In this chapter, I have considered certain examples and some thought experiments which will persuasively argue for the moral dilemma from which our research question has emerged.

In this chapter, I have observed that breaches of ethical conduct are occurring when Artificially Intelligent Systems are deployed. To elucidate my position, I have cited some examples. In this context, I have followed the 'case-based approach'. I have discussed some thought experiments as well which depicted this dilemma.

Moreover, from these thought experiments and examples, I have found a shift in the way in which we think about digital ethics. In this century; we are not only frightened with physical evil but mental evil too. And perhaps this has forced me to think of artificially intelligent machines as an 'agent' and these 'agents' as 'labourers'[5] which can be conceived to have certain ethical claims. Since ethics emerges from society and primarily helps us to

[5] In the whole thesis I have used artificial agents and artificial labourers interchangeably.

deal with certain situations where conflict of duties arise, the question however remains, is it possible to compute the unknown situation beforehand, so that the 'ethical machine' can take a proper decision?

To understand the problem, I have classified the chapter into two sections. In the first section, I have asked, what are the breaches of conduct that are occurring in the context of deployment of Artificial Intelligent systems? In this context, I have cited some newspaper reports that pointed to the Socio-Technical Problems in AI.

In the second section, I tried to find an answer to the question: Why do some ethicists demand ethics for artefacts? To understand this, I have discussed the Trolley Problem in AI.

The third chapter consists of a Literature survey. In the previous chapter, I have discussed that there is a shift in the way in which we think about machine ethics. It may seem that previously matters regarding ethical development of machines, usage of technology and a kind of technological paranoia stemming primarily from the usage and development of machines were at the core of human thinking about these issues. Technological development has been perceived as an ominous sign for humans in these situations. We have discussed this in the previous chapter.

As time flows, however, the need for the ethical dimension of machines themselves preoccupies ethical and technological theorists. We have traced the shift in the human thought process in a paper 'Contextualizing Ethics in the Realm of Robotics'. (Bandyopadhyay, R. and Guha, M. 2008) jointly written by me and one of my supervisors Dr Maushumi Guha, in the year 2008. There is literature aplenty, which also reflects this tension/s. In both cases, human beings engaged their reasoning to interpret a situation where sentient beings interact

with machines. But what if, a group of robots interact with another group of robots along with humans? We can easily imagine a situation where interaction between man and machines takes place and simultaneously interaction among machines also occurs. Consider the accident that took place in Mountain View, California on 26th September 2016. (Curtis, S. 2016) A driverless car had hit with a van. The *Mirror* in its report describes this accident as the worst involving an autonomous vehicle. What if in such an accident two driverless cars had collided? What if, in such an accident, a person inside one of the vehicles would have died? Such a situation would involve human-machine and machine-machine interaction.

This prompts us to ask: how to implement the 'anthropocentric harm principle' in a situation where robots interact with robots and humans? By 'anthropocentric harm principle' I mean the ethical doctrines which put the interest of humans at the centre and protect the humans from any kind of harm inflicted on them. Do we need to develop different kinds of ethical parameters to account for the emerging situation (robot-robot and robot-human interaction)? Given the current situation and the shift in ethical analysis following from them, is there any more any distinction between a person who performs a moral act (moral actor/agent) and the one at the receiving end of such an act (patient)? Let us consider this as the central question of this chapter. To delve deep into the question, I have clarified certain ideas regarding machine ethics in general.

We have seen how the notion of 'ethics', as well as 'agency', have evolved from time to time. The journey of ethics from 'anthropocentrism' to 'ontocentrism' is a long one. There are certain landmarks in this journey and I have tried to critically revisit these milestones to

understand the next turning point. 'Ontocentric' Information Ethics formulated by Luciano Floridi plays an important part in my thesis. I have discussed this theory in this chapter.

This confronts us with a question: could ontocentreism be the end of the road in ethical discourse? This paves the way for our next chapter. In this chapter I have asked a question: Can we say that if we accept the logic of ontocentrism, then artificial agents would be qualified for gaining human-like agency and thereby become laboures? I have illustrated this with the example of Actor-Network Theory (ANT) proposed by Bruno Latour and observe how the notion of ANT has subsequently been expanded to incorporate AI agents. The incorporation of AI in the actor-network raises some legitimate concerns over its ethical implications. Using Marxist ethics as a case study I have illustrated the problem. In the realm of the application of artificial agents, we have already encountered some piquant problems and I have discussed those specific cases towards the close of the chapter. At the end of this chapter, I have discussed that if we rely too much on ontocentrism, i.e treating 'nonhuman' like 'humans' and ascribing human-like agency to them and thereby replace human labourers, then it may lead to some ethical problems. One problem is regarding the generation of 'value'. Since an Artificially Intelligent Agent is a human creation, it is a congealed form of dead labour. Following Marx, I can say that dead labour, cannot generate new value.

In this dissertation, I have used the concept of value after Karl Marx. In *Capital Volume I* Marx discusses "exchange-value and the commodity", which are the very "foundations of the capitalist system" (Marx. 1996). According to him, value takes the form of exchange-value, which masks its origins in labour. He writes, "value of a commodity is

relative, and not to be settled without considering one commodity in its relations to all other." (Marx. 1996.)

Think of a hypothetical situation. In a firm, the workers face a complex situation. To get rid of this situation, they have to solve some difficult mathematical problems. The problems are so complex that they will need considerable time to solve. However, an AIA can solve this problem in a trice. The opponent may ask, is it not value generation? Do AIAs contribute to generating epistemic value? To find an answer, I will discourse on what Marx would have understood as 'value'.

Marx considers value as 'realized, fixed, crystallized social labour'. (Marx. 1996.) He thinks, to calculate the exchangeable value of a commodity, one must add the amount of labour previously expended in the commodity's raw material, as well as the labour expended on the implements, tools, machinery, and buildings that aid such labour (Marx. 1996.).

In my case when an AIA calculates some difficult sum, we need to look at the making process of that particular AIA which involves a huge amount of labourers to make— from its mechanical structure to developing software that calculates in a trice. So when an AIA calculates a difficult sum, it actually transfers the value. The 'dead labour' within the AIA helps in calculating the difficult sum so quickly. It may seem that it is generating 'epistemic value' but Marx would have said that it is the 'crystallization of the quantity of labourers previously realised' in the AIA that generates value and machines like AIAs only transfer it.

> AIAs, in my opinion, are part of the production methods that offer themselves anew as a constituent part of the value of the commodity. In other words, the AIAs value is retained by being transmitted to the creation. (Marx.1996).

> AIAs will not transmit much value to the new commodity than it lost in the labour-process by the obliteration of its use-value (Marx.1996). Throughout the labour process, the AIAs will lose value in the character of their old use-value. How much dropping of value (that is transfer of value) can they endure in this process? That would be determined by 'the amount of the original value congealed in them, which is the 'socially necessary labour time' expended in their making. According to Marx, the means of production (read AIAs) just cannot pour further value into the item than they already have, regardless of the manner in which they aid. (Marx.1996).

It can only distribute it to the commodities it creates. I have cited some examples that show that in some cases excessive dependence on AI tells upon productivity. Why machine cannot generate value have been explained by taking a cue from Marxist ethics.

The name of the fifth chapter is, 'Can AI agents as labourers replace human labourers in some situations?' At the very outset, I have treated this as a question stemming from the lifeworld. I have conceived the notion of the lifeworld in Husserlian sense[6]. We know, there is a close relationship between ethics and the lifeworld as moral dilemma has always emerged from the lived experience in the lifeworld. Ethical questions (like, Does AI agents have agency? or,

[6] By lifeworld I mean the phenomenological world of intersubjective experience. Individual, social, perceptual, and practical experiences are all part of the lifeworld. Through experience, humans learn whatever they could learn. Furthermore, human experience tells her that reality is too diverse to be explained algorithmically. To establish my position, I have designed some thought experiments. I have shown that understanding others through mental simulation occurs by participating in the lifeworld. In the conclusion, I have discussed this in detail.

can AI agents as labour replace human labour?) are questions that essentially originate from the everydayness of the interaction between man and machine. I have given a possible explanation of this taking a cue from Marxist ethics. In this chapter, I have proposed that until and unless AI acts in tandem with lifeworld, it will be difficult for them to 'replace' humans in some situations.

At this juncture, the question arises; is learning algorithm sufficient in acquiring practical wisdom? Wouldn't people ought to practice anticipated, affective, and social techniques that allow them to apply their overall conception of happiness in some respect that are appropriate for every situation? I have asked, can an Artificially Intelligent Agent gain 'a predictive hold' over its behaviour so that it could attribute emotions, beliefs, desires and thoughts to one another? Adam Morton (2003) claims that we comprehend others because we can collaborate. By this, he means that one can anticipate, describe, and comprehend action in part because one can participate in a collaborative activity. This has prompted us to ask, can artificial agents involve in these cooperative activities that take place in the lifeworld and would eventually replace humans in some situations?[7] These are the lines of thought we follow in this chapter. Thus the chapter comprises of two sections:

1. Are Artificial Agents as labourers going to replace humans in some situations?
2. Are they going to supplement human capacities in some respect?

In the conclusion I have addressed our research question---Can Artificial Agents/labour replace humans in some situations? I have, however, discussed that Artificial Agents cannot generate value. They only transfer it. On the other hand, humans

[7] In terms of technology cooetrative robots exist. The question is not technology can build such robots or not, the question is whether such technology can replicate the essence of human cooperation in the lifeworld.

create value. This is one reason Artificial Agents cannot replace humans in some situations.

Moreover, another reason may be, their failure to participate in the lifeworld just like humans. I have visualized some thought experiments and tried to show that until and unless AI acts in tandem with lifeworld, and learns to distinguish between acts of mental simulations like ‘pretensions’ and ‘intentions’, it will be hard to replace humans by them in some situations.

Nevertheless, the question which confronts us at this juncture is, can nuances of practical wisdom be acquired by a learning algorithm? Don’t we need to acquire through practice those deliberative, emotional, and social skills that enable us to put the general understanding of wellbeing into practice in ways that are suitable to each occasion?

From one point of view, we have gathered our ‘knowledge’ through evolution. This has taught us to differentiate between ‘pretension’ and ‘intention’. This is a question of lifeworld that emerged from a folk context. We have raised a question: can a robot gain 'a predictive hold' over its behaviour so that it might ascribe thoughts, desire, belief and emotions to others?

Hence, I conclude that it will not be a relationship of replacement; they are going to be there along with humans. Hence, it will be a relationship of ‘conjunction’. Borrowing the term from bi-valued logic, I proposed that the relationship would be of ‘conjunction’ (leading perhaps to complementariness) (Shramko, et al. 2020.) where if both the operands satisfy the conditions (true) then the entire system will be satisfiable or satisfactory (true) and if any one of the operands malfunctions, the entire edifice will

collapse like a house of cards. Without the co-existence of humans and Artificially Intelligent Agents value generation and value transfer will not be possible.

In the introduction, I have proposed that this question stems from a certain ethical framework i.e. soft-morality. Several issues fall within the scope of soft-morality. The replacement of human labour by AI agents is one such concern. Since I propose that the connection between sentient and artificially intelligent agents is of 'interdependence', the ethical concern that emerges from here falls under the periphery of soft-morality. Hence, the title of the thesis is a broad one—Soft-morality and Artificiality.

Chapter 2

Socio-Technical Issues in AI: Can an Artificial Agent differentiate between doing and allowing?

Breaches of ethical conduct are occurring in the context of the deployment of an Artificial Intelligent System. To elucidate my position, I want to cite some examples. Moreover, I will follow the case-based approach in this chapter. Additionally, I want to show why some ethicists want to build Artificially Intelligent ethical machines that claim to replace human labour in some situations. The question, however, remains, is it at all possible to build ethical artificial agents? To reach a conclusion, I will cite some examples from real life and see how ethical conduct is breached in certain situations.

This, however, will pave the way for our next section, where I will be discussing some thought experiments that will help us better understand the problem.

From these examples and thought experiments, we will be able to find a paradigm shift in the thinking process regarding machine ethics. *Frankenstein*'s ghost, however, is not the only concern, to add to it, we are often being challenged with our 'intelligence' itself! Some of these intelligent machines challenge their creator's intelligence and pose a threat to the labour market. I would like to mention that; we conceive a car driver or trolley driver as a 'labour'.

What do these examples from real life as well as the thought experiments show? They show that Artificially Intelligent machines claim to have 'agency'[8]. That means, they can

[8] Regarding the notion of agency Anscombe and Davidson may disagree, but both of them hold that 'action is to be explained in terms of the intentionality of intentional action.' Furthermore, we will call an agent a moral agent only when the agent can be held responsible for its actions.

perform certain actions. They can take the decision of their own and their actions are intentional. (Schlosser, Markus. 2019)

Thus, I will address two central questions in this chapter:

i) What are the breaches of conduct that are occurring in the context of the deployment of Artificial Intelligent systems? To find an answer, I will cite some newspaper reports that point to the Socio-Technical Problems in AI.

ii) Why do some ethicists demand ethics for artefacts? To understand this, I will discuss the Trolley Problem in AI.

I

Socio-technical Problems in AI

Today, there are news galore regarding AI in the media. It is true that we can track societal changes by reading the news in newspapers or on news websites. However, there is news that can have good and bad effects on society. It is observed that a lot of new information published about recent scientific breakthroughs over the past few years. It has also been noticed that Artificially intelligent agents are being applied in different spheres. These include deep ocean exploration, aerospace engineering, and health sciences, to name a few. Moreover, one could easily imagine how the use of drones has changed warfare. However, it has also been observed that how autonomous cars functions. Furthermore, news pours in about how the software agents like bots control financial trade or how deep learning in medical science helps in a major operation. (Adams. 2017)

This demonstrate the development of mankind from the time of the industrial revolution or the scientific invention portrayed in Charlie Chaplin's movie *Modern Times*. Without a doubt, machine learning and Big Data help Artificial Intelligence reach a new height. (Mani, Chithrai. 2020) We have found a new subject called Mechatronics (Winterstein, Dave, 2022). This subject focuses largely on the development of Artificially Intelligent agents. These 'autonomous' agents exhibit a great grade of independence. These artefacts allow the sentient to communicate more closely with them. We have also got computer-brain interface machines. So many things from science fiction become possible. (Li G. Zhang D. 2017).

Furthermore, in some cases, smart technology can be used to replace or assist humans. Consider the use of 'Google assistance' in smartphones. It can assist you in looking on the internet for anything. Furthermore, technology can converse with customers via online call centres. In some cases, a robot hand can outperform a human at repetitive tasks. Smart systems can buy and sell stock in a flash. It can also direct your vehicle to park at a safer place. These Artificially Intelligent machines are no longer programmed in a linear fashion. According to a report published by the European Commission, Google Brain develops Artificial Agents that supposedly build Artificially Intelligent agents better and faster than the sentient. (Li G. Zhang. 2017.). AlphaZero can learn chess rules from being known as a tabula rasa to a world champion level in a few hours (Aguayo, Carlos. 2020). These machines can 'teach' themselves thanks to deep learning and generative adversarial network approaches (Brownlee, Jason. 2019). As a result, their actions are frequently unpredictable, and they remain unintelligible (Silver, David. et al. 2018). According to experts, there may be a gap between the initial algorithm and the final result. The report shows, their effectiveness is rooted in the data that was utilized throughout the educational process and might not be found later. As a result, preconceptions

and mistakes made in the past get ingrained in the mechanism. (European group on ethics in science and new technologies. 2018.)

If we go through the newspapers or news portals, we can find many examples of AI failures as well. From the news published in different media, we can find that these autonomous intelligent machines themselves have ushered in increasingly complex questions of growing concern. At the beginning of this chapter, I have stated that I will follow a case-based approach, so we will cite certain examples published in some of the reputed newspaper that confronts us with trolley-like situations. The question stemming primarily from this discussion is: In such complex socio-technical systems where is the moral agency located and who is going to be responsible for any untoward outcomes?

Legislators, law scholars, and producers in Europe are debating whether such devices or living beings should endure final the brunt of their deeds. A European Parliament's draft report was published in 2017. According to the reports, self-developing artificially intelligent agents may be given the status of 'electronic personalities'. It reads, "Creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good or any damage they may cause, and possibly applying electronic personality to the cases where robots make autonomous decisions or otherwise interact with third parties independently…"[9] According to the report, such a status could enable robots to be

[9] The European Parliament Committee on Legal Affairs voted 17 to 2 on January 12, 2017 to accept a proposed report published in May 2016 by Luxembourg MEP Mady Delvaux with suggestions to the Commission on Civil Law Rules on artificially intelligent agents. The report advocated for the institution of electronic person status for robots, as well as the inclusion of Asimov's Laws into European robotics law. It is now proposed for a vote by the entire European Parliament.

individually held accountable if they go wrong and start biting folks or vandalizing belongings.

Despite the fact that the European Commission's latest overview of a machine intelligence plan does not grant personhood for artificially intelligent agents, still, this draft report speaks a lot about the shift in the human thought process. I have mentioned earlier that there is a shift in the human thought process.

However, the European Parliament's action compelled numerous specialists to issue a public letter. In this letter, it has been urged to the Commission to disregard the Parliament's plans and discard the proposal for giving AI the status of 'electronic personality'. (Nathalie Nevejans, 2018.) The letter says it would be inappropriate, ideologically, to publicize any lawful standing. It is nonsensical and non-pragmatic. According to this letter, the Natural Person model cannot be used to derive a lawful standing for an artificially intelligent agent. The letter argued, if it does, then the Artificially Intelligent Agents civil liberties like the right to equality, the right to integrity, the right to compensation (wage), and the right to citizenship would be secured. Moreover, the letter tells that it will directly confront human rights. This, they argue, is in violation of the EU's Charter of Fundamental Rights and the Convention for the Protection of Human Rights and Fundamental Freedoms.

Again, according to the experts, the lawful grade of an Artificially Intelligent Agent cannot be derived from the Legal Entity model. Because it presupposes the presence of a sentient behind to portray and guide it, which an Artificially Intelligent agent does not have. However, the letter acknowledges the European Union's identification of a problem. It appeals to the its member to establish a framework for implementable advancement and build dependable Artificially Intelligent Agent in order to spur even larger advantages for European

citizens and the European Union's trading bloc. As a result, the Commission has highlighted its upcoming plan for dealing with the issues related to artificial intelligence. The phrase ‘electronic personality,’ as used earlier, is not mentioned in this report.

In the statement on ‘AI, Robotics and ‘Autonomous’ system’ EC, an institution of the EU, made a remark on AI, robotics and autonomous system. It discusses in detail the ‘autonomy’, as this is the keyword in the debate. According to the statement, the term ‘autonomy’ emerges from Philosophy and narrates sentient beings’ ability to constitute laws for themselves, to devise, imagine, discover social rules, and legislation to abide by. It includes freedom. That is, to define one’s personal benchmarks, select targets and achievements of one's own life. The thought functions that aid and encourage this are the ones that are most tightly linked to people's self-respect, and agency. They usually include elements of introspection, self-consciousness, and self-creation based on reason and values. As a result, autonomy in the associated ethical context must be assigned to sentient beings. Trying to apply the term ‘autonomy’ to a simple artefact, even if it is an extremely sophisticated, complex dynamic, and even intelligent system, is somewhat misleading. Because no intelligent agent or system, no matter how advanced and complex, could be termed ‘autonomous’ in the authentic ethical sense, they cannot be bestowed with the ethical standing and inherent dignity of the sentient. In current controversies about Lethal Autonomous Weapons Systems (LAWS) and Autonomous Vehicles (AVs), there appears to be broad agreement that Meaningful Human Control (MHC) is required for moral responsibility. This implies that living beings, not Artificially Intelligent Agents should finally retain control and thus bear moral responsibility. (Committee on Legal Affairs. 2017). Here lies the debate within the European

Union itself as to whether these autonomous systems could be an agent like that of human beings.

IEEE, the largest international technical professional organization devoted to the advancement of technology for the benefit of humanity, recently released the first version of a report (K. Shahriari and M. Shahriari. 2017). It inspires scientists and engineers to take priority ethical considerations when developing intelligent and autonomous systems. This prompts us to think about the need of building an ethical machine.

I will illustrate my position with two specific examples. The first situation is such that self-driving car slays a wayfarer.

The second is two bots started chatting with each other which its programmer could not understand.

A wayfarer was killed by a self-driving car

A self-driving SUV hit and slays a female pedestrian in Tempe, Arizona. It's the first-ever recognized pedestrian killing on a city street caused by an automated car. It has been reported that the vehicle was in self-directed mode. Though there was a human safety driver present there. (Wakabayashi, Daisuke. March 19, 2018.)

Later it was exposed, one of its software did not function properly after the car's sensors noticed the person. As per a report, at the time of the accident Uber's self-sufficient (autonomous) mode disabled the manufacturer's (Volvo) automatic emergency braking system.

After this incident, Uber postponed self-driving testing in North America. It has been reported that companies clogged their self-driving road tests in the US. Eight months after the accident,

however, Uber declared its intention to restart self-driving safety checks in Pittsburgh. What the firm would do with its self-driving project is not clear.

The question, however, remains, what kind of ethics do we incorporate or programme beforehand in dealing with such a situation? What could be the ethical principles that will be followed by the 'ethical machine' in the aftermath of the accident? And where is morality located?

Two bots chatting to each other

The second example is from Facebook's AI research lab. In the year 2017, we learned that researchers at the Facebook AI research lab had to close down 2 Artificially Intelligent bots after it was found that they were communicating in a strange language that only they could 'comprehend'.

This strange incident came to light after Facebook confronted the chatbots to negotiate a trade with themselves. The bots were instructed to exchange books, hat, and balls, which all had different values. According to reports, when the chatbots decided the English language was really not nice enough for them, the experiment quickly got out of hand and evolved peculiar attributes. The individuals who were allotted to look after them had no idea what they were conveying! (Griffin, Andrew. 31 July, 2017.)

An even more sobering analysis is revealed in a study conducted by Facebook's Artificial Intelligence Research division. According to the article, the bots managed to learn to negotiate in very human-like ways. Bots could presume to be fascinated by a particular thing in order to claim afterwards that abandoning it was a significant hardship. (Griffin, Andrew. 31 July, 2017).

In his book, Nick Bostrom (2014) discusses the 'Intelligent Explosion', an occurrence that will happen when devices much smarter than humans start developing devices of their own, creating a vicious cycle. Superintelligence, according to Bostrom, is an 'existential threat', a power that can be difficult to combat. In his book, he focuses on how we can stay alive in our unavoidable encounters with it. Deep learning as well as Machine Learning [the ground-breaking 'neural' algorithms which accurately reflect a person's brain function] advanced much faster than expected in past years. That is undoubtedly a significant cause why this has recently become such a hot topic. Humans could foresee phenomena going ahead in the technological sphere and remain anxious about what will happen after that, he says (Adams, Tim. 2016). But when the data will be manipulated by the machines, then this confronts us with an 'ethical problem', as we have seen earlier.

The problems persist not only in the two domains that have been discussed earlier. In the media, the problem has its manifestations. As an employee of print and later digital media, the writer of this dissertation can perceive that AI can wreak havoc. It could deteriorate the trouble caused by fake news. It would escalate the animosity and bigotry that social media bots are presently capable of spewing. Additionally, it might overwhelm you with emails, making this difficult to differentiate between genuine and automated emails. Experts have long expressed concern about the unintended social repercussions of widespread artificial intelligence. Elon Musk has long been admonishing us about how robotics and AI would eventually rule the planet. In the past, he has referred to AI as our 'greatest existential threat' and described its development as 'summoning the demon'. He expresses concern that AI could pose a 'fundamental risk to the continuation of human civilisation' (Gibbs, Samuel. 2014). According to Ray Kurzweil, smart artefacts will be able to outsmart people by the year 2029.

(cited in Cadwalladr, Carole 2014). According to Stephen Hawking, once people develop complete AI, it will start taking off by itself and remake its progeny at an accelerating rate. (Cellan-Jones, Rory. 2014). The media is full of alarmist opinions about the scary potential of general AI. As a result, these post-apocalyptic predictions were coupled with requests for more moral innovation in Artificial Intelligence. Some AI specialists claim that we can educate our future robot rulers to distinguish between good work and wrongdoing. They urge to build AI like a 'Good Samaritan' which would perform morally and assist people in need. However, it is more difficult to teach robots the concept of morality since humans are unable to express morality objectively in terms of quantifiable criteria that are simple for a computer to understand. Even the idea that humans possess a solid moral philosophy on which we can all agree is debatable.

When a conflict of duties arises, humans approach the issue from different perspectives. Someone may address the problem based on their best guess rather than extensive cost-management calculations. This is evident from the fat-man scenario of the trolley problem. In contrast, devices require specific and aim performance measures that can be assessed and optimized.

We can illustrate this with an instance. An Artificial intelligence-based chess player could indeed thrive in gameplay with straightforward guidelines and limits by repeatedly playing the game and acquiring knowledge to maximise the scoring rate. Alphabet's DeepMind had defeated the greatest professional performers after experimenting with profound reinforcement learning on Atari video games. (Garisto, Dan. 2019). We can remember that Deep Blue, IBM's chess computer, is regarded as the pioneer machine in defeating a ruling world titleholder in a six-game match.

The examples that we have cited above confront us with a question— how to build 'ethical machines' when we could not have the same understanding of ethics?

Now we will move on to the next section.

II

Trolley Problem in AI

In the previous section, it has been observed that the Artificial Intelligence system can decide on the fields like the stock market, medical field, manufacturing sector, and so on. In most cases, Artificially Intelligence Agent's algorithm is monitored by some persons who take the final decision.

However, imagine a situation where the final call would be taken by an Artificially Intelligent Agent. To make our imagination complex, let us imagine further that the situation is such that the question of life and death for humans is involved. In deontological ethics, the trolley problem is one of history's most famous thought experiments. We can alter the original thought experiment a bit to replace human pilots to incorporate Artificially Intelligent agents. Other basic features of the thought experiments remain the same.

Let's begin this section with the Trolley Problem in AI and its implications. We will discuss this after David Edmond's (2014) book.

In a nutshell story of the trolley problem is that an out-of-control trolley is putting the lives of several folks in danger. If the pilot of the trolley does nothing, then that would have killed those persons. But if he wants to save those people and divert the trolley in another

direction, then it might well lead to the death of others. For nearly three decades, there are many dimensions added to the main problem and these make the problem more intricate. Edmonds, however, begins the book with a historical example. He transports us to the close of the Second World War.

Edmonds commences the book by saying that during the fag end of World War II, Germany used their flying bomb to create havoc and large-scale destruction over England. However, there were two difficulties for the Nazis. Firstly, the bombs fell a few miles south of the centre.

Secondly, the Nazis did not know where exactly those bombs were dropped. If the bombs would have fallen north of the centre, it could have created much more trouble. The British administration, however, decided to befool the Nazis by spreading the news that the bombs were hitting the targets. The British did the trick so that in the future the Nazis could not alter the target. Several double agents helped the allies in this respect. The military supported the operation. But it had been a difficult situation for the political leaders. As the working-class people lived in the south, the bombs and the consequent destruction caused untold suffering to these hapless people. Here comes an ethical question— '…Politicians determining who was to live, who to die.' (Edmonds, David. 2014).

It is true that 'without the double-agent subterfuge' the devastation would have been much higher. So Churchill was perfectly aplomb, with no compunction or prickle of conscience for taking their decision. Nonetheless, the event is important because it captures the framework of a well-known epistemological riddle. (Edmonds, David. 2014).

The dilemma that is referred to as 'spur'

The thought experiment 'trolley problem' raises several important ethical issues that are directly related to artificial intelligence and ethics. Philippa Foot (1967) in her paper introduced an interesting problem. She asked us to imagine that a person is the motorist of a runaway trolley.

The trolley is near a curve where five workmen are repairing the track. So the fatal accident is looming. The trolley must be stopped but the breaks of it were not working. Then the driver observed a subdivision of the track. If he moves in that way, he would go to the right. If the trolley car was diverted to the right, then the five workmen would have been saved. But unfortunately, on the right side, there is one track workman who would be killed instead. Now the moral question arises, is it morally acceptable to divert the trolley? (Foot, Philippa. 1967)

We can, however, alter this thought experiment and replace the human pilot with an Artificially Intelligent Agent. The question that Foot asks, however, remains the same.

Fat Man scenario

Another version of the dilemma was formulated by Judith Jarvis Thomson (1985). He asks us to consider a case. Assume you're on a pedestrian bridge above the tram tracks. There's a runaway trolley rushing down the rail. It was out of control. There were 5 workers working on the track. But there's no way to stop it. A fat man, on the other hand, is waiting beside you on the pedestrian bridge. You're certain that if you throw the fat man, his body will halt the trolley. So, will indeed you force the guy onto the rail, sacrificing him in order to halt the trolley and save five other people?

However, Winston Churchill's riddle as referred to in David Edmond's book and the problem of the spur are not identical, though there are similarities. It was a matter of choice

for the British government. However, they opted for the option as a result of which different people and fewer people died. In the case of the spur the changed direction of the train would have saved five people, as a result, one other person would perish. To most people, it was 'morally obligatory' (Edmonds, David. 2014).

Philippa Foot's fourteen-page articles published in an abstruse periodical undoubtedly created a stir in the academic circle and it sparked a controversy that is still going on today. Even Foot could never have imagined it. Moreover, these philosophical debates reverberated in the minds of important moral thinkers—like Thomas Aquinas to Immanuel Kant, from David Hume to Jeremy Bentham. This, however, indicates the fundamental tension in our moral outlook. Philosophers make us stand face to face with some bizarre scenarios from which emerge philosophical puzzles. To understand the ethical dilemma mentioned at the beginning of this chapter regarding Artificial Intelligence, we need to have a clear idea about these thought experiments first.

It should be noted that Kwame Anthony Appiah invented the expression 'Trolleyology' to describe the condition mentioned in Foot's article. (Edmonds, David. 2014)

In this context, we want to cite an example. This will, indeed, show how trolleyology has entered the popular consciousness. In 2009 an interviewer asked the British Prime Minister to imagine that he was on a vacation and spending time on a beach. Suddenly he came to know that a disaster like an earthquake or a Tsunami was going to happen soon. The Prime Minister knew that on one side of the beach there was a relative of 5 Nigerians on one side of the beach and only one British person on the other. The Prime Minister could not let both of them know about the impending danger. How will he cope with the situation? Whom

does he alert first? The Prime Minister, however, replied, ‘‘modern communications alert both!’’ (Edmonds, David. 2014, p 10)

On the contrary, in reality, we cannot save everybody. The politicians would have to take the momentous decision with far-reaching implications. Not only politicians but sometimes health officials also face the same dilemma as health resources are not limitless. All these fall under the ‘Trolley problem,’ of course with a few variations.

Trolleyology, as we have seen in our earlier discussion, has a subtle but important distinction. For instance, on the one hand, choose between rescuing 1 and killing 5, or destroying 1 and saving five 5.

Furthermore, this subtle philosophical puzzle has already permeated into the realm of real politics. ‘Just war theory’ emanates from it. The cadets who come to the U.S. Military academy in New York are exposed to trolleyology. Sometimes military installations are targeted; as a result, some civilians may be killed. It is known in the parlance of International Politics as ‘collateral damage’. On the other hand, sometimes the Civilians are deliberately attacked. Indeed, there is a difference between these two types of situations but it comes under the purview of trolleyology.

Philosophers sometimes doubt whether problems of this type at all come under the precinct of trolleyology. But we should remember that trolleyology is no longer the exclusive domain of armchair philosophers. Moreover, it has already permeated to other fields, such as psychology, law, linguistics, anthropology, neuroscience, evolutionary biology and experimental philosophy so on and so forth. From Israel to India to Iran we come across trolley-related studies.

In this sanctum-sanctorum of trolleyology literature, there are some core questions: What is right and what is wrong, how we should behave in a particular situation, what is more, important in a given situation etc.

To understand the implications of trolleyology better, we need to look at the background of Philippa Foot in formulating this theory in a nutshell.

Background

It was 1920 when Phillippa Foot was born. The violence of the Second World War marked a lasting impact on her. Moreover, her ethical outlook was moulded by it (Edmonds, David. 2014). At the time she started teaching Philosophy in 1947, 'subjectivism' still ruled the roost and according to Foot, it hurt academia. Subjectivism tells us that there cannot be any objective moral truths. Moreover, the Vienna circle gave strong intellectual support to subjectivism.

Later they developed logical positivism. It claimed that either it must give us a concrete result, for example, 2+2=4 or like these statements 'Buses are nothing but vehicles'.
According to logical positivism, a proposition must be verifiable in principle through experimentation. Except for these two kinds, all other statements are meaningless. (Edmonds, David. 2014, p 14-15) The question, however, remains, where does the moral assertion or ethical statement stand?

There was, indeed, an alternative approach. It is the ordinary-language philosophy. As it is clear from her life story, Foot had very little time for this approach. Moreover, ordinary-language philosophy gave more emphasis on how 'language is deployed in everyday speech.' (Edmonds, David. 2014, p 10) Philosophers, however, would spend much time deconstructing subtle distinctions of our various expressions. According to them, before

resolving the different philosophical problems, this is necessary. Foot taught this approach to her students but in a casual manner.

However, after the war, Foot came into contact with Elizabeth Anscombe who had a vital though indirect role in trolleyology. Later, both of them came into contact with Iris Murdoch. One can find that their approach to philosophy was almost the same. They attacked the meta-ethics and were preoccupied with the 'virtues' in any particular moral dilemma and our approach to it.

One approach relates to moral obligation and duties (Categorical Imperative). For example, in any situation, we should not take recourse to lie.

Another approach is Utilitarianism. It asserts that the outcomes of an action are the most essential issue. We will have to see; whether any act protects the greatest lives or harvests more pleasure. Anscombe, however, introduced the word 'consequentialism' in the realm of Philosophy. This trio was attacked by a 3rd approach. It stresses the importance of character. Even though Crisp argues, 'Virtue Ethics' is a branch of deontological ethics there is no denying the fact that this third way of thinking was inspired by Aristotle and Aquinas and later John Rawls. (Edmonds, David. 2014, p 10).

In this context, we must mention the contribution of the legendary Austrian Ludwig Wittgenstein. Wittgenstein, indeed, had a lasting impact on Anscombe. According to Wittgenstein, the philosophical puzzles were the result of conceptual confusion. They were natural and easy to make. However, they were dissolvable by language analysis. In his own words, demonstrating how to get the bird from the cage was the aim of Philosophy. (Edmonds, David. 2014, p 10).

Moreover, Wittgenstein was more interested in the foundational issues of logic and language. He was skeptical that philosophy could make a contribution something to ethics. So this trolleyology problem would have been a bit alien to him.

However, many philosophers believed that moral philosophy was more than just an esoteric exercise, confined mainly to endless verbiage among intellectuals. It had a definite role in our day-to-day existence. Among others Foot also identified the problems in applied ethics. She wrote about them. However, she was studying the logic behind two things---abortion and euthanasia. Foot discards Utilitarian views of value.

On the other hand, two examples will prove without an iota of doubt that Anscombe was greatly influenced by politics and current affairs. During this time, 33rd American President Harry S Truman was once offered an honorary degree at Oxford University. Anscombe, however, protested against this. She delivered a passionate talk against the prize and forcefully told that the man who instructed the falling of a Nuclear Bomb for the very first moment in history ever could never be given any kind of award. For Anscombe, it wasn't just murder, it was a pogrom, as thousands and thousands of people were killed and other thousands were subjected to untold sufferings.

Anscombe's fury revolved around the concept of 'intention'. According to Truman, he intended to accelerate the end of the war, not to kill innocent civilians. Anscombe dissected the concept of 'intention' and to her Truman's declaration was not correct. From this incident, Anscombe's views on other moral issues took shape.

Regarding contraception and abortion, Foot and Anscombe possessed diametrically opposite views. Their relationship was permanently damaged because of this.

One can remember that the sixties were regarded as the decade of sexual liberation and feminist liberation. Anscombe, a devout catholic, was fervently defending the Roman Catholic churches' prohibition of contraception. To her, any pregnant woman who opted to have an abortion is a murderer.

Foot and Anscombe both did write philosophical research papers about a foetus's moral status. It was, indeed, a matter of strong disagreement among Philosophers. However, the constitutional right to abortion is now established. In the United States, abortion is now a legal right. (Edmonds, David. 2014, p 10).

It must be noted that abortion is legal in Indian law if the progression of the pregnancy would endanger the pregnant woman's life or cause harm to her mental or physical well-being. While the Supreme Court of India specified that Article twenty-one of the Constitution of India implicitly guarantees the right to privacy, a right to abortion can also be interpreted in this light. The Medical Termination of Pregnancy Bill was passed by both Houses of Parliament on August 10, 1971. Following that, it was approved by India's President. An unintended pregnancy can be terminated by a registered doctor in a government-established or maintained hospital or by a government-approved facility, under certain conditions.

In Britain, however, the legislation was enacted in 1967. Philippa Foot (1967) published her essay in the same year. In this article, she uses the trolley problem to explain the ideology of the double effect and explicitly distinguishes between doing and allowing.

Foot's article argued that the doctrine of double effect (DDE) could not be used to criticize abortion. Foot thinks that there is a distinction between envisioning an effect and intent. According to her, the DDE is grounded on the difference between whatever a person anticipates through his intentional act and what he aims. According to the author, the person

includes both those things that the person seeks as ends and those that he seeks as means to his ends. Foot claims that Bentham used the term ‘oblique intention’ to contrast it with the ‘direct intention’ of ends and means, and folks may as well use his jargon. (Foot, Philippa. 1967) The terms ‘double effects’ refer to the two effects that an action can have. These two effects are:

a) intended

b) predicted, not intended.

By the DDE, Foot means the proposition that it is occasionally allowable to give rise to by oblique intention what one does not straightforwardly intend. As a result, the distinction is important when making moral decisions in difficult situations. (Foot, Philippa. 1967)

One may note that DDE was proposed by St. Thomas Aquinas, whom most Catholics regard as their religion's preeminent theologian. (McIntyre, Alison. 2019). Even liberal intellectuals recognize his profound contribution in fields ranging from Philosophy of mind to transcendental studies and natural law theory. His body of work in ethics claims relevance even today. He settled on the principles which were required for a war to be turned into ‘just’. He declared that deliberate murder cannot ever be defensible. However, for the sole purpose of self-preservation killing could be morally permissible.

Here we can refer to a literal example cited by Edmonds from Nicholas Monsarrat’s book the *Cruel Sea* (Edmonds. 2014. p 28). It was a tale of World War II and the scene was the Atlantic Ocean. A British merchant convoy was attacked by German torpedoes. Ships were destroyed but there were many surviving members in the ocean who had to be picked up. The British commander was in a difficult situation. He was, in fact, in a pickle. It was necessary to sink the German U-boat. Otherwise, it would create havoc sinking ship after ship. But in

doing so, the resultant massive explosion would kill the survivor. So here is a paradox. What would he do? He eventually decided to destroy a U-Boat. In this situation, the commander foresaw the plight of the survivors but it was not his intention to kill them.

Abortion is only permissible in rare instances according to Catholic theology. For example, imagine that a pregnant woman is diagnosed with a tumour in her uterus. In such a case, a hysterectomy could be the only option to save the life of the woman. An operation here has no purpose other than to remove the tumour, not the foetus. The DDE has been incorporated into legal terminology, medical practice, and military rules. There is a difference between straight or purposeful intention and oblique intention. This thought experiment demonstrates the complexities of morality by differentiating between assassinating and allowing someone to end up dying (doing versus allowing) — a concern with impacts on our legislation, actions, science, police enforcement, and warfare. 'Right' and 'wrong' is not as simple as it's often made out to be.

Thompson, on the other hand, had a different point of view. Moral theories grounded solely on consequences, like consequentialism or utilitarianism, she claims, are insufficient to demonstrate why the certain act of killing is justified and not others. As far as she is concerned, if everyone has equal rights, then to sacrifice one even if we intended to save five would be wrong.

We must mention that Joshua D. Greene et al. studied how the brain functions when folks consider the first two variants of the trolley quandary (Edmonds. 2014. p 28). The conclusion is that the initial variant stimulates our rational and reasonable, areas of mind, and thus if we made the decision to press the lever, it was to save more lives. When we think of pressing the fat man off the bridge, however, it is emotional reasoning kicks in, and we think

differently about assassinating one to spare five. Now the question arises, are the emotions in this instance leading us to the correct action? To put it in other words, should we avoid sacrificing one, even if it is to save five?

It is believed that chemical and biological factors are involved in ethical decision-making. Co-relational research is still in the primary stage, but it is progressing rapidly. The role of oxytocin, testosterone, vasopressin, serotonin etc have been in the area of research interest for a long time. Researchers are keen to observe, how these chemicals alter behaviour, how they change the attitude towards certain things such as risk-taking, negotiation, bargaining or cooperation. But in this dissertation, we will not delve deep into this research, as they fall outside our central research theme.

In this context, we must mention that, whatever may be the reason, some of the experiments suggest that majority of the participants in trolley experiments would divert the trolley in the Spur and most of them would not push the fat man. A study conducted by the BBC online among sixty-five thousand participants shows that roughly four out of five agreed that the trolley would be diverted down the spur and one in four participants would agree that fat man should be thrown over the footbridge. Other studies also show that close to 90 per cent of the participants would not push the fat man rather; they preferred to divert the trolley.

We can situate the trolley problem in a driverless car or driverless train as well. The Google driverless cars are an example of new-age technological development. The questions arise, in a similar situation like the trolley problem, what would a driverless car or a driverless train do? Will those artificially intelligent machines ‘choose’ between killing five and killing one? In *Nicomachean Ethics* Aristotle distinguishes between types of wisdom---some are theoretical and some are, as Aristotle puts, ‘phronesis’, which can be translated as practical

wisdom. According to neo-Aristotelians, a person with practical wisdom can sense what is the right thing to do (Kraut, Richard. 2018). Can those driverless cars have situational appreciation? From the previous discussion, we conclude that, since self-driving cars offer one of the most transformative examples of the impact of artificial intelligence on society, the prevalence of the trolley problem in the context of self-driving cars is a bit problematic.

Moreover, the trolley problem is hypothetical---a thought experiment. It is easy to adapt to other situations, such as the Moral Machine project which incorporates this for the road (Awad, E. Dsouza, S. Kim, R. *et al.* 2018). But the original problem's basic simplification—- a trolley or a vehicle can be manoeuvred left or right, so the choice is therefore binary and binding—also ends up making it troublesome in aspects of Artificial intelligence - based reasoning.

It can be said that there is a risk of ascribing to AI a 'thought processes' or 'decision-making system'. The trolley issue tends to be a moral thought experiment in which we are forced to examine our beliefs and prejudices. But in real life, these values and biases are important and proved to be determinant factors in making a decision. If we look at some of the variants of the trolley problem, then we can better understand this.

The fat villain

In this scenario, we can imagine a fat villain where originally the fat man stood. And this person is responsible for putting the other five persons in peril. So, putting the bad guy to die, particularly if it protects the lives of 5 guiltless people, appears morally legitimate, if not even absolute necessity. This is similar to some other thought exercise recognised as the ticking time bomb situation, in which one is compelled to select between two morally dubious acts.

It offers a concept or ideas that explain our strong reactions and can teach us something about the fundamentals of morality. Foot and Thomson cast off an appeal to the DDE, but this principle, which was first recognized by Thomas Aquinas, has influential instinctive resonance. At its core, however, there is a distinction between 'intending' and 'foreseeing' (Edmonds. 2014. pp 39-42). This distinction does not carry any weight among the Utilitarians, but most non-utilitarians would agree that the nature of an intention is relevant in the judgement of an action. If it is the case, then how do these Artificially Intelligent devices make a distinction between 'intending' and 'foreseeing'?

To summarize, Philippa Foot gives a couple of examples. Firstly, she urges us to imagine a situation where a judge has to make choice between framing and murdering a blameless person and letting five guiltless to be executed in an uprising. Secondly, she asks us to imagine another scenario where a trolley pilot has to decide if he could turn a trolley to run over a naive man linked to a track or allow the trolley to drive and kill five innocent citizens. Foot does not support killing in the first scenario. On the contrary, she gave a different opinion in the second case. Foot, however, thinks that in the second case people might accept the doctrine of double effects. The doctrine of double effect singles out between the harm that is purely intended and harm that is only foreseen. Foot, asserts that the cases can be clarified by the difference between doing and allowing harm. (Woollard, Fiona. Frances, Howard-Snyder. 2021).

According to her, the judge in the first case needs to choose between killing one and allowing five to die. In the case of the trolley, the pilot has to opt for slaying one person and murdering other five people.

Judith Jarvis Thomson, on the other hand, changed the case slightly. She introduced one bystander, rather than the driver, to make the decision. The distinction was significant because the bystander is clearly choosing between murdering and letting die, but it still appears allowable to turn the trolley. However, later she argued for giving permission to the bystander to turn the trolley was erroneous. She presents a third option. In this option, one may turn the trolley on to and kill oneself. (Woollard, Fiona. Frances, Howard-Snyder. 2021)

We can see from the discussion that there is no agreement on how to solve the trolley problem. Many academics attempted to approach this issue from various angles. Some try to complicate it even more. But the fundamental question remains the same.

Can an Artificial Agent differentiate between doing and allowing?

We have seen that Artificially Intelligent machines are evolving at a great pace and thereby reshaping our infosphere. (Floridi, Luciano. 2018) Furthermore, our lived experience shows that the domain of offline and online is, indeed, getting blurred day by day. Floridi terms this 'onlife'. According to him, we live in the 'infosphere', which is analogue and digital, offline and online. (Floridi, Luciano. 2018) In the infosphere, however, we interact with humans as well as with artificial agents. Since we include artificial agents in our ethical framework, we arrive at a different situation, than we have seen before. One of the differences is, earlier we dealt with machines or instruments, which were not Artificially Intelligent. We can take the example of the film *Modern Times* again. (Chaplin, Charlie. Director. 1926.) This is an American silent comedy film, written and directed by Charlie

Chaplin. The machines or instruments depicted in the films were controlled by humans and not Artificially Intelligent Agents. We saw what happens when these machines malfunction.

On the contrary, these days, we come across machines that can learn by themselves. These machines can evolve by themselves. Even some intelligent machines control other machines. This happens within the scope of 'infosphere' and gives rise to a different situation where 'conflict of duties' (Frankena, William, K. 1973) arises involving artificial agents and humans. Hence on the one hand we have such intelligent machines and on the other hand, we have situations like trolley puzzles.

In this context one may ask, what about the example of a driverless car accident that I mentioned earlier and the trolley puzzle show? They point to a moral quandary. The questions are: if one person would be sacrificed to save many? Another question would be; can we kill one innocent person? In this case, we are not taking into account the aftermath of this action. These problems do not have a simple solution.

For example, Donagan (1977) says when someone chooses amongst duties, the person should choose which inflicts the least harm. Karl Popper (1966), however, termed this as the minimization of suffering. When people's rights and obligations clash, we must decide which one to uphold. Additionally, these thought experiments involve the subject making a quick and important decision. The exercise is useful because it demonstrates how difficult it is for a human being to make such a decision in practice. After all, there are so many variables involved. Hence, Trolleyology has raised an important ethical question: how should we treat others and go about our daily lives? It is a question that requires us to introspect and appeal to our intuition when we face a moral dilemma in day-to-day life. As

we have seen, any machine is less prone to introspection (some would say, cannot introspect at all), it is quite natural, that a self-driving car can execute a decision in a trice, but its decision-making process is unlikely to operate like a human being. It is a fact that self-driving cars take on a set of data from the surrounding using cameras, radar or other devices much as human drivers do use their sense organs and neural network. We know that they can identify objects and predict pedestrians by mimicking the human brain. Since the external environment is continuously evolving, is it practically possible for a programmed machine to keep pace with changes?

Moreover, the decision making process of human beings involves so many things---personal biases, gender differences, upbringing, education, hegemonic forces, current political scenario, culture, etc. It also involves neurochemical processes. Will a machine be able to take into account every element related to decision-making? Hence, among other things, the trolley puzzle in AI also indicates this 'realizability' problem.

The trolley problem, indeed, shows a fundamental tension/s between two predominant schools of moral thought---the utilitarian and the deontic. In a broad sense, the utilitarian perspective says that the most appropriate action is the one that achieves 'the greatest good for the greatest number'. Meanwhile, a deontic would assert that some actions–like killing an innocent person or telling lie– are wrong, even if they have good consequences. In both versions of the trolley problem, utilitarians might say one should sacrifice one to save five persons, while deontologists say the opposite.

According to studies, most people agree with utilitarians in the first version of the problem. They believe it is morally acceptable to kill one to save five. But in the fat-man

case, people lean on the deontological point of view and believe it's not acceptable to push a fat man to his death – again killing one to save five. (Edmonds. 2014) How will this difference be addressed in a situation where Artificial Intelligent Machines are involved?

Moral intuitions, as we know, have evolved to make us good social beings, which is the prime need of a just society. From a very tender age, we learn the principles like, 'do not harm anybody', 'do not tell a lie', etc. We learn from childhood that violence towards others is punishable and our intrinsic moral intuitions tell us it is wrong to take actions that physically harm others. Fat man scenario involves physical contact, harming one to save many is generally less acceptable in deontic doctrine, though this situation is just and acceptable in utilitarian ethics.

Another crucial difference between the spur and the footbridge case is that the latter involves using a person as a means to an end, which a Kantian might not accept. Treating others as individuals with their rights, wishes, needs and their rational agency, rather than simply as objects to be used at will, is a key aspect of becoming a good social being. Whether a Kantian or not, it is a fact that, people distrust those who use others as a means to an end. Again, our moral intuitions seem to accord with this principle.

So, the question remains how would you programme these nuances and put these into an artificial agent? At this juncture, it seems we are making a conscious effort to make things difficult ethically for research work in AI. But one would be mistaken to interpret our effort negatively. Much as an official hacker attempts to break into a secure system, to debug security lapses, our intention here is to provide a much stronger philosophical and ethical foundation for AI. Why we begin with trolleyology, can be better understood with the examples from real life.

Here is a resume of our discussion:

i) There is a need for building ethical machines. We have established this from the previous sections. We have seen that an Uber self-driving car killed a pedestrian and after that incident discussion of ethical machines gained additional momentum. But when we talk about ethics in the realm of artefacts, problems lie with their implementation, as there is disagreement on moral standards and, it is very tough to predict a situation. This paves the way for the second section of our discussion.

ii) The question that confronts us in the second section is, when we arrive at a certain situation like that of the trolley problem, what is kind of 'ethics' do we incorporate into the 'ethical machine'? We have seen that in real life, optimization problems are more complex. For example, how do you teach a machine algorithmically to overcome racial and gender biases in its training data? As a result, when we get to real-life situations, the problem becomes more complicated. Even the definition of morality varies, and there is no agreement on this.

This paves the way for the next chapter which consists of a literature survey. As time flows, however, the need for the ethical dimension of machines themselves preoccupies ethical and technological theorists. In a paper 'Contextualizing Ethics in the Realm of Robotics' (Bandyopadhyay, R. and Guha, M. 2008,), we traced the shift in the human thought process. There is literature aplenty, which also reflects this tension/s. In both cases, human beings engaged their reasoning to interpret a situation where sentient beings interact with machines. But what if a group of robots interacts with another group of robots in addition to humans? We can easily imagine a situation where interaction between man and machines takes place and simultaneously interaction among machines also occurs and artificially intelligent

machines inflict harm on humans. This begs the question, how should the anthropocentric harm principle be implemented in a situation where robots interact with both robots and humans?

By ‘anthropocentric harm principle’ I mean the ethical doctrines which put the interest of humans at the centre and protect humans from any kind of harm inflicted on them. Do we need to develop different kinds of ethical parameters to account for the emerging situation (robot-robot and robot-human interaction)? Given the current situation and the shift in ethical analysis following them, is there any more any distinction between a person who performs a moral act (moral actor/agent) and the one at the receiving end of such an act (patient)? Let us consider this as the central question of the next chapter.

So far we have seen how the notion of ‘ethics’, as well as ‘agency’, has evolved from time to time. However, the journey of ethics from ‘anthropocentrism’ to ‘ontocentrism’ is a long one. There are certain landmarks in this journey and we have tried to critically revisit these milestones to understand the next turning point in ethics. ‘Ontocentric’ Information Ethics formulated by Luciano Floridi plays an important part in my thesis. We have discussed this theory in the next chapter. Luciano Floridi sees Allan Turing’s theory and the ontocentric information ethics which follow from Turing Machine as the fourth revolution. According to him, after Copernicus, Darwin, and Freud, this metaphysical shift in ethics represents nothing less than a fourth revolution. Why does he say so, why does he criticise the existing doctrines of ethics, what are the basic features of information ethics and why according to Floridi Information Ethics is unique and ushered in the fourth revolution, have been discussed in the next chapter.

Chapter 3

Ethical Theories: Literature Review

In the previous chapter it has been discussed that there is a shift in the way in which we think about machine ethics. It may seem that, previously matters relating to ethical development, use of machinery and a kind of technological paranoia stemming primarily from the usage and development of machines were at the core of human thinking about these issues. In these situations, technological development has been perceived as an ominous sign to humans. It has been discussed in the previous chapter.

As the flows, however, the need for the ethical dimension of machines themselves preoccupies ethical and technological theorists. There is literature aplenty, which also reflect this tension/s. In both cases, human beings engaged her reasoning to interpret a situation where sentient being interact with machines. But what if, a group of robots interact with another group of robots along with human? We can easily imagine a situation where interaction between man and machines takes place and simultaneously interaction among machines also occurs. Consider the accident that took place in Mountain View, California on 26th September 2016. (Curtis, S. 2016) The accident involved a driverless car made by Google and a commercial van. According to the report published in the *Mirror* this accident was thought to be the worst involving an autonomous vehicle yet. What if, two driverless cars had collided?

What if, in such case accident a person inside one of the vehicles would have died? Such a situation would involve human-machine and machine-machine interaction.

This prompts us to ask: how to implement 'anthropocentric harm principle' in a situation where robots interact with robots and human? By 'anthropocentric harm principle' I mean the ethical doctrines which put interest of human in the centre and protect the human from any kind of harm inflicted on them. Do we need to develop different kinds of ethical parameters to account for the emerging situation (robot-robot and robot-human interaction)? Given the current situation and the shift in ethical analysis following from them, is there any more any distinction between a person who performs moral act (moral actor/agent) and the one at the receiving end of such an act (patient)? Let me consider this as the central question of this chapter. In order to delve deep in to the question, we need to clarify certain ideas regarding machine ethics in general.

In a pioneering work 'What is Computer Ethics' James H Moor (1985) points out that when computers were introduced they created immense possibilities. With the help of computers people could do things, they could not do before. This, according to Moor, creates a 'policy vacuum'. In the chapter five of the book *Philosophy of Computing and Information*, Deborah G. Johnson (2004) illustrates this position of Moor.

According to Johnson, when technology is involved in the performance of an 'act type', a new set of 'act tokens' may become possible. With an example he furthers his position. He asks us to envision a scenario in which we could play chess while seated face-to-face with a computer without engaging some other sentient beings. As a result, when individuals perform these tasks with computers, new sets of 'act tokens' become feasible, and these new

‘act tokens’ have characteristics that are distinct from ‘other tokens’ of the similar ‘act type’. It is possible to imagine that when technology alters the characteristics of tokens of an ‘act type’, the moral character of the ‘act type’ can shift. (Johnson 2004) It is known to all that chess is a mind game. In the game of chess ‘eye contact’ with the opponent is a pretty important tool to read her mind and at the same time to influence others. Many believe that a world champion and a chess prodigy, Magnus Carlsen has achieved his success due to ‘hypnotic abilities’ which he does with the help of eye contact only**.** Japan-born American Grand Master Hikaru Nakamura surprised everyone by wearing a sunglass for a game against Magnus Carlsen in 2013. This is a well-thought-of strategy that Nakamura developed against Carlsen. But if we replace Carlsen with a computer or a robot, then this strategy will have no effect. This is what I imply when I assume that when software alters the properties of tokens of an ‘act type’, the moral character of the ‘act type’ changes as well. Hence, the need to reformulate the existing ethical discourse.

Similarly, we can imagine that when a robot interacts with another robot (let me introduce the term Robot-Robot Situation or R-R Situation) the properties of ‘act type’ also changes, which eloquently speaks that the moral character of that ‘act type’ also changes. This should be discussed in more detail in the next two sections. To sum up, there are two issues that are going to be discussed in next two sections i.e. section I and section II:

i) What would be the nature of ethics when human being interacts with Robots? Since Robots are the offspring of technology, first we need to delve deep into the detail of the human-technology relationship and see how it changed with the time. I will discuss how the

ethical discourse also changed with the changes in the interaction between the humans and machines. (Let me call this as a 'Human-Technology Situation' or H-T Situation).

ii) What would be the nature of ethics when Robots interact with Robots? To put it in other words, is it possible to conceive of any ethical discourse when a Robot interacts with a robot? Is the ethical principle in the H-T Situation applicable to R-R Situation or do we need a different kind of ethics for the Robots? Since R-R Situation is an extension of the H-T situation; we have to discuss two issues separately. In the H-T situation what we have is a linear relationship with the machine which is not too complex to make out. But in the case of the R-R situation what we have is a machine often interacts with another machine simultaneously with the human. So Information and Communication Technology (ICT) has profoundly altered our way of living. We are living in a very complex phase of man-machine interaction. It is non-linear in the sense that interactions with both man and machine and machine and machine take place simultaneously. This, in a sense, affect our moral life also. Issues like 'privacy', 'property rights', 'surveillance', 'dependence', 'agency' etc. have altogether got new meanings. Hence the need to reformulate ethics.

Section I

In this section, we will review the literature of cognitive science and see what would be the nature of ethics when human beings interact with Robots. Let us recall that there are two issues or two levels of discussions going on here---one is the apparent continuity of the technologies involved in the H-T and R-R situation and the other is the possibilities of continuity between the ethics involved in the two situations.

So, I will start with H-T Situation. In this section, it will be discussed how the relationship between human beings and technology changed with time and ethical discourses also changed in order to give this transformation a metaphysical foundation. In the Cognitive Science literature, there are various perspectives on the moral relationship between sentient beings and technological artefacts in general, and computer-like devices in particular. I have divided this section into two sub-sections. In the first section, i.e. Section I (a), I will discuss classical theories of Ethics that are applicable to Cognitive Science as well as the theories of Cognitive Science that are used to understand the problem at hand (I have termed this section as Ethical Behaviour of Machine: Classical Approach).

In the second section, i.e. Section I (b), contemporary views on moral agency of artefacts will be discussed (I have termed this section as Views on Artificial Agency: Contemporary Approach). As we know, 'ethics' presupposes an 'actor' or 'agent' (in this dissertation the term 'actor' or 'agent' will be interchangeably used), at the core of these two sections there is one central argument that binds these two sections. That is---the

change in the nature of the interaction between human beings and technology provoked us to reformulate the ethical discourse from time to time. Again, this will pave the way for the discussion of my primary question----What would be the nature of ethics when a human being will interact with Robots in the latter half of twenty-first century?

Secion I (a)

Ethical Behaviour of Machine: Classical Approach

I will call the first approach as a classical one, as it involves a few classical theories of Ethics. This approach claims that artificially intelligent agents have properties that impact an individual's actions and the way they come to a decision. This approach looks at the relationship between humans and technology from the outside, where there is a both way impact. (Voort, V. D et. al. 2015).

In the second place, there are a few studies that see the relationship from a phenomenological perspective (Voort, V. D et. al. 2015). In this method, the role of technology in the directedness of sentients to the world is explored. Don Ihde mentions four relationships between sentient and technology. (Ihde, D., 1990)

According to Ihde, the first relation is the embodiment relationship. In this relationship, technology becomes part of the human. In other words, human being establishes a relationship with the world through technology. He cites magnifying glass as an example. According to him, this glass forms this kind of relationship with a person.

The hermeneutic relationship is the second one that Ihde mentions after discussing the first one. He tells that in this case technology interprets the world. After interpreting technology provides a representation to the user of the technology. He cites the example of a Thermometer. Ihde believes that a Thermometer interprets a specific property of the environment. This property is the temperature. Without the Thermometer, the temperature cannot be measured.

Next comes alterity relation. In this relationship, as Ihde holds, technology is viewed as a quasi-other. For instance, he cites the example of robots. To him, this artificially intelligent agents can be called quasi-other.

The fourth is the background relationship, where a human being is not necessarily conscious of the presence of technology. Though technology influences experiences the human is supposed to have with their environment. An example of technology that has this background relationship with the sentient is the presence of lights in a room. It is the light that impacts how the people see the space in the room, but the viewer is not always aware of the absence of dark.

Let us see how from the phenomenological perspective, the relationships between sentient and technology is seen by Peter-Paul Verbeek. The first approach, according to Verbeek, is 'Externalism'.(Cited in Voort, Van de. 2015.) In this approach, the interaction between a person and technology is explained as 'means and goals'. According to this, a person uses technology as means to attain the goal. It can be called as instrumentalism also.

The second view is 'Substantivism'. In this view, it has been discussed that it is the technology that governs our culture and develops certain independence or autonomy in its

own development thereby creating an overwhelming development of technology. Marlies Van de Voort and et al. (2015) coin this as a technological imperative.

Let's see what is 'transhumanism'. It is the second approach to the relationship between technology and sentient. By 'transhumanism' Verbeek means that it is impossible to detach humans and technology, as well as differentiate between the two. According to Marlies Van de Voort (2015) and et al., some transhumanists believe that the homo sapiens sapiens will become extinct soon because it will be surpassed by technology. In the first chapter, I discussed Stephen William Hawking's concerns about the advancement of AI. I have also mentioned a certain concern raised by Lociano Floridi too. The proponents of this view also question the value of human life in the transhuman world.

Now we will see what is the third approach that Verbeek (2006) talks about. Technological mediation is the third approach. It is said in this approach that machinery acts as the mediator between a person and the environment. As a result, in this perspective, technology is neither an impartial artefact nor a transhuman substitute for the sentient. Thus, Verbeek defines Ihde's embodiment and hermeneutic relationships as technological mediation. However, Magnani and Bardone (2008) use the idea of mediation to level technological artefacts as moral mediators. In their article, they argue that our moral decision-making can be improved by externalizing part of our moral tools into moral mediation. Under conditions of uncertainty this can be helpful. The authors cite internet as an example. They hold that internet can provide relevant data which can hegemonic effects moral decision-making by human. With this example the authors establish that artificially intelligent agents or artefacts can be a moral mediators.

In the classification of Ihde and in the description of Verbeek, the influence of technology on the sentient is recognized. However, this influence only exists as part of the relationship between the person and the environment, via technology. But in these theories, the technology is not conceived as an 'agent' making decisions by itself.

This notion, however, towards technology has changed. Verbeek furthers his position by introducing a broader conception of mediation. (Voort, Van de. 2015) At this juncture, he introduces the notion of cyborg intentionality. As we know, in the philosophy of mind literature, mental states are called 'intentional states' as it is directed towards something. This is called intentionality. Verbeek borrows this term from those the literature. By this, he means an intricate alignment of both humans and technology, rather than a clearly distinguishable human being directed towards the world. Verbeek however, confines himself to situations where fragments of technology are truly fused with the human physique. Therefore, this approach is not directly applicable to cases where the decision-making process containing experience and action, is shared within a cyborg construct. However, the technology that is considered in this study has some sort of intentionality. This intentionality is independent of this connection with the human or the presence of the human. In this scenario, the technology is dedicated to performing certain task and can function autonomously. This is called 'composite intentionality' according to Varbeek. (Voort, Van de. 2015) According to him, technology is directed toward the phenomena (world), that are different from the human one. Thereby it generates its own representations of the world. This may then construct a certain kind of representation for the sentient.

Hence, as mentioned earlier, we find that there has been a distinct shift in ethical thinking. Earlier ethical behaviour of human manufacturing and using of technology was at

the core of ethical thinking. Now, the ethical behaviour of the machine itself demands much more attention. The draft report of the European Parliament stated in the preceding chapter also shows that.

In the literature of cognitive science, machine ethics is considered to be a field of applied ethics but later this view also changed and now this is called as 'macro ethics' which is also a form of applied ethics (Floridi, L. 2010).

It has grown rapidly in the last decade. In the last decade, the focus of ethical thinking has expanded. Ethical behaviour of the machine itself is at the core of ethical thinking now. It is assumed that advanced robots behave ethically. This paves the way to delve deep more into the discussion regarding the Ethical Behaviour of Machine.

In order to account for the ethical behaviour of the machine, different approaches are discussed. There are various disagreements regarding moral standards in daily life. Now, it can be said that disagreement among ethical theorists about which norms moral agents ought to follow and disagreement about what it means to be a moral agent are interrelated. In this section, I will discuss some widely debated and at the same time hegemonic approaches to ethical theory. This will show the interrelations between disagreements about norms and about what it means to be a moral agent. Because the connections between these two disagreements become clearly visible as soon as two approaches are stated. The two approaches are: utilitarianism, on the one hand, and Kant's use of the 'categorical imperative' on the other (in other words, deontologism in general).

Utilitarianism (Stuart, Mill John. 1957 [1861]) is the notion which says that the best action is one that produce the best aggregate outcomes. By Mill's time, the utility

principle had a long history that dates to the 1730s, with origins as far back as Hobbes, Locke, and even Epicurus. Three British intellectual factions explicitly invoked it in the 18th and early 19th centuries. Though everyone agreed that an action's consequences for general happiness determined its rightness or wrongness, the reasons for accepting that principle and the applications to which it was put varied greatly. Nevertheless, even utilitarians disagree on which outcomes are important in this assessment. As Colin Allen et al (2000). hold, "the classical utilitarians were sentientists, holding that effects on the consciousness of sentient beings are ultimately the only events of direct moral significance." According to traditional utilitarian theory, the finest acts are the ones that result in the most happiness for the largest number of people. Mill held that "it is better to be a human being dissatisfied than a pig satisfied." (Stuart, Mill John. 1957). He believed that one of the qualities that set humans apart from animals was their ability to exercise moral agency. According to Colin Allen et al., this provides a real idea of 'morally good' that a follower of Utilitarianism could put into an agent's actions regardless of how the agent made the decision.

Now I will turn on to Immanuel Kant. For Kant, an action will be ethically decent, if it is completed 'out of the respect for the categorical imperative'. Kant contended that the superior moral principle is a standard of rationality that he termed as the 'Categorical Imperative'. Kant defined the Categorical Imperative as 'an objective, rationally necessary, and unequivocal criterion' that one must every time obey regardless of her natural desires or inclinations. (Johnson, Robert and Cureton, Adam. 2017) This principle, according to Kant, justifies moral prerequisites. All bad acts, Kant holds, are irrational, as they infringe the Categorical Imperative. Kant, on the other hand, did agree

with a number of his precedents that an examination of practical reason uncovers the requisite that rational agents follow instrumental principles. However, he contended that adherence to the categorical Imperative is a non-instrumental principle. As a result, moral requirements can be shown to be necessary for rational agency. He gets support for this reasoning in his ideology, which states that a rational will needs to be considered autonomous, or free, in the sense of being the author of the law that unites it. (Johnson, Robert and Cureton, Adam. 2017). Kant believes, the essential principle of morality, the categorical imperative, is the rule of an independent will. According to the preceding discussion, there is an idea of reason at the heart of the moral philosophy of Kant. For Kant, it is this sovereign reason within persons that unblock roads for seeing everyone as having equal worth and demanding equal respect.

There is a wealth of literature having a debate on what Kant means by the categorical imperative and what actually he means by acting in its honour. Allen directs everyone to acknowledge one explanation of it. He says that categorical imperative alludes that the agent acting because it determined that the act under consideration is consistent with the categorical imperative. In this context, unless an agent reasoned in certain ways, an action cannot be morally good. One can get extremely varying moral standards in Kant and Mill's work. It also expresses very new viewpoints about what a decent moral agent would be like. According to John Stuart Mill, an agent can be said to be morally sound to the degree that its performance donates to the overall well-being of the moral community. This is not the case for Kant. Inspired by this position, Colin Allen et al. conclude that an artificially intelligent agent can be termed a morally good agent if

it is programmed to do job steadily following the principle of utility, irrespective of the outcome. (Allen, Colin et al. 2000.)

A Kantian might not agree. To him, any claim that an agent is morally good points toward claims about the agent's inner deliberative processes. Kant may think that, in order to build a genuine AI moral agent certain specific cognitive processes should be implemented by the designers and we will have to incorporate this into the agent's decision-making process. (Allen, Colin et al. 2000.)

From the above discussion, it is clear that in everyday situations, we are guided by some ethical principles. It may be deontic, utilitarian or some mixed principle that we follow, or following Virtue ethics claim, we often say that "act as a virtuous person would act in your situation". Hence there are disagreements among ethicists regarding the moral standard. In the literature on Cognitive Science, the disagreements are classified at two levels:

1) Practical: On the level of the actual moral principle that we need to follow,

2) Ontological: In order to decide what it means to be a moral agent.

In order to solve this, two approaches will be considered.

Top-down Approaches: These are often called as 'Theoretical' approaches. (Powers, T.M. 2006.) Normative theories, such as the three laws of robotics (1.A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not

conflict with the First or Second Laws. [Asimov, I. 1990.]) may be programmed into machine, with the expectation that the machine will act according to these ethical guidelines. From the point of view of a programming engineer, this is a comparatively an easier job. It is plausible, but hard to execute as, lengthier will be the programme, the efficacy of the machine will be put to a question mark. If we consider the robots playing football, we can better understand this. Though day by day these robots are becoming faster, still it seems, achieving the efficacy of human being will remain at a tantalizing distance.

We can illustrate this position with an example. Patrick Lin, Keith Abney And George A. Bakey (cited it Taylor, Josua and Bringsjord, Selmer. 2012) adopted this Top-Down approach to solve the problem at hand. First they outlined the necessary and sufficient conditions for an ethically correct robot. To quote, "The engineering antidote is to ensure that tomorrow's robots reason in correct fashion with the ethical codes selected. A bit more precisely, we have ethically correct robots when they satisfy the following three core desiderata." (Taylor, Josua and Bringsjord, Selmer. 2012. pp 86-89)

D1. Robots only take permissible actions.

D2. All relevant actions that are obligatory for robots are actually performed by them, subject to ties and conflicts among available actions.

D3. All permissible (or obligatory or forbidden) actions can be proved by robots (and in some case, associated system e.g., oversight system) to be permissible (or obligatory of forbidden), and all such proofs can be explained in ordinary English. Then the authors discuss

four top-down approaches to solve the problem. (Taylor, Josua and Bringsjord, Selmer. 2012). For our present purpose, I am not going in to the detail of this discussion.

Bottom-up Approaches: These approaches are often called 'Modelling' approaches. (Wallach, W. & Allen, C.2009.) This strategy is based on the machine's capacity to recognize ethical behaviour. The proponents of this view claim that a machine can learn in the same way in which a child learns what is right and what is wrong. In Cognitive modelling, for example, we have seen that a machine can learn a language. Likewise, they can learn moral languages also.

I can illustrate this position with an example. We could model Linguistic Perception on the machine through parsing using context-free grammar (The man killed a deer) and context-sensitive grammar (Birds fly.) (Konar, Amit. 1999.) We could develop similar moral sentences and can show that machines could learn these too. At least this is logically or theoretically possible.

The main challenge of these approaches is that, the world where the robot is situated is constantly changing. In order to cope with the change, the programme needs to be modified constantly. Still, this will be inadequate to perform in an error-free way. Are any or both types of approaches to artificially intelligent machines like robots applicable to the contingent realities of life? I will discuss The Frame Problem of AI in this context. (Shanahan. Murray. 2016)

The environment of a robot is not stagnant; it is constantly evolving. Many different actions can lead to changes to it. The frame problem in artificial intelligence is based on the problem of pushing a robot to adjust to these changes. The knowledge base's information and the robot's conclusions combine to form the input for the robot's future action. However,

Dennett (1984) observes that a good selection from its facts can be made by discarding or ignoring irrelevant facts and eliminating results that may have negative side effects. An artificially intelligent agent should measure facts pertinent to a specific situation in order to perform a specific action. An artificially intelligent agent must first assess its current situation before looking for the best programme to help it decide what to do next. The artificially intelligent agent must look for slightly changing facts. After that, it analyses these to see if any of these have altered since the last inspection. Change can be of two types:

a) Relevant Change: Consider the consequences of an action.

b) Irrelevant Change: Do not examine facts that have nothing to do with the current task. Facts can now be investigated on two levels. The first is the Semantic Level, followed by the Syntactic Level.

In the Semantic Level type of information is being interpreted. The assumptions of how an object should behave could lead to obvious solutions. Some proponents of a pure semantics theory believe that accurate data can be derived from meaning. However, this theory needs to be substantiated. Now I will discuss in short what happens at the syntactic level. As we have seen that syntax in Philosophy of language means word-word relation. Similarly, in this case it determines the format in which the information is analyzed. To put it another way, it develops strategies grounded on the order and patterns of factual information. (Raredon, J and Blais. M. 1998.).

Various issues can arise when examining factual information. Maybe an inference is overlooked. It takes time to contemplate all of the factual information and its consequences. Some facts are investigated when they are not required. But the question still stands: will an

artificially intelligent agent be able to cope with the rapidly changing natural environment? Though there are few objections against the Frame problem of AI, but the basic question remains the same, how to condition an artificially intelligent agent properly to cope with the fast changing world? This is perhaps, one of the biggest problems of AI. It is difficult to incorporate sense of relevance in to artificially intelligent agents and since these agents are devoid of intuition they will be devoid of empathy and sympathy. Since moral dilemmas are non-algorithmic in nature how these agents will solve them?

Often it has been objected, that the average speed of doing some task of a new generation robot is still slow if we compare it with that of a human, because of its lengthy programme. Scientists are working on it. From present discussion, it can be said that this is considered to be one of the main problems of both Top-down and Bottom-up approahes.

So, if we take up either Top-down approaches or Bottom-up approaches, then it is hard to explain the fast changing real world. How can the artificially intelligent machine account for, adapt to or have an impact on this change? The situation is quite difficult and complicated because a robot may be programmed by several programmers and some robots are programmed to learn. Because of these facts, it is nearly impossible to predict how a robot will behave in any given situation. This confronts us with a question---how would a robot apply the notion of situational morality? Someone might say that even though different people may programme a robot, under the code of conduct in software engineering, the programme code has to be made available and in order to make the robot function, an individual or group of people have to be responsible for collating the

programme code, check for errors and run the robot. In doing so, its malfunctions (which are very likely to include ethical ones) can be fixed by the programmers.

If this is the case, then a new problem also arises. There is a difference between a 'programme' and an 'agent'. The interaction of human beings takes place with the 'agents' and not with the 'programme'. An 'agent', like a robot, is an entity in which a 'programme' is realized. 'Programme' is often formed or conceived by the programmer who is a human being. On the contrary, an 'agent' is something that acts. Here the 'programme' is the guiding principle through which an 'agent' acts. Even if we incorporate ethical notions or principles into the programme of a robot (using top-down or bottom-up approach), will it be effective in a real world situation where the robot's action takes place?

Last but not the least, these two approaches are anthropocentric, putting human interest at the centre.

It is clear from the previous discussion that, both the top-down theoretical approaches and bottom-up modelling approaches have their own difficulties. The main problem lies in their realization. So in order to avoid this and to solve our original question, the ultimate objective of building an artificially moral a6gent should be to build a morally praiseworthy agent. Colin Allen et al. believe that giving a morally commendable agent enough intellectual ability to evaluate the consequences of its deeds on sentient beings and use those evaluations to carry out effective decisions is essential. (Wallach, W. & Allen, C.2009.)

In order to build morally praiseworthy person, the discussion on the debate regarding the moral status of the artefacts is needed. It appears that the main question in this debate is: is non-organic ethical agency possible?

In order to understand this, we will discuss three views.

a. Strong View according to Verbeek (2006)

b. Moderate View after Illies & Meijers (2009)

c. Neutrality Thesis following Peterson & Spahn (2011)

    a. According to the strong view, both sentient and artificially intelligent agents can be moral agents, and technologies embody morality. Peter-Paul Verbeek (2006) is the proponent of this view. This view states that, technologies actively shape people's being in the world. Humans and technology do not have a separate (moral) existence anymore. Technologies have intentionality. Therefore, moral agency is distributed over both human and technological artefacts.

    b. According to the moderate View, artefacts have moral relevance. But they are not morally responsible or morally accountable for their effects. Technological artefacts are not moral agents. Artificially Intelligent Agents take a causal role in the sequence of events at times. They have relevance in moral action. This view is supported by Christian Illies and Anthonie Meijers. (Illies, C.F.R. & Meijers, A.W.M. 2009.)

c. The neutrality thesis says that technological artefacts are neutral tools. They have instrumental value. Artefacts sometimes affect the moral outcome of an action. But artefacts cannot be held responsible for its action. Technologies are not active. They are passive or neutral. Technology does not possess intentionality. This view is advocated by Peterson and Spahn (2011). They dismiss the possibility that artefacts having intentionality. It seems from their discussion that they are very much against non-organic agency.

So, there is no consensus regarding the moral status of the artefacts. This paves the way for detailed discussion on artificial agency. Because, without 'agency' moral status of the artefacts cannot be determined.

Section i (b)

Artificial Agency: Contemporary Approach

The modern approach to artificial agency necessitates determining whether or not artificially intelligent agents have emotive content, and enjoy a degree of autonomy. Likewise, whether it has mental states or free will, among other requirements of moral agency. In this dissertation, I will discuss some markers of agency for the time being, excluding emotion as a necessary and/or sufficient requirement of agency because it would be too broad a topic to cover.

I will start with Daniel Dennett's Conditions of personhood. (Dennett, Daniel. C. 1976.) and move on to other theories of Cognitive Science.

Dennett (1976) observed that we can ascribe agency to an artefact if it satisfies the conditions of moral personhood. Dennett here wrote on the concept of a person and outlined six conditions of personhood. It may appear that the metaphysical and moral notions of personhood is two different but interlinked conceptions. If we take the metaphysical sense, then we can say that if an agent is conscious and intelligent then it will qualify as a person. And in a moral sense, a person is thought to be morally accountable. Hence one can critique the person, or the person can be praised. Rights and responsibilities can be assigned to the person. Dennett thus asks, do the two notions of a person overlap? By the two notions, he means the metaphysical and moral conceptions. His question is: does the notion of a conscious and intelligent agent coincide with the notion of accountability and responsibilities? Or, he asks, being a person in the metaphysical sense is necessary but not sufficient for being a person in a moral sense? Dennett again asks, is being an entity to which states of consciousness or self-consciousness are assigned the same as being an end in itself, or is it just one precondition? Should the derivation from the original position be viewed as a demonstration of how metaphysical persons can become moral persons in Rawls' theory of justice, or as a demonstration of why metaphysical persons must be moral persons? (Dennett, Daniel. C. 1976.pp 176-177) He, however, does not attempt to solve this problem, instead outlined six conditions for a thing's being a person in the moral sense. These conditions are:

i. The First is that the entity can have rationality. It is the first criterion. It is conceived that the first and obvious theme is that persons are rational beings. In other words, the entity must have rationality.

ii. The second is the intentional stance that can be taken towards it. Individuals are said to be beings to whom states of consciousness are ascribed. Psychological, mental, or intentional predicates are ascribed to these individuals.

iii. The third theme, according to Dennett, is it must be the target of a certain kind of attitude. Dennett explains that, whether something counts as a person depends in some way on an attitude taken towards it, a stance adopted with respect to it. According to him, this implies that once we have established the objective fact that something is a person, we treat him or her or it in a certain way, but that our treating him or her or to it, this certain way is somehow and to some extent constitutive of him or her or it being a person.

iv. Fourth is the capability of reciprocity and returning attitude. According to Dennett, the object towards which this personal stance is taken must reciprocate. This act of reciprocation is expressed through a phrase: to be a person is to treat others as persons.

v. The fifth is verbal communication. According to Dennett because of this, non-human animals cannot get full personhood and enjoy moral responsibility. His contention is that; this is implicit in all social contract theories of ethics.

vi. Next comes the sixth. This is self-consciousness. According to Dennett, a person can be distinctive from others by being conscious in some unique way. So there must be a way in which we are conscious and in this way no other species is conscious. This, according to Dennett, is self-consciousness. (Dennett, Daniel. C. 1976.pp 177-178)

Though, Dennett thinks that these conditions are necessary and are not together sufficient condition for personhood. Following Fredric C. Young (1979) I conclude that,

if an artificial agent satisfies these conditions, then according to Dennett it can be said that it acquires personhood in moral sense.

From the last conditions of Dennett, it is assumed that if it is conscious then an artificial agent is said to be a moral agent. According to some cognitive scientists, this alone would qualify as the criterion of agency, more specifically artificial moral agency. According to Himma, K.E (2009), an artificial moral agent's possibility is dependent on whether it is conscious. The author argues that each of the various elements of the necessary conditions for moral agency presupposes consciousness, i.e., the capacity for inner subjective experience like that of pain or, as Nagel (1974) puts it, the possession of an internal something-of-which-it is-is-to-be-like. As a result, the authors state that whether or not the artificial moral agency is possible is dependent on whether or not ICTs can be conscious. According to Himma's article, even though the standard account of moral agency does not explicitly mention consciousness, it is reasonable to believe that each of the necessary capacities presupposes consciousness. The concept of accountability, which is central to the standard account of moral agency, should be limited to conscious beings. That is, the standard account of moral agency applies only to conscious beings, whereas non-standard accounts may not. (Himma, K.E. 2009) The author gives some reasons. The author begins by stating that it is a conceptual truth on the standard account that an action is the result of some intentional state - and intentional states are synonymous with mental states.

Second, the author cites Jaegwon Kim's (2006) work, which contends that if we lack access to the mental states that constitute reasons, we will lack the first-person self-conscious perspective that appears to be required for the agency.

Third, the author believes that, as a matter of substantive practical rationality, it makes no sense to praise or condemn something that lacks conscious mental states, regardless of how sophisticated its computational abilities are. Appreciation, incentive, disapproval, and punitive measures are all rational responses intended to prepare one for conscious states such as pride and shame. Furthermore, the author contends that the conditions for agency and moral agency, as well as the moral conditions for accountability, all presuppose consciousness. And the conclusion is that, while determining whether an artificial agent is conscious and a moral agent involves difficult epistemic issues, consciousness is a necessary condition for an artificial agent to be a moral agent. (Himma, K.E. 2009)

Dennett however, argues that it is unlikely that a robot would be conscious in principle just like the human, but later illustrates with a thought experiment which shows that sufficiently complex robot would be conscious. Dennett (1998) however, holds that conscious robots are impossible in the true sense of the term. Another thing is that, conscious robots would cost too much to build. He then reviews some reasons for the impossibility of conscious robots. Though these arguments are not error-free, still Dennett believes that these are the main challenges for a robot to be conscious.

1) Robots are purely material things and consciousness requires immaterial mind stuff. This is the argument from old-fashioned dualism.

2) From the standpoint of materialism, by definition robots are inanimate (inorganic). On the other hand, in an organic brain consciousness can exist. Artificially intelligent agents such as robots are artefacts, and consciousness abhors an artefact; only something natural, born not manufactured, could exhibit genuine consciousness.

3) Robots are always far too simplistic to be aware.

Though Dennett says that there are many criticisms cited against these arguments, still he believes that a robot cannot become conscious in principle. However, he says he would happily protect the conditional prediction: if an artificially intelligent agent cultivates to the extent where it can have well-controlled chats which come close to natural language, it would undoubtedly be able to compete with its individual monitoring devices (and the scholars who examine them) as a foundation of knowledge about what and why it does and feel. Can we not ascribe consciousness to a complex robot if it does that? This, perhaps, prompts Dennett to examine the problem from a different point of view, and supporters of Strong AI might find this argument interesting.

Dennett (1995, pp 422-426) however presents a thought experiment that protects the claim of strong AI. Followers of strong AI claim that artificial intelligence matches or surpasses the intelligence of sentient. Dennett asks us to suppose that one wishes to live in the twenty-first century, and that the only technology available to her is to place her body in a cryonic chamber, where she will be frozen in a medically induced coma and later awoken. Furthermore, the individual must create a super system to protect and power her capsule. The individual would now need to make a decision. The person could find an ideal fixed location that will supply whatever the person's capsule requires, but the disadvantage would be that the person would die if something bad happened at that location. It would be preferable to have a mobile facility to house the person's capsule, which could move in the event of an emergency. We can imagine placing herself inside

a huge robot. According to Dennett, these two techniques summarize the difference between static plants and movable animals.

If the capsule is placed within an artificially intelligent agent, the individual wants the automaton to select techniques that advance his preferences. This does not imply that the automaton has free will, but rather that it follows venturing directions to ensure that when choices are presented to the programme, it selects the ones that best function the person's interests. Provided these situations, the individual will indeed lay out the hardware and software to protect himself, as well as empower it with the necessary sensory systems and self-monitoring abilities. The super system should also be created to adapt to new situations and look for new sources of energy.

To make matters worse, while the individual is in deep freeze, other artificially intelligent agents are aware of what is going on in the outside world. As a result, the individual ought to design his automaton to know when and how to collaborate, form coalitions, or battle other animals. A ploy such as always collaborating will almost certainly get you murdered but never collaborating might not even end up serving your self-interests anymore, and the scenario may be so perilous that the person's automaton must start making numerous good decisions. The end result will be a self-controlling robot, an independent agent that emanates its own goals from the individual's initial target of continued existence; the desires that it was bestowed.

Advocates of Strong artificial intelligence argue that this automaton is not driven by its own needs and wants or intentions, but rather those of its developer. Dennett refers to this as 'client centrism.' (Dennett, Daniel. 1995) The automaton, according to client centrists, does

not have consciousness. Dennett, however, discards this idea. He says, if this logic is followed, then the logical conclusion you have to settle the same thing about yourself. One can come to the conclusion that she is a survival machine designed to protect her genes. She is therefore not fully conscious. We must accept that sufficiently advanced robots have intentions, objectives, and consciousness in order to avoid these unpleasant results. They are similar to us in that they are independent survival machines that have evolved through interaction with the outside world. Critics like Searle might admit that such a robot is possible, but deny at the same time that it is conscious. Dennett responds that such robots would experience 'meaning' as real as our 'own meaning'. They would have transcended their programming just as we have gone beyond the programming of our selfish genes. He concludes that this view reconciles thinking of yourself as a locus of meaning, while at the same time being a member of a species with a long evolutionary history. We are artefacts of evolution, but our consciousness is no less real because of that. The same would hold true of our robots. It seems that this debate will never end. But for our present purpose, I can conclude from the previous discussion that if a robot is conscious, then we can call it a moral agent, just like its human counterpart.

J. P Sullins (2006) argues that robot can be moral agents if it shows i) significant autonomy in terms of programming. ii) Ascription of intention. iii) Behaviour that shows understanding. iv) Responsibility to other agents. So according to the authors, 'autonomy', 'intentionality', 'understanding' and 'responsibility' are the markers of agency. If artificial agents show all of these qualities, then it can be said to be a moral agent.

If selfhood is synonymous with agency, then it can be said after Dennett, that if the (electronic) agent has the ability to narrate, then it will be considered as having agent hood. (Dennett, Daniel. C. 1992) That is not to deny the possible normative (moral, rational, autonomy-based) evaluation of agency even of the artificial type. That is, even artificial agents may be evaluated against some normative standards. We may, for instance, ask them: Are they moral? Are they rational? Are they free? The point we are making is however that the notions of agency and selfhood are related. Dennett in another thought experiment illustrates this position.

To say the least, an agent is a (or some would say, has a) self. This may be negatively put as: 'no selfhood, no agency'. But what does being a self (or having a self) consist in? Selfhood, according to Dennett consists of the ability to produce a 'self-narration'

Dennett (1992) seems to suggest that an artificially created artefact, like a computer programme is also a 'self-interpreter' and hence an agent: it can provide its own account of its activities. Dennett asks us to imagine that a novel writing machine (a computer created for that purpose) writes a story. The story begins with a sentence, 'Call me Gilbert'. Here Gilbert is a fictional, created self but its creator (the novel writing machine) is not a self in any conventional sense. Dennett's thought experiment-based story about Gilbert may serve to put this point across. As of now, we've envisioned the narrative, The *Life and Times of Gilbert*, clacking out of a computer which is nothing more than a box in the corner of some laboratory. Now Dennett alters the storyline slightly. He asks us to imagine that the computer does have hands and feet, or wheels. It can roam around the

nearby vicinity. It is endowed with a television camera which can serve as its eye. The device thus starts with the sentence ‘Call me Gilber’ and starts telling a narrative

Gilbert's excursions now include an impressive and ostensibly unrelated link to the expeditions of this robot traveling across the globe. Dennett claims that if someone hits the robot with a bat, the Gilbert's narrative will also include the individual. It narrates by trying to describe being hit by an individual whose depiction is comparable to the individual who just strike him. The robot occasionally becomes stuck in the closet and proclaims, ‘Help me!’

Who, according to Dennett, requires assistance? The answer is simple. Gillbert is the one who requires assistance. But who is Gilbert, exactly? Is Gilbert the robot, or just the robot's imagined self? If the robot receives assistance, it will leave a thank-you note. We will be unable to ignore the fact that the fictional Gilbert's career carries an intriguing similarity to the "career" of this simple automaton moving through the world at the moment. We can still argue that the robot's brain, or computer, knows nothing about the universe. It is not a self. It's simply a clumsy computer. It has no idea what it is doing. It has no concept that it is creating a fictitious character. (The same is true of one’s brain; it has no concept of what it is doing.) Nonetheless, the trends in the computer-controlled behaviour can be interpreted by us as accumulating biography—telling the story of a self. We are, however, not the only decoders. Of course, the robot novelist is also an interpreter: a self-interpreter who provides its very own account of its actions in the planet. (Dennett, Daniel. C. 1992)

According to Dennett, it is a grave error to ask about the spatio-temporal location of the self. The self is in that strict spatio-temporal sense, a fiction, just as the notion of a centre of gravity is a fictional thing. But like that notion, the idea of self has some practical value: its value lies in its social, legal, cultural, linguistic, historical and evolutionary applications. There is no ontologically distinct entity called 'self' but there is a self as far as all practicalities are concerned. We talk about it. It has instrumental value. This idea of an ontologically non-existent but practically necessary thing is neither unphilosophical nor illogical. But it has an important implication that Dennett explores at some length throughout his philosophical career. If the idea of self is intimately bound with the idea of narration, then even an artefact capable of producing a self-narration may be looked upon as having a self. If, as Dennett says, there is no absolute category called 'self' then an agent has a, or is a self in the mere sense that it can tell a story about itself. Gilbert, the robot, is an agent by the above standards because it is able to generate a narrative about itself. So from the previous discussion, I can conclude that, if the robot has the ability to narrate, then it has selfhood and can be called an agent---a moral agent.

Some may think that, we can ascribe moral agency to an agent if it is intelligent enough. At this juncture, we will discuss about the Turing Test and see how Dennett contextualized it. In 1950 British mathematician Alan Mathison Turing (Turing, A. M. 1950.) wrote an article which is considered to be one of the pillars in AI literature. In the article Turing begins with a question. The question is, could machines think? This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining

how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition, I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the problem can be described in terms of a game which we call the 'imitation game.'' (Turing, A. M. 1950.)

Turing Test refers to Turing's (1950) proposal for dealing with the question of if machines could think. Turing thinks about the issue of whether machines could indeed think to be 'too meaningless' to merit consideration. However, if we take into account the much more accurate somehow related question of whether a digital computer can operate well in a particular kind of game described by Turing as 'The Imitation Game', we have a question that allows for exact conversation. According to Turing, the Imitation game is like this: Assume we have a machine, an individual, and an investigator. The investigator is detached from the other individual and the machine in a room. The goal of the game is for the investigator to figure out which of these two is an individual and which is a machine. The investigator recognizes the other person and the machine by the labels 'X' and 'Y', but does not know which of the other person and the machine is 'X', and at the end of the exercise tells whether if 'X' is the individual and 'Y' is the machine or 'X' is the machine and 'Y' is the individual. The investigator may ask the individual and the machine the following questions: 'Will X please tell me as to if X plays chess?' Whichever of the machine and the other individual is X must respond to questions directed at X. The machine's goal is to trick the investigator into thinking the machine is the other individual; the other person's goal is to assist the investigator in identifying the machine. Turing believes that in about 50 years, it will be possible to

configure computers with a huge storage space to perform the imitation game so well that an average investigator will have no more than a 70% chance of making the accurate detection after five minutes of questioning. He believes that by the end of the century, things will change so much that one can speak of machines thinking without fear of being contradicted. (Turing, A. M. 1950)

In 2014, claims emerged that, as the computer program Eugene Goostman had fooled 33% of judges in the Turing Test 2014 competition, it had 'passed the Turing Test'. But there have been other one-off competitions in which similar results have been achieved. Back in 1991, PC Therapisthad fooled 50% of judges. (Amoth, D. 2014)

Now, I will see how Dennett defences Turing test. Dennett (1995) claims that Turing test is strong enough as a test of thinking. His claim is that critics have failed to recognize what the test actually is all about and as a result, they have dismissed it unjustifiably. According to Dennett, it is important to realize that failing this test is not supposed to be a sign of a lack of intelligence. He thinks that, it is a one-way test; failing it proves nothing. He thinks that Turing wanted to put an end to the discussion on intelligence. What Descartes did to Metaphysics, Turing did with Artificial Intelligence---to provide a foundation for AI by designing a philosophical conversation-stopper. Turing proposed, as Dennett thinks, a simple test for thinking that was surely strong enough to satisfy the sternest skeptic.

Dennett, however, opines that Turing's proposal had an opposite effect of that which Turing has envisioned. According to Dennett, Turing did not design the test as a useful tool in scientific psychology, but he designed it to be nothing more than a

philosophical conversation stopper. Dennett thinks that Turing conceived this test as a pretty strong one. Dennett shows that Turing was inspired by Descartes, who on Discourse on Method argued that there was no more demanding test of human mentality than the capacity to hold an intelligent conversation. To quote from Descartes, "It is indeed conceivable that a machine could be made so that it would utter words, and even words appropriate to the presence of physical acts or objects which cause some change in its organs; as, for example, if it was touched in some spot that it would ask what you wanted to say to it; if in another, that it would cry that it was hurt, and so on for similar things. But it could never modify its phrases to reply to the sense of whatever was said in its presence, as even the most stupid men can do." (Descartes, Rene'. [1637] 1960.)

Descartes emphasized the importance of ordinary conversation as a test for intelligence. So, intelligence, as Vincent Homburg (2008) thinks, can be conceived of requiring a number of abilities. These are:

i) Take coherent discourse (as opposed to isolated sentences) as input.

ii) Make inference and revise beliefs.

iii) Understand plans and make plans for conversations (to ask and answer questions, to respond to questions and to initiate conversation).

iv) Learn about the world and about language, in part via conversation.

v) Have background knowledge and add to this base through conversation.

vi) Remember what it heard, learned, inferred and revised. (Homburg, Vincent. 2008)

During his time technology was not as developed as it was in later age, so it is understandable that Descartes thought that machines cannot use natural language and cannot engage in conversation. Turing, however, in his article substituted the original question with the question of whether computers can use language.

Dennett believes, based on Descartes's strong suspicion, that ordinary conversation would put artificial intelligence under as much strain as any other test. Hence Turing concurs. Dennett, on the other hand, believes that Turing was prepared to formulate the assumption that nothing could probably get through the Turing Test by getting a win the imitation game instead of being capable of carrying out an infinite number of other highly intelligent actions. Dennett (1995) refers to this as a 'quick-probe assumption'. He believes that failing on the Turing test does not anticipate failure in the other areas, but success does. According to Dennett, Turing's test was so difficult that he believed anyone who passed it would disappoint us in other ways. Dennett concludes that "Turing test in unadulterated unrestricted form as Turing presented it, it is plenty strong if well used. I am confident that no computer in the next twenty years is going to pass an unrestricted Turing test. They may win the World Chess Championship or even a Nobel Prize in physics, but they won't pass the test fair and square." (Dennett, Daniel C. 1995.) However, if at all some computers can pass it 'fair and square', or in Dennett's word, "computer that actually passes the unrestricted Turing test" then logically it can be said that those computers have 'intelligence', or they are 'theoretically a thinking thing'. (Dennett, Daniel C. 1995) However, Dennett thinks that it is not possible and asserts that Turing did not conceive this test like this. Turing's point, according to Dennett, was that

we should not be species-chauvinistic or anthropocentric about the inner workings of intelligent beings. He reasoned that there could be non-human ways of being intelligent.

Secondly, Dennett (1995) thinks that we sometimes overestimate the cognitive prowess of the machine we are using. So, to him, it is not only a problem of Philosophy, but has a real social impact.

Whatever Dennett's point may be, the foundation of the Turing test lies in what is called as Functionalism. Functionalism is the doctrine which claims that 'mental states are functional states'. (Churchland, P. 1994.) What distinguishes something as a particular type of mental state is not its internal structure, but rather the way it functions or the role it plays in the system of which it is a part. As a reaction to identity theory and behaviourism, functionalism emerges. Functionalism, in contrast to behaviourism, maintains the traditional notion that mental states are internal states of thinking creatures. In contrast to identity theory, functionalism proposes that mental states are realised in multiple ways. (Dennett, Daniel C. 1995)

John Searle (1980) objects to this with his famous Chinese room experiment. Searle criticises strong AI's claim. Strong AI asserts that a computer is more than just a device for studying the mind; rather, an adequately programmed computer is a mind in the sense that computer programmes are able to comprehend and have other cognitive states. Searle specifically opposes this claim of strong AI which says that a properly programmed computer has mental states and the programme explains an individual's cognition. He explains this with the help of a thought experiment namely the Chinese room argument.

To quote Searle:

> Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch 'a script', they call the second batch a 'story', and they call the third batch 'questions'. Furthermore, they call the symbols I give them back in response to the third batch 'answers to the questions', and the set of rules in English that they gave me, they call 'the program'. Now just to

complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view - that is, from the point of view of somebody outside the room in which I am locked - my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese. Let us also suppose that my answers to the English questions are, as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that I am a native English speaker. From the external point of view - from the point of view of someone reading my ‘answers’ - the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program Now the claims made by strong AI are that the programmed computer understands the stories and that the program in some sense explains human

understanding. But we are now in a position to examine these claims in light of our thought experiment.

1. As regards the first claim, it seems to me quite obvious in the example that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, Schank's computer understands nothing of any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing.
2. As regards the second claim, that the program explains human understanding, we can see that the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and there is no understanding.

The formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output. (Searle, John. R. 1980.)

According to Searle, syntax is neither necessary nor sufficient for semantics. We can draw a conclusion from here. That is, programmes are neither necessary nor sufficient for minds. (Searle, John. R. 1980.)

There are many objections to Searle's notion. But his basic tenets remain the same. So if we say that 'intelligence' is the marker of agency (moral agency), then Searle may object that 'intelligent machines' that pass the Turing test cannot 'understand' anything or they lack semantics.

If Searle's argument holds its ground for Strong AI, one can move on to the argument made by the followers of Weak AI. Strong AI claims that an artificially intelligent system can think and has a mind. On the contrary, Weak AI asserts that a system can behave as if it can reason (think) and has a mind. As a result, the goal of artificial morality is to create artificial agents that can act as if they are moral agents. (Allen et al. 2006.) This can be called an *as if* approach. James Moore (2006.) also gives cognizance to this approach and says that we cannot be sure that machines in future will lack the qualities we now believe uniquely human ethical agents possess.

Now we have to look for what Mark Coeckelbergh (2009) had proposed. Coeckelbergh argues that we can replace the question about how moral non-human agents really are with the question about the moral significance of appearance which according to Johansson (2011) is a kind of the *as if* approach.

Coeckelbergh opines that we might be agnostic about what really goes on behind the scene and can focus on the 'outer' scene, the interaction and how the interaction co-shaped and co-constituted by how artificial agents appears to humans. He goes on saying

that human beings are justified in ascribing moral agency and moral responsibility to those non-humans that appear similar and we ascribe moral status and moral responsibility in proportion to the apparent features. In order to do this he introduced the term 'virtual agency' and 'virtual responsibility'.(Coeckelbergh, M. 2009) Coeckelbergh (2009) says the responsibility that sentient beings ascribe to each other and to some non-human beings is dependent on how the other is experienced and appears to them.

From this discussion it can be concluded that robots should be considered *as if* they are moral agent if it passes moral version of Turing Test suggested by Colin Allan et al. (2000) as it would mark it as if it has intelligence.

Colin Allan et al. (2000) conceive Moral Turing Test as a functionalist method for deciding if someone or something is a moral agent. The authors argue that if two systems are similar in the respect of input and output and if they have the same moral status, then one is moral agent and the same applies to the other. In this section it has been discussed that in both ethical theory and day-to-day talk about ethics, people disagree about the morality of various actions. There are disagreements about the moral standard also that we have mentioned earlier. For example, Kant claimed that it is always immoral to lie, no matter what the consequences are. Though, Singer argues that Kant's own principles do not entail this conclusion. But in moral philosophy, Kantian ethics is considered as opposed to utilitarian ethics. A utilitarian, on the contrary, would hold that lying is justified whenever its consequences are sufficiently good in the aggregate. The authors hold that "life is rife with disagreements about the morality of particular actions, of lifestyle choices and of social institutions. In the face of such diverse views about what standards we ought to live by, an attractive criterion for success in

constructing an AMA (Artificial Moral Agent) would be a variant of Turing's (1950) 'Imitation Game' (aka the Turing Test)." (Allen, Colin et al.2000.) The authors discuss that in the typical type of the Turing Test, an 'interrogator' has to distinguish a machine from an individual. The interrogator has to do this depending on the interaction with the sentient and machine via written dialectic. A machine could pass the Turing Test if, when paired with a sentient, the investigator cannot identify the sentient at a level above chance. And if this happens again and again then we can say that it has passed Turing Test. However, Turing wanted to conduct a behavioural examination that avoids differences about standards defining intelligence or fruitful gaining of natural language. Likewise, the authors proposed a Moral Turing Test (MTT) which might likewise be planned to set aside differences about ethical norms by restricting the standard Turing Test to conversations about morality. The authors argue that if a sentient investigator cannot recognize the machine with greater than chance accuracy, then the machine is a moral agent under this criterion. (Allen, Colin et al.2000)

There may be objections to conceiving the test like this. One limitation of this approach is discussed by the authors. This is the focus on the machine's capacity to communicate moral judgments. Followers of Kant can be pleased with this emphasis, because Kant needed that a good moral agent acts not only in a particular manner, but also as a result of reasoning in a specific way. Similarly, both a utilitarian approach and common sense indicate that the MTT places far too much importance on the capacity to articulate one's reasons for actions.

Mill believes that numerous actions are morally decent regardless of the agent's motivations. Some people believe that young kids, or even pups, are moral agents despite their inability to articulate the reasons for their actions. An alternative MTT could be planned in such a way that the investigator is given pairs of explanations of actual, morally significant

activities of an individual and an AMA, ousted of all examples that would identify the agents. The machine fails the test if the investigator recognises it at a higher level than chance. According to the authors, one issue with this variant of the MTT is that differentiating is the incorrect criterion. Because, according to them, the machine may be recognisable for consistently acting better than a human in the same situation. Instead, the writers suggested that the 'interrogator' be asked to determine whether one agent is more moral than the other. If the machine is not recognised as the less moral member of the pair significantly more often than the human in this situation, it has passed the test. This is known as the 'comparative MTT' (cMTT). (Allen, Colin et al.2000.)

So for the present purpose, we have come to a decision that if an agent passes the comparative Moral Turing Test or moral version of the Turing Test, then the agent should be called as moral agent.

In this context, we need to look at the conception of 'agent' and 'agency' in the engineering literature. In these literature, an 'agent' is something which does some purposeful activity. In his article 'Agent orientation in Software engineering', Gerhard Weiß (2012) defines an agent as: "An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives."

In this sense, a software (software agent) or may be a hardware is an 'agent'. There is a given target objective and an 'agent' is used to meet that target. So it can be said that, in order to fulfil a given target objective, an 'agent' is used. In this sense, a hardware using any form of energy can be an 'agent' in this literature. If it has autonomy, then it can be called as

an agent. (Joseph, Sam and Kawamura, Takahiro. 2001.) Individually or together it does some purposeful activity. For example, we can use a robot which sorts out the potato from the field. So it is an ‘agent’, since it serves some purpose. (Konar, Amit. 2016.)

We can conceive of ‘agents’ who work in a team. Supposing that, in a real estate farm a robot is used to mix cement, sand etc. and then it passes the mixture to another robot which carries it to the destination. In this case, robots work as a team and we can call each of them as ‘agent’.

In these literatures, these ‘agents’ do have ‘autonomy’ like that of a human being. When these robots have some learning ability, then they can be said to be ‘autonomous agents’. So according to some, learning ability is the prime mover for the ‘autonomy’ of Robots. Through this ability, it senses the environment and learns accordingly. When learning ability is not incorporated, it works in fixed programme architecture and there will not be autonomy for those ‘agents’. It can learn from its mistakes just like that of human beings. When it successfully senses the environment and alters its learning process, then we can give the robot some rewards in the form of a certain score. So, it can learn the penalty-reward mechanism.

From the previous discussion I can colnclude that, in order to be an ‘agent’, it should learn the hitherto unknown environment. Of course, in the initial state, it will commit some mistakes and from these mistakes, it will learn and in this way the process of its learning continues. One may object that, while sensing and learning the programme, the database will be large enough to handle. But that is not the case since we consider domain-dependent learning. In order to perform a certain task, it needs to prepare its environment in a trial and

error process. It needs some time to learn. After learning, the possibility of committing mistakes minimizes.

If we consider 'autonomy' as the marker of agency, can we say that human beings always enjoy autonomy? Consider a person who books a ticket at a station, can we say that he is autonomous? In the natural environment, the autonomy of the human being is curtailed in many cases. Only a few people enjoy autonomy. Robots have autonomy just like a human being does have in most of the cases. (Konar, Amit. 2016.)

In most of the AI literature, human's autonomy is mimicked and we represent it in the realm of Robotics. So, the artificial agency is nothing but mimicking natural agents. It is a 'copy' of the natural agency. Since artefacts do not understand ethics in the sense a human being does, we need to develop different kinds of ethics for those manmade agents. For that, we need to develop an ethical programme layer which will remain in the outer domain of the programme. We need to develop domain-dependent ethics for them.

In the natural environment also, we get domain-dependent and subjective ethics. We can say that the ethics of a teacher is different from that of a businessman. Just like that, we need different kinds of ethics that applies in a different situation. But in the case of robotics, it is difficult to conceive of some generalized ethics as the programme may get slow and as a result, the robot may not serve the purpose. So instead of some generalized ethics, we need to develop domain-dependent ethics for artificial agents. (Konar, Amit. 2016.)

This paves the way to discuss another problem mentioned in the beginning of the section which arises when a robot/electronic person, now a moral agent, interacts with another robot and with the human simultaneously. In this situation, new sets of 'act tokens' also

become possible and these new ‘act tokens’ may have properties that are distinct from ‘other tokens’ of the same ‘act type’. So, it is possible to imagine now, that when technology changes the properties of tokens of an ‘act type’, the moral character of the ‘act type can change’ in this situation. We have seen this in previous sections. Now, we will have to analyze the other side of the story in the next section.

Section ii

Logic for R-R Situation

Since this field is changing thick and fast, the nature of the interaction between human and machine also changes. We confront with a situation, where robots as electronic persons interact with other robots while interacting with humans. Sometimes this interaction takes place simultaneously, sometimes it overlaps. Hence, this situation is an extension of the previously discussed H-T Situation. Since the problem in this situation arises differently, we have to discuss the application of ethical principles that can be applied in this situation separately.

Consider a report which states that one of Google’s self-driving cars was involved in an accident with another car in El Camino Rea and Phyllis Ave in Mountain View, California. (Gibbs, S. 2016.) We have mentioned this incident in the previous section. Inspired by this incident, I can design a thought experiment. I can imagine a similar situation some times in future when a self-driven car may collide with another self-driven car. And then one of the cars collided with another one which was driven by

a human being. In order to complicate the situation further, we can imagine that a person dies in each of the two accidents. This confronts us with a few questions. Firstly, who would be held responsible for the accidents; the car owner or the manufacturer or the programmer/s or the end-users? Who would be morally responsible for the death of two persons? This is what I mean when I say that technology changes the properties of tokens of an 'act type', the moral character of the 'act type' also changes. And this prompts us to re-contextualizing ethical discourse to accommodate and/or explain the emerging situation.

If we conceive ethics purely from an anthropocentric viewpoint, then it would be difficult for us to account for the situation where robots interact with robots or humans and robots interact simultaneously at a given situation. If a moral dilemma or conflict of duty arises in such situation, then what would be the ethical principle that needed to be applied in order to resolve this? Sometimes we associate 'morality' with the 'mind' or with the nature of 'a sentient being'. But we have seen in the previous chapter that the nature of the interaction between man and machine changes. Situation emerges where robots interact with robots in course of doing something. As a result, thinking regarding the 'discourse of ethics' also shifts. In order to address computer related ethical problems, like privacy, property rights, pirated software uses and virtual harassment, we have computer ethics. In order to give computer ethics a metaphysical foundation and in order to set free the notion of 'morality' from anthropocentrism, Luciano Floridi. (1999) used the concept of information ethics.

Now I will discuss the concept of Information Ethics developed by Luciano Floridi. Floridi discussed at length why existing ethical doctrines cannot provide a philosophical base for computer ethics. In his article Floridi (1999.) illustrates his position. He says that in the virtual world or what I have called the R-R Situation an agent's action is like that of 'role playing'. Because of the process's distances, the nonmaterial essence of information, and simulated interplay with faceless people, the infosphere is envisioned as a mystical, 'political, social, and financial-dreamlike environment' in opposition to the real world. (Floridi, L. 1999, P2.) As a result, a person may believe that her actions are unreal and unimportant. As in murdering a rival in a virtual reality world. So the person does not feel responsible for her action and she has a 'moral sanction' for her action. So, the situation calls for a new ethical discourse. Folridi's arguments are as follows. Virtual nature of the action makes it undetected and leaves no perceptible effects behind. Modern information and communication technology distances its agents from its action. As the 'action' takes place in a computer-mediated way, 'actor' diminishes her sense of direct responsibility.

The high level of control and compartmentalization of actions tends to restrict them and their evaluation to specific areas of potential misbehaviour. Some CE case studies show that human nature when left to itself is much more Hobbesian and Darwinian than Consequentialism. The increasing numbers and varieties of computer crimes committed by perfectly respectable and honest people show full limits of an action-oriented approach to Computer Ethics. The infosphere is constantly changing and it is complex in its nature. So any reasonable calculation or forecasting of a long-term aggregate value of the global consequence of an individual action is impossible.

The individual and his/her rights acquire increasing importance within the information society, not just as an agent, but also as a potential target of automatically tailored actions, yet individual's rights are something that consequentialism has found difficult to accommodate.

In order to develop his own position and to provide the metaphysical foundation for Computer Ethics, Floridi first criticizes Virtue ethics, Consequentialism, Contractualism and Deontologism and then tries to establish a new ethical paradigm, which is known as Information Ethics. According to Floridi, Virtue ethics is 'agent-oriented', 'subjective' ethics. It is intrinsically anthropocentric and individualistic. It is only applicable to 'agent'. According to Floridi, this has Greek roots in the individualist conception of the agent and metaphysical interpretation of his functional development and partly because of contemporary empiricist bias. But according to Floridi, we live in a culture based on ICT where this ethical doctrine cannot be applicable. (Floridi, L. 1999.)

Floridi maintains that Consequentialism, Contractualism and Deontologism are three well-known theories that concentrate on moral value of action performed by the agent. (Floridi, L. 1999.) They are 'relational' and 'action-oriented' theories, intrinsically social in nature. They focus on the moral value of human action very differently. He thinks that, while Consequentialism and Contractualism perceive moral value from a posteriori (i.e through the assessment of their consequence in terms of global or personal welfare) view point and Deontologists look at it from a priori point of view (i.e through universal principles and individual's sense of duty). According to Floridi, agent-oriented, intra-subjective theories and action-oriented inter-subjective theories are inevitably anthropocentric. But he mentions Kantian version of Contractualism which take in a relative interest on the 'patient', which is,

as Floridi holds, the third element in a moral action. Medical Ethics, Bioethics, Environmental Ethics are example of this non-standard approach. They attempt to develop a patient-oriented ethics in which the 'patient' may be not only a human being but also any form of life. (Floridi, L. 2008.)

Now Floridi discusses the shift between Computer Ethics and Information Ethics. He asserts, information ethics provides basis of computer ethics. Information ethics, according to Floridi, will not be beneficial to solve the problem, but it will provide the grounds for the moral ideologies that will escort the problem-solving procedures in computer ethics.

Information Ethics, as an entity-oriented and ontrocentric theory accepts following principles and concepts. To quote Floridi (1999), these are "Uniformity of becoming, Reflexivity of Information Processes, Inevitability of information processes, Uniformity of being, Uniformity of agency, Uniformity of non-being, and Uniformity of environment."

Information Ethics, like any other non-standard ethics, maintains that each object is an expression of being. Every object has dignity, which is constituted by its mode of existence and essence. This deserves to be respected, and thus places moral claims on the interacting agent and should contribute to the constraint and guidance of his ethical decisions and behaviour. This, according to Floridi (1999, p3), is the 'ontological equality principle'. It indicates that any form of reality (i.e. information), just for the fact of being what it is, enjoys an equal right to exist and develop in a way which is appropriate to its nature. The 'ontological equality principle' (Floridi. L. 1999, p3.) presupposes a perspective, which is as object-oriented as possible.

So, Floridi's information ethics is based on three fundamental concepts:

a) information ontology

b) the agent/patient pair and

c) the infosphere. (Floridi, L. 2008.)

Before diving deep in to Floridi's ethical theories these concepts need to be clarified. In another article Floridi (2004.) begins by saying that, 'Information can be said in many ways' just as being can (Aristotle, Metaphysics T 2) and the correlation is not accidental. Information in its cognate concepts like computation, data, communication etc. play a key role in the ways we have come to understand, model and transform reality." According to him, information is a 'multi-layered and polyvalent concept'.

First and foremost thing is that, Floridi's information ethics looks at information as an 'entity'[10]. So he endorses an ontological approach in order to explain it. We can easily imagine looking at the universe from a chemical perspective. According to this viewpoint, each object and process will match up a specific molecular summary. A sentient, for instance, can be described as having somewhere around 45% and 75% water. From an ontological information-based standpoint, the same sentient being is characterized as a constellation of data, that is, as an information-based entity---information coded in the gene, information that passes through the neuron, physical information that are being received through sense organ etc. So, everything on earth can be described through information. It is noteworthy that, while describing human I have described the term 'can be described as'. These are significant because they emphasize that the object is characterized in those aspects, not that a sentient - or any other entity - is essentially or solely a large group of data. It is simply a way of viewing entities or, more precisely, the explicit selection of a Level of Abstraction. (Floridi, L. 2006.)

[10] In this section 'entity' and 'object' has been used interchangeably.

The agent/patient couple represents any information-based object that either causes changes in the environment (an agent) or is the receiver of these adjustments (a patient). Floridi wilfully dissociates his doctrine from that of anthropocentric ethical theories stated in the previous section. Floridi's ontological method can offer a very broad definition of an object. An information-based entity, according to him, is not required to be a living thing, have consciousness, or even be embodied. At a given level of abstraction, an information-based entity can be an agent and patient or both. It can be an individual, animal, plant, or anything that has existence, from an artwork and a novel to a planet and a rock; anything that could or would exist, such as a future generation; and anything that was but is no longer, such as one of our forefathers or an old civilization, or even an ideal, intangible, or intellectual object. From this point of view, information-based systems, instead of simply living systems in general, are elevated to the role of agents and patients in any morally significant action, with environmental processes, changes, and interactions described equally informationally. Floridi defines the 'infosphere' as the sum of all information-based entities and their relationships. It can be assumed of as the informational equivalent of the biosphere, as long as we recollect that the biosphere can also be regarded informationally at a given level of abstraction. Floridi defines information ethics as: "an ontocentric, patient-oriented, ecological macroethics." (Floridi, L. Forthcoming) Information ethics treats every entity as an expression of being. It says that every entity has a dignity. They are constituted by their mode of existence and essence. Floridi thinks this as an ontological equality principle. By this he means that any form of reality, has a minimal, initial, equal right to exist and develop in a way suitable to its nature. It is like we might go beyond extending ethics from human to animals so that we take in anything –any 'informational' object. Floridi holds that biocentrism will be replaced by

ontocentrism. Ontocentrism holds that there is something even more rudimentary than life, which is being and something even more foundational than suffering. This, he terms as ‘entropy’. (Floridi. L. 2006.)

Entropy describes any type of exhaustion of informational object. Corruption, destruction and pollution come under this head. By Entropy he means ‘any form of impoverishment of reality’ (Floridi. L. 2006.). Floridi does not use the term from thermodynamics.

The Method of Abstraction has been formalized in Floridi and Sanders work. (2004) The jargon has been impacted by Formal Methods, a branch of computer science that uses discrete mathematics to specify and analyze the behaviour of data systems. But in Information Ethics, the idea is not technical. Floridi illustrates this with an example. To quote from his own words, “Let us begin with an everyday example. Suppose we join Anne, Ben and Carole in the middle of a conversation. Anne is a collector and potential buyer; Ben tinkers in his spare time; and Carole is an economist. We do not know the object of their conversation, but we are able to hear this much: Anne observes that it has an anti-theft device installed, is kept garaged when not in use and has had only a single owner; Ben observes that its engine is not the original one, that its body has been recently re-painted but that all leather parts are very worn; Carole observes that the old engine consumed too much, that it has a stable market value but that its spare parts are expensive. The participants view the object under discussion according to their own interests, at their own levels of abstraction (LoA). They may be talking about a car, or a motorcycle or even a plane. Whatever the reference is, it provides the source of information and is called the system. Each LoA makes possible an analysis of the system,

the result of which is called a model of the system. For example, one might say that Anne's LoA matches that of an owner, Ben's that of a mechanic and Carole's that of an insurer. Evidently a system may be described at a range of LoAs and so can have a range of models." (Floridi, L., 2006, p26.)

Information ethics can be equated to certain other environmental methods. Biocentric ethics typically bases its interpretation of bio-entities and eco-systems moral worth on the innate merit of life and the inherently zero tolerance of distress. It aims to create a patient-centered ethics in which the 'patient' is not always a sentient being. Any type of living thing can come under its purview. When compared to other interests, any type of living organism is fated to appreciate several basic moral rights that are required to be honoured. According to Floridi, biocentric ethics asserts that the welfare of living entities should start contributing to steering the agent's ethical choices and restricting the agent's moral behaviour. But this is not what Floridi wants to mention here. Hence patients are placed in the middle of the ethical discussion as a source of moral concern, whereas agents are pushed to the fringes. If we replace 'life' with 'existence', it should be clear what information ethics wants to accomplish. Information ethics is an ecological ethics that substitutes 'biocentrism' for 'ontocentrism', implying 'being' is more elemental than life itself. Hence it believes in the existence and well-being of all entities. It also believes that entropy is more elemental than suffering. By the word 'entropy' Floridi means any form of obliteration. Information ethics assesses a certain ethical agent's responsibility with reference to its role in the development of the infosphere. Any action that adversely affects the entire infosphere and contributes in the enhancement of entropy level is not desired according to this doctrine. (Floridi, L., 1999.)

Who is a moral agent according to this ethical discourse? Floridi says if an agent can interact, if it can do work autonomously and can perform morally quantifiable actions, then it can be said as moral agent. (Floridi, L. and Sanders, J. W., 2004).

In Floridi's information ethics morally right and wrong are determined by for laws--- "a) entropy ought not to be caused in the infosphere. b) Entropy ought to be prevented in the infosphere c) Entropy ought to be removed from the infosphere; d) The flourishing of informational entities as well as of the whole infosphere ought to be promoted by preserving, cultivating and enriching their properties." (Floridi, L. 1999)

These principles express how does an agent should act in the infosphere. The laws are arranged from most important to least important.

In information Ethics, "the duty of any moral agent should be evaluated in terms of contribution to the sustainable blooming of the infosphere, and any process, action or event that negatively affects the whole infosphere – not just an informational object – should be seen as an increase in its level of entropy and hence an instance of evil. The four laws are listed in order of increasing moral value. They clarify, in very broad terms, what it means to live as a responsible and caring agent in the infosphere." (Floridi L. 2008.)

Concept of Floridi's Information Ethics emerges after Alan Turing's theory. Floridi sees it as the 'fourth revolution,' following the Copernican Revolution, Darwinism, and Freudianism. Human beings have been pushed from the centre to the periphery in each previous 'revolution.' In Copernican Revolution humans were thrown away from the centre of the cosmos. That is to say living beings just aren't motionless at the center of the solar system. In Darwanism humans were removed from the biological kingdom. This means living

beings aren't really separated from the rest of the animal kingdom. Freud shows that not all the parts of the mind are rational. That is to say living being isn't logical thinkers entirely open to themselves. So humans were removed from the centre of rationality as well. Fourth is the Turings invention which brings about a change in the realm of Ethics. Turing shows that living beings are really not separated actors, but information-based species or 'inforgs'. They share an atmosphere that is essentially informational. This is called the infosphere. (Floridi 2021.) Based on this foundation Floridi develops his theory. The replacement of biocentrism with ontocenrism is an example of this. According to Floridi, non-humans can come to the centre of ethical discussion. For this reason, there needs to be a doctrine. Information ethics can serve this purpose. (Floridi, L. 2014.)

It is a fact that Information Ethics is not without problems. Floridi wanted to adopt an ontocentric stance and tried to explain different ethical problems in through his theory. The major thesis of Floridi is that he tried to explain things from a different perspective where humans are not at the center of ethical actions. It is the non-humans who are at the centre. He tries to develop a doctrine where these non-human objects can have agency, can have ethical claims. In order to summarize his position Floridi cites a letter written by Albert Einstein. Five years before his death, Einstein received a letter from a 19-year-old girl. The young girl lost her younger sister. She did not know how to cope with the grief. The young woman wished to know what the famous scientist might say to comfort her. On March 4, 1950, Einstein wrote to this young lady:

"A human being is part of the whole, called by us 'universe,' a part limited in time and space. He experiences himself, his thoughts and feelings, as something separated from the rest, a kind of optical delusion of his consciousness. This delusion is a kind of prison for us,

restricting us to our personal desires and to affection for a few persons close to us. Our task must be to free ourselves from our prison by widening our circle of compassion to embrace all humanity and the whole of nature in its beauty. Nobody is capable of achieving this completely, but the striving for such achievement is in itself a part of the liberation and a foundation for inner security". (Floridi, L. 2014.p34.)

Now, if we take up Floridi's position, then does it open up a possibility of reformulating the Functionalist's position as an information processing state and where does that lead us to? What is happening here is that, the information processing state probably might narrow down the causal mechanism. It is indeed true that we are information-processing entities of a certain sort. When we talk about information, then it can be said that it is neither mental nor physical. But there is a problem; information is also connected to the physical system. If we recognize that, then we can realize that it is connected to the way that the physical world is configured.

One can illustrate the position with a simple example. If we have an infinite set of '0' and '1', it contains very little information. We can create software and keep producing it. So from one vantage point, if you take '0's and '1's as physical marks, all it says is that you have one dimension of physical marks of a certain sorts of extremely symmetrical system. If you take the physical configuration is much less symmetrical (0-1, 1-0, 1-1, 0-0) then that kind of symmetry is not there. But it carries more information. So, less symmetric it is, the more informative it would be. That would get connected to a kind of physical system and gets formal organization (not formal, but if you take symmetry as a formal feature, then formal organization.) Now it can be said that it is a physical system with a formal feature. Now, would that create a problem for functionalism?

That way Floridi's position opens up a very interesting door for us and asks, can we reconfigure Functionalism in some way? Because, in some way we need a system like that as we want to connect to the lifeworld, we are saying that humans are moral agents, Robots are 'electronics persons' which are in some sense agents if not moral agents. If we take this position, then problems like the Sorites Paradox can be handled. Now, we can say that all of them are physical systems, all of them are organized in a certain way and all of them codify information. In order to be a moral agent, do we need a certain kind of physical system which embodies certain kinds of information and which subsequently allows a certain physical process of processing that information, as the physical world runs only through physical mechanisms?

If information is the key words and based on broadly speaking physics or natural phenomena you will have to depend upon natural processes. (Chatterjee, A. Basu, P. and Guha M. 2017.) It is clear from this discussion that, even if we are talking about Functionalism, we cannot exclude causality. We have discussed this in chapters five and six.

In addition, in the following chapter, we need to explain more thoughts about the difference between a 'Moral agent' and a 'Moral patient'. These days, when a robot interacts with a robot, what we have is a situation where a series of actions take place just like a network. Where robots, their programmer (or multiple programmer), implementer, human, moral-patient---interact with each other which gives rise to an action that has moral significance. What I mean is that, in these days an action of an 'Electronic Person' takes place in the form of a network.

We can better understand this with an example. Think of the Robot that drives a car. Multiple programmers programmed it. After that, the instructions are given in order to run it.

Commoners like us would follow those instructions and implement certain actions so that the Robot could perform the given task, i.e. drive a car. There is a group of people involved in hi-tech surveillance and can interfere if something goes wrong with the driver-robot. So they are also performing certain actions simultaneously with the robot driver. This will be illustrated in the next chapter in detail. Now, if the car met with an accident, as it happened in California a few years back, then it also affects the network of multiple actors involved in running the car along with the robot driver. But this confronts us with a question: can we say that the notion of 'actor' has also changed these days and it is not an 'actor' but 'actors' involved in a certain situation and thereby the difference between a moral agent and a moral patient has also been narrowed down? Can we call this an amalgamation of the moral divide? This question has not been addressed in any of the previous literature, not even in Information Ethics. Traditional theories of ethics are concerned with the agent. Floridi, in his Information Ethics, put the moral agent on the periphery and the patient into the centre. But the 'agent' and 'patient' distinction is still very clear in Information Ethics. But as we have seen in the previous chapter, and the example that I mentioned in previous paragraph, the notion of 'agent' or 'actor' has also changed. In this context, I will discuss the 'actor-network theory' by Bruno Latour and try to develop my position in the next chapter.

Chapter 4

Could ontocentrism be the end of the road?

We have seen in the previous chapter, in Ethics, the anthropocentric barriers that focused on human subjectivity were replaced by a new theory, namely Information Ethics. It, however, situates the objects into the centre of discussion. Moreover, the proponent of the theory, Luciano Floridi, suggested replacing biocentrism with ontocentrism.

Ontocentrism, as we have seen, suggests that there is something even more elemental than life, namely 'being' and something even more fundamental than suffering, namely 'entropy'. I have discussed this in the previous chapter.

Moreover, Floridi uses the new theory to gain a better understanding of the issues surrounding 'agency' and AI's ethical claims. It should be mentioned, ontocentric information ethics emerges as a critique of anthropocentric ethical theories.

In this context, we can remember that a few decades ago in Sociology, we saw a familiar turn. We will discuss this in the next section. The main theme in these theories, however, is that: humans along with other things (non-humans) occupy the centre of discussion.

This, indeed, confronts us with a question: could ontocentrism be the end of the road in ethical discourse? Further, can we say that if we accept the logic of ontocentrism, then artificial agents would be qualified for gaining human-like agency and thereby would replace human labour in some situations?

In this chapter, however, I will illustrate this with the example of Actor-Network Theory (ANT) proposed by Bruno Latour and observe how the notion of ANT has subsequently been expanded to incorporate AI agents.

The incorporation of AI, in actor-network, moreover, raises some legitimate concerns over its ethical implications. Furthermore, in the realm of the application of artificial agents, we have already encountered some piquant problems which show; after using AI, the rate of profit of some companies declined. I will discuss those specific cases towards the close of the chapter and try to explain these phenomena using Marxist ethics of value as a case study.

Thus, this chapter comprises of three sections:

In section I, I will discuss in a nutshell the main theme of ANT proposed by Bruno Latour.

The second section deals with, how the idea of ANT was expanded by a scholar to include Artificial agents.

The last section deals with the problems we face if we include Artificial Agents in the original scheme of Latour's ANT taking a cue from Marxist ethics.

Section i

Actor-Network Theory

Since our main focus is to find out whether artificially intelligent agents would replace human labourers in some situations or not, I will not discuss the nitty-gritty nuances of Actor-

network theory or ANT (Latour, Bruno. 2005), instead, I will focus on how Bruno Latour had originally conceived the idea of it.

Actor-network theory (ANT) is an approach that looks at ‘material’ and breaks the so-called anthropocentrism by reducing human subjectivity. Sociologists Bruno Latour, Michel Callon, and John Law developed Actor-network theory as an alternative to ‘anthropocentric’ traditional sociology. However, the theory also provides a wonderful parallel narrative on how things (or objects) interact with one another. Thus, the theory starts the discussion by reducing the emphasis on the subjectivity of the sentient at the beginning of any process.

According to this theory, objects have some kind of agency as respondents in a network's chain of connections with other objects. Furthermore, this theory states that a social world built on connections between actors, who can be either human or other than human (non-human). Latour (2005) says that, we should not restrict in advance the type of being populating the social world. Indeed, other than the human actors (non-human actors) within Actor Network Theory would be the attention of this section in order to assert how Actor Network Theory relates to AI technologies. Furthermore, a question arises: what ethical quandary will we face if we do so? The concern for social sciences, according to Latour, is how things, people, and ideas become linked and gathered into greater units. He thinks that Actor-network theory (ANT) is an escort to the process of addressing this question. To him, it is not a doctrine, but rather a theory of how to examine the social. Thus; the discussion regarding ANT will help us in conceptualizing our understanding regarding ethics of the inanimate as well.

In the introduction Latour establishes his position. He says that the book's argument is unpretentious: when social scientists add the adjective 'social' to a phenomenon, they

designate a stabilised state of affairs, a bundle of ties that may later be mobilized to account for another phenomenon. (Latour, Bruno. 2005) However, problems arise when the adjective 'social' means a type of material as if it were roughly comparable to other terms such as 'wooden', 'steely', 'biological', 'economical', 'mental', 'organisational', etc. According to Latour, the meaning of the words breaks down at that point because it designates two entirely different things: first, a movement during the process of assembling; and second, a specific type of ingredient that is supposed to differ from other materials. In ANT 'social' can't be an arbitrary starting point. It investigates the formation of the social. Here, the shift is from structure to process.

Latour, in the first part of the book, tries to explain, how to use debates about the social world meritoriously. He uses the word 'controversy'. By 'controversy' he means that the notion of the 'obvious' has been shifted. For Latour, each disagreement is a source of ambiguity. He wants to explore it. In the next five chapters, he discusses five sources that are of great relevance.

To him, the first revolves around the position of groups. Thus he asks, do they truly occur, or are they being repetitively shaped and reshaped again? Moreover, he interrogates the 'certainty' of the 'structure'. It is quite obvious that ANT takes the latter option and hence quite accomplished to show that the initial feature of the social world is this continual outlining of boundaries by persons over other persons.

Sociologists of the social believe that the most important characteristic of this world is the indisputable presence of boundaries, regardless of who is trying to track them or with what techniques. (Latour, Bruno. 2005. p 28.)

Now Latour would say that ‘organization’ is the name of a unit and to him, each organization study enhances the steadiness of this unit. Otherwise, it might be on the edge of dissolving or recovering. So there are constant mergers and acquisitions. This may baffle the young researchers for because of this volatile nature of ‘the organization’, it may seem to them that it longer exists. However, in the case of ANT, there is no such dilemma because their main thrust is not on the group but the doings of the group. It studies how it is being make and unmake. In doing so, their modus operandi are to follow an actor and note the name used for the position they attain.

The second source of uncertainty, according to Latour, is regarding agency. When perceiving an action, the fundamental question is who or what is acting. The actor-network suggests that an observer’s finding as an ‘actor’ may be a whole network. However, the researchers discover inconsistencies in the accounts given by those who appear to be the ‘actor’. As a result, in order to investigate this source of uncertainty, the researchers select only those actor accounts that can be incorporated into a theory with the perfect clan.

To accomplish this, the researchers erase the symbol of the multiplicity of agencies that may be of great interest to them. There may appear to be some ambiguity regarding how the agency should be described.

This is where the third source of uncertainty, that objects can be seen as having agency, should come into play. We can see here that the definition of ‘social; is broadened from ‘human only’ to ‘all actants that can be associated’. (Latour, Bruno. 2005. p 28.)

This extension of definition, indeed, is nothing new to fiction or everyday life and its need is quite obvious. In our everyday life, old companions like dogs and horses have been replaced by computers and iPods. Social scientists, however, are very meticulous in

differentiating between humans and non-humans. They regard humans as their concern and non-humans as the concern of other disciplines. This emerges from the fear of social scientists of losing their domain.

In cooperative projects run by sociologists and economists, this self-definition of organization studies is often revealed. When the word 'money' appears on the scene, sociologists lose interest. Economists who live in space between social and natural sciences are expected to come forward for inspection. Thus the objects stabilize. This is the special role objects play in associations. Money plays a role in organisations and it has an impact on people.

Contracts are chosen to write, obituaries are engraved in stone, and technical instructions are constructed into the equipment to guide users in a specific manner. Rather than society, we live in a team in which there are groups of humans and nonhumans. Many perplexed critics claim that ANT ignores so-called 'power relations'. However, rather than ignoring it, ANT explains it.

The wealth of the people, on the other hand, is inextricably linked to the ownership of capital. According to Latour, to say that people are wealthy because they have capital—- is a tautology. As a result, ANT is left wondering on the question 'How did they form the bond?' This distinguishes 'matter of fact' from a 'matter of opinion', as well as a 'cause for concern'. This relates to the fourth source of uncertainty: the state of facts. The difference between 'matter of fact' and 'matter of concern' is in the making. ANT wants to study this process of 'making'. Concern has the power to turn suppositions into facts and politics as the power to send them into oblivion. In discussing the source of uncertainty Latour explains why ANT

abandoned the level of 'social constructivism'. ANT, on the other hand, interpret 'social' as 'not individual' and 'construction' as 'not creation'. In broad terms, however, this reads as 'human contrivance'.

Next comes the fifth source of uncertainty. It is a reassurance to literary inventiveness and a cautionary against using words that may be shrewd but not sincere and at the same time, it is also cautionary against pride. (Latour, Bruno. 2005. p 127.)

However, redistribution of the local is necessary. Fetishism gives much more credence to non-human contributors to society. Likewise, ANT also focuses on non-human contributors to society. According to traditional sociology, there is also a social 'context' in which non-social functions are performed. It is a distinct realm of reality. Ordinary agents are always embedded in the social world that surrounds them; they can be 'informants' about this world at best, and blind to its existence at worst. (Latour, Bruno, p. 176) According to Latour, 'social' is more concrete to traditional sociologists than it is. He believes that social forces do not exist and that the only thing social scientists can do is describe how different actors interact within ANT. He argues that social forces do not arise and that all social scientists can do is explain how various actors communicate within Actor Network Theory.

According to Latour, there is no material social dimension as other sociologists have envisaged, so the associations between actors must be at the core of discussion. According to ANT, there is nothing unique to social order; there is no social dimension, no social context, no distinct domain of reality to which the label social or society could be applied. It propagates the notion that actors are never merely informants because they are never embedded in a social context. The social, rather than being a constant realm in which everything occurs, is made up of connections between actors. To put it another way, the social is always changing.

As a result, Latour believes that the social exists since these connections exist. When one connection fails, it must be restored with a new one, else the term ‘social’ will become obsolete. Latour's classic example of ANT is the gunman. It depicts the interactions between human and non-human actors that shape the social. He writes, when a person and a weapon like gun are linked together, a new entity is formed. We can call this new entity as the gunman. It is impossible for a human being to shoot others on his own. However, one cannot say that gun is the source of all evil. Guns that can fire on their own are extremely rare. What connects man and the gun resulting in the creation of a gunman? This is the link that ANT wants researchers to focus on. According to this theorist, a gunman is distinct from both a man and a gun in that a gunman can shoot someone, whereas neither a man nor a gun can. (Latour, Bruno. 1999.)

A man and a gun make up the gunman: a human actor and a non-human actor. The inclusion of non-humans as possible actors is a key component of ANT. Actors in ANT include objects, ideas, processes, corporations, institutions, and people.

However, the social relationship is viewed differently in traditional sociology and actor-network theory. The constant realm of the social is the focal point of traditional sociology. On the other hand, according to Actor-Network-Theory, the social is an ephemeral realm that exists when connections exist. The social is formed by the interactions and associations of human and non-human actors. As a result, the presence of ‘social’ is linked to the actors. According to Latour, if an agency is mentioned by someone, then she must provide an account of her actions. As a result, a connection between actors necessitates some form of action.

Thus, for the ‘social’ to exist within ANT, there must be connections made among both actors. Those connections are impossible to establish unless the actors demonstrate agency through action. Consider the gunman's transformation: in order for the man and the gun to become the gunman, the man must mentally decide to pick up the gun. The gun must then be physically picked up by him. Each of these steps requires action. The gun must then be fired in order to complete the conversion of man and gun into the gunman, resulting in yet another instance of action. Each of the above actions has resulted in the gunman.

Actor-Network Theory generally doesn't really try to explain why a network remains; rather, it is more engaged in the connectivity of actor-networks, how they are created, how they might collapse apart, and so on.

On the other hand, ANT, integrates what is recognized as ‘a principle of generalized symmetry’ (Bruno Latour, 1996.) It states that what is sentient and what is non-human (e.g., artefacts, organizational structures) must be incorporated into the same conceptual framework and given equal agency. According to Latour's Actor Network Theory (ANT), any structure we confront can be confronted most successfully if all of the components natural, technological, or sentient viewed as interconnected and active members of the system. Furthermore, humans, technology, and natural factors such as sunlight, air flow, heat, and so on all play an equal share in the Actor-Network Theory system.

Moreover, in ANT, every situation is a network. By network, it refers to interconnected elements which affect each other. This network is composed of ‘actants’. By ‘actants’ Latour refers to components of the network, that play a role. This network also has ‘connections’. The ‘connections’ are how all the parts interact.

It is not necessary for an object to be an actor in this sense to have contentful mental states, but rather to be able to carry out acts as a type of behaviour describable under some specific intent. As a result, actors can have a variety of interactions and relationships. In particular, some actors can reshape other actors (these transformations are sometimes called translations). A 'network' is an accumulation of actors in which the steady actors have relations and translations that determine the actors' positions and operations within the network.

When the network is settled, it means that no other actors or relationships will be able to join it. It invites the possibility of scientific knowledge gathering as a result of translations within the network. Scientific belief, theory or facts emanates from the standpoint of the actor-network theory that places the actors in a firm network.

Section 2

Does ANT fit in incorporating Artificial Agents?

This section of my thesis owes much to Ms Mallory Reed (2018) who, in her dissertation argues that to accommodate artificial agents, ANT should be expanded.

She begins her dissertation, however, with the example of Robotic Honey Bees from the Netflix science fiction anthology series *Black Mirror*. She goes on to say that the episode 'Hated in the Nation' depicts a hypothetical situation in which technology fails. In this emerging situation, robotic honey bees can operate without human intervention. Furthermore, she uses *Black Mirror* as an example to demonstrate how artificial intelligence technology can go beyond the definition of an 'object'. Reed contends that, while Actor-Network Theory appears to be suitable for artificial intelligence due to the action present, it does not fully account for the degree of agency seen in technological developments such as robotic honey

bees. As a result, Latour's Actor-Network Theory needs to be expanded to accommodate artificially intelligent technologies. Even though there are different kinds of actors in ANT, an actor's agency is limited to only certain types of actors within ANT.

According to Reed (2018), with the advancement of artificial intelligence and the expansion of its everyday applications in recent years, non-human smart actors are increasingly becoming a part of society. She cites smart home systems, self-driving cars, chatbots, intelligent public displays, smartwatches, Alexa, and other examples. Furthermore, machine learning methods and neural network models form the foundation of today's artificially intelligent agents. At this point, one might wonder why AI agents are held responsible for ethically complex behaviour or held accountable for people's employment.

To proceed further, we need to clarify certain ideas regarding Latour's position. Hence, in this context, I will discuss in a nutshell the basic idea of Peter-Paul Verbeek's (2005) book *What Things do*. On reading the book one may find that the philosophy of technology is the main topic Verbeek is concerned about. In this book, Verbeek discusses the work of Martin Heidegger, Don Ihde, and Bruno Latour. For our current purpose, I will not review the first part of this book, where Verbeek has engaged himself with Heidegger's work. Instead, I will follow the argument given in the second section of the book where Verbeek engaged himself with Don Ihde and Bruno Latour's work.

It should be noted that Don Ihde, as Verbeek believes, develops the postphenomenological approach. Ihde referred to this as a type of phenomenology without transcendental pretensions. Verbeek, on the other hand, describes it as a relational ontology in which sentient and things, subjects and objects, are mutually constitutive. Thus, Verbeek claims that Ihde's approach can preserve classical philosophy of technology's existential

concerns while also incorporating technology studies' concept of the co-construction of both society and technology.

Verbeek casts a shadow over Ihde when describing the intentionality of technologies in terms of how they influence perception and action. The technical element which shapes both subject and object are included in the 'I-world relation'. Thus, Verbeek summarised Ihde's account of this relationship, which can take several different forms. Moreover, it is analyzable at the 'micro-level of perception and the macro-level of culture'. He agrees with Ihde that technological innovations are only what they are within a sociological perspective and thus do not have any essence 'in themselves'. Indeed, what it means to be human is ascertained by context. Thus, technologies have a contribution to shaping the cultural structure which forms individual's life. Verbeek's book culminates in an attempt at a synthesis of this postphenomenological approach and Bruno Latour's actor-network theory. (Verbeek, Peter-Paul. 2005.)

I would like to mentioned that Latour's hypothesis seems to be anti-essentialist and relationalist, making it consistent with Ihde's notion of phenomenology. Things exist only in their framework, according to Latour in this scenario the framework of the network into which they are integrated. His human-non-human symmetry postulate has the negative goal of preventing any recourse to pre-existing essences that would decide the network instead of being co-constructed by it.

Thus, Verbeek describes Latour's mediation principle, that clarifies how networks are shaped through the programmes of actors and how those programmes are translated into the actions of hybrids moulded by individuals and their apparatuses. (Verbeek, Peter-Paul. 2005) Ihde, however, emphasises the structure of experience while Latour on the action. Verbeek

thinks, their methods seem harmonizing. According to Verbeek, the contrast of subject and object is overcome in both cases.

There is, moreover, a difference in emphasis. Ihde's main area of interest is the formation of the relationship between subject and object. Latour, on the other hand, believes that the network is what brings subjects and objects to the forefront. Thus the actor-network theory deals with the making process. It studies the postphenomenology of experience as well as the formation of subjects and objects. Verbeek thinks the two facets are linked but not identical.

Although both types of actors are included in Latour's theory, there is still a distinction between how human and non-human actors can act. As a result, Reed believes that Actor-Network theory formulated by Latour could be extended. According to Reed, it is because of two components of Latour's theory that preclude the inclusion of artificially intelligent agents within ANT.

For starters, as Reed has pointed out, Latour's theory categorises actors as non-social. Furthermore, Latour differentiates among actors who have internal inertia and those that require external inertia to perform. Reed points out that Latour ignores objects which can influence things autonomously of other actors, rather choosing to focus on actors as non-social entities. Reed urges us to take the example of the gun in the gunman case. Latour holds that, the social appears to exist only when associations form between actors. But the actors themselves are not social. According to Latour, the social is only visible through the remnants it leaves when a fresh association is formed between components that are not social. The connection between the actors, according to Latour, is more crucial than the actors. As a consequence, the actors are not social, but their associations create the social. Even though

the actors are crucial, it is the associations that are more fundamental, not the actors themselves. (Reed. 2018)

Conversely, Reed cites the example of *Black Mirror.* In this fiction, robotic honey bees save pollinated plants' future. At the same time, they cause tremendous harm to the human population. It should be remembered, to pollinate plants is the main purpose for which the robotic honey bees were created. It was designed with the goal of increasing pollination capacity in mind. Moreover, Reed's observation is that, according to Latour's scheme of things, we cannot say the bees themselves are social. In addition to the humans who created them and the plants they pollinate, bees are actors. As a result, the social is formed by the interactions of these actors.

On the contrary, Reed shows that, in the case of *Black Mirror,* the individual bees themselves are social. In this case, the bees act against their initial intended purpose. They stop pollinating plants after a certain period of time. Instead, they directly target humans and kill them by going deep into their brains. They halted pollination. Concurrently, they severed ties with their former masters. Those connections and associations between plants and humans appear to be irrelevant at this point in time. Instead, only the robotic honey bees remain. They are the only actors who have left the network where they previously worked. Rather, just one connection they make is the act of killing humans. The connection between the actors, according to Latour, is more essential than the actors themselves. As a consequence, the actors are not social, but their associations create the social. Although the actors are important, it is the associations that are more fundamental, not the actors themselves. (Latour. Bruno. 1999.)

This becomes even more troublesome when we take into account artificially intelligent agents, such as robotic honey bees, because these bees become the star of the show when they

turn on their inventors and begin attacking and executing humans. According to Latour, determining the origin of action is impossible. The action is 'dislocated', and it is composed of a complex web of actors collaborating to create a single action. However, because the bees strayed from their original mission, they left the web and are now acting independently. The second issue that the robotic honey bees face with ANT is the classification of some actors as intermediaries and others as mediators. Intermediaries are actors who conveyed meaning or force without modification. In other words, they are carriers of meaning or action that do not intervene or interact with it. They make no changes to the information they receive and serve only as a link in the chain.

On the contrary, the second type of actor is a 'mediator', who 'transforms, translates, distorts, and modifies the meaning or the elements they are supposed to carry'. (Latour. Bruno. 1999.) Mediators, as opposed to intermediaries, start engaging with meaning and action, which they may or may not alter. They are unpredictable since an outsider cannot predict what will happen after interacting with a mediator because the mediator has the ability to change the information.

According to Latour, both objects and humans can act as mediators or intermediaries. He does, however, claim that there is a distinction between how humans and objects switch between being intermediaries and mediators. As per Latour, it is difficult to stop humans from becoming mediators again. In other words, humans become mediators and remain mediators, whereas objects become mediators for a short time and then transform into intermediaries.

In terms of 'inertia possession', Latour distinguishes between humans and objects further. The ability of an actor to remain a mediator is related to the presence of inertia. Latour

distinguishes two types of groups. One that is endowed with some inertia and those required to be constantly maintained by some group-making effort. (Latour, 1999, p. 35.) According to Latour, humans have inertia while objects cannot. It appears to be conceptually important, at least how Reed chooses to deconstruct Latour. Humans can easily become mediators due to this inertia, whereas objects struggle to become or remain mediators.

However, one reason humans continue to be mediators is their capacity to speak. Humans have the power of speech. Since humans can express themselves, they cannot be denied the role of mediator. To make humans intermediator, it is urgent to stop them from speaking which is not possible. It should be noted, what a human is going to say is known only to her. Hence one cannot make any prophecy about her speech. She will be able to affect the world through her freedom to talk.

Objects, on the other hand, do not have this capability. These days artificially intelligent agents are on the path of achieving general intelligence and superintelligence. One of the primary goals of this is to achieve the ability to communicate like humans. According to Reed, new-age technology can have the power of speech just like its human counterpart. In this way, they can become mediators like humans.

At this juncture, we can think of the Robots in Isaac Asimov's (1950) science fiction *I, Robot.* Asimov's various robots can think and speak. For example, there is a robot called Cutie in the chapter 'Reason'. Asimov's Cutie could speak. Cutie wonders about its existence as well as the existence of humans at various points in time. Not only that, but convincing Cutie that it was created by humans is difficult. Hence Cutie does not pay hid to the humans. It might work for her. As a result, it works on the space shuttle. Moreover, Cutie can reason and communicate. Hence, it decides not to obey the humans' command. So, one can say that Cutie

is a mediator. Its creator cannot make Cutie become an intermediary. The reason is Cutie does not pay heed to the humans who created it.

According to Latour's Theory, objects are not decent at remaining mediators. Reed, on the other hand, believes that even when humans try to intervene and make them intermediaries, Isaac Asimov's robots and the Black Mirror's robotic honey bees act as mediators. Cutie, Asimov's character, would ignore humans and take over their jobs on the spacecraft. Cutie acts in this manner since it believes it can outclass human beings.

An almost similar situation has been raised in *Black Mirror*. In that series, humans attempted to rewire the programme for the bees. They want bees to return to their initial assignment. However, they were unable to do so. As a result, there were casualties. Thus, robotic honey bees in Black Mirror are objects, according to Reed. She believes they will continue to be mediators in ways that Latour does not anticipate.

Section iii

Problems of extension taking a cue from Marxist ethics

The question that confronts us is: can we extend Latour's original theory to incorporate AI? To proceed further, we need to discuss some of the research that has been conducted in the domain of AI on labour productivity. (Damioli, G. Van Roy, V. & Vertesy, D. 2021.)

However, some recent researches show that AI will increase the profitability of a company. For instance, Accenture conducted a study labelled 'How AI Boosts Industry Profits and Innovation'. (Purdy, Mark and Daugherty, Paul. 2017.). Furthermore, the economic growth rates of sixteen industries were compared in this study. Following that, it forecasted the impact of AI on global economic growth through 2035. In the study, the GVA (Gross Value Added) technique was employed as a rough estimation of GDP (Gross Domestic

Product). According to the study, the more AI is integrated into economic processes, the greater the potential for economic growth. The research also claimed that AI can increase economic growth rates by a weighted average of 1.7% across all industries through 2035. (Purdy, Mark and Daugherty, Paul. 2017.)

According to these researches, artificial intelligence combines labour and capital factors. As a consequence, labour and capital are both engaged, creating a labour-capital hybrid. (Purdy, Mark, and Paul Daugherty. 2017.) The researchers demonstrate that this labour-capital hybrid has the potential to boost any company's corporate profitability. Furthermore, these studies proposed that integrated artificial intelligence technologies can detect, understand, respond, learn, adjust, and develop. As a result, they have significant advantages over human labour and traditional capital. The study also identified the role of AI in profit and innovation. The research shows, in three ways AI increase profit. I will discuss these in short.

The first is intelligent automation. It promotes the application of instruments, methods, and techniques to automate processes and, as a result, removes the requirement for labour. The second is labour and capital expansion. The main philosophy of labour and capital augmentation, on the other hand, is the application of techniques that work alongside human workers, complementing their work and improving their skills. The third factor is the spread of innovation. The central focus of innovation diffusion or spread of innovation is that artificial intelligence inventions in one industry disperse to and impact others. (Purdy, Mark and Daugherty, Paul. 2017.)

Without a doubt, AI innovation benefits industries. However, we also come across some contradictory research. According to some studies, with the advent of smart technology, current profitability is decreasing. Erik Brynjolfsson et al. (2017) argued in an article that the recent patterns in aggregate productivity growth highlight an apparent contradiction. According to the report, there are some examples of potentially transformative new technologies that could greatly increase productivity and economic welfare.

New progress in the performance of artificial intelligence (AI) is one of the most noticeable initial tangible indications of these technologies' assurance. At the same time, evaluating productivity growth has slowed noticeably in the last decade. This slowdown is significant, cutting productivity growth in the decade preceding the slowdown by half or more. As a consequence, we have seemed to be confronted with a Redux of Solow's (1987) paradox. It implies that revolutionary new technologies can be found everywhere except in productivity statistics. (Purdy, Mark and Daugherty, Paul. 2017.) As a result, the document sees the slowing of contemporary productivity as a conundrum.

Moreover, other researchers also echoed this concern. They refer that many digital systems and new technologies, such as the Internet of Things and cloud computing, have emerged in the last decade. Economic growth in many of these advanced nations, such as the United States, is eventually gradually decreasing. Isn't that a paradox? (Lyu, Yitian & Zhang, Chenrui. 2019)

It is true that technological know-how has progressed significantly. Cloud computing and new service-based business models can be helpful these days. Recent improvements in AI are largely based on machine learning. (Sakovich, Natallia. 2020.)

In addition to this, we must mention that the third wave of cognitive science has arrived. In the new era, machine learning represents a sea change from the first wave of computerization i.e. symbol manipulation. Though machine learning is advancing at a good speed, still it is not up to the level of professional human performance yet. Let us use an example to demonstrate this. Using convolutional neural net sequence prediction techniques, Facebook's AI research team recently improved on the best machine language translation algorithms available. To develop this, deep learning methods are combined with reinforcement learning techniques. In this manner, new techniques for generating control are employed. The main philosophy behind this is to train autonomous agents in such a way that they could act all by themselves in a given environment. Though many of these developments are in the embryonic stage, advances in this field are impressive. (Hazelwood, Kim. 2018.)

Thus, I have discussed some of the notable technological milestones. The advancement of technology is intimately linked to the economic landscape. Moreover, it has the potential to generate new business opportunities. Additionally, it helps in saving money. I can demonstrate this with a few examples. A deep neural network-based system was run against 21 board-certified skin specialists and paired their results in detecting skin cancer. Besides, take the example of Facebook. It uses neural networks for over 4.5 billion translations each day. This, indeed, can be marked as a momentous departure from the early days of computing. We know that most computer programs were created by codifying human knowledge. Initially, they were designed to map inputs to outputs as defined by the programmers resulting in a rule-governed process. But on the contrary, these days' machine learning systems use neural networks, that is to say, general algorithms to work out relevant mappings on their own. However, they toil with a gigantic set of data. With the help of

machine learning methods machines have made impressive gains in perception and cognition. These two are considered to be the essential skills for most types of human work. It can be said that these technologies are trained on historical data to uncover patterns. They learn from examples. They learn to predict and classify generalized future outcomes for decision-making. They are grounded on large data sets, which are normally referred to as Big Data. Moreover, they analyse Big Data at speeds and scales that exceed the human brain's ability to analyse. (Leonelli, Sabina. 2020.)

The technological automation discussed above have enormous potential. However, according to these studies, there is little evidence that they have impacted aggregate productivity. According to the researchers, labour productivity growth rates fell in the mid-2000s and have since stayed low. In the United States, for example, cumulative labour productivity growth averaged only 1.3% per year from 2005 to 2016. Between 1995 and 2004, less than half of the annual growth rate of 2.8% was maintained. The study found that 28 of the 29 other countries for which the OECD has compiled productivity growth data followed a similar pattern. (Syverson, Chad. 2016.)

Furthermore, annual labour productivity growth rates in these countries were 2.3 % from 1995 to 2004. It fell to 1.1% between 2005 and 2016. Moreover, real median income has been stagnant since the late 1990s, and non-economic measures of well-being, such as life expectancy, have declined for some groups. (Case, Anne and Angus, Deaton. 2017).

What caused this to occur? There are four major explanations given in this study: 1) false expectations, 2) miscalculation, 3) concentrated distribution and rent dissipation, and 4) implementation and restructuring lags. (Brynjolfsson, Erik et al. 2017)

The main theme of False hope is like this: modern technologies appear to have lofty aspirations for us. But the situation demonstrated beyond a shadow of a doubt that our expectations were excessive. These innovations are extremely successful effective on a small scale and in a particular company or industry. Their total impact, however, is negligible. This is, without a doubt, a logical contradiction that will hopefully be settled in the future. I hope in one day, the optimist will see the truth.

The crux of the matter of mismanagement is as follows. In terms of productivity, we discover a disparity between output and productivity. There is no wonder that we profit from the fresh opportunity opened by modern technologies. These advantages, however, are not sufficiently measured. So there's reason to be optimistic. But our past, which we are well aware of, is a major hurdle to our elation. This is referred to as the 'mismanagement hypothesis'. Analysts claim there has been a continuous decrease over the last decade. This explanation, however, is inaccurate. This is evident in a number of works and contains an important point. Smartphones, online social networks, and downloadable media are examples of low-cost technologies. As a result, their GDP contribution is small. They are, however, extremely useful. According to recent research, this presumption is unable to determine the cause of the slowing economy. (Brynjolfsson, Erik et al. 2017.)

The third option is 'concentrated distribution and rent dissipation'. We appear to benefit from emerging innovations, but we must remember that these advantages are dispersed carefully as well as through dissipative attempts. For this reason, it's natural to believe that technological advancements benefit only a small portion of the economy. On medium work, the effect is, to put it mildly, subpar. For example, online advertisement targeting and pricing,

automated trading of financial instruments, and the two most profitable applications of AI generate very little or no profit. As a consequence, it is natural to believe that these avatars of technology benefit only a small part of the economy. However, because of the inherent nature of these emerging innovations, we are witnessing a massive rush for these technologies, comparable to the mad dash for Mackenna's Gold! There are two clusters. One is the beneficiary group. Another is the beneficiary but does not want to share the benefit. (Brynjolfsson, Erik et al. 2017.)

The next one is implementation and restructuring delays. There are compelling reasons to think that future productivity growth will be hampered. We've reached this point as a result of the recent trend of slower productivity growth. However, the situation is not as precarious as it appears. Regardless, future growth may be promising. It's possible that things will change. Total factor productivity growth is occurring concurrently with overall output growth. However, changes in labour and capital input are possible. Furthermore, this factor cannot account for such a scenario. Forecasting based on past experiences is impossible. (Brynjolfsson, Erik et al. 2017.)

Based on these studies, I can conclude that it may take time for new technology to demonstrate its worth. Simultaneously, an investment in the proper operation of the new technology is crucial. Otherwise, these emerging innovations will not produce the desired results. As a result, we must first acknowledge the importance of technological development. Without this modern innovation, it will be impossible to travel a long and complex path. However, we should remember that along with technological advancement many other things are required. (Brynjolfsson, Erik et al. 2017.)

In the meantime, we found some news indicating that profitability is declining. Consider the following two recent events. This will assist us in comprehending our hypothesis. The first of these was published in the *Guardian* (2020). It is a story about a faux pas by a Microsoft's robot editor.

Microsoft's decision to replace human journalists with robots backfired when the company's machine intelligence software depicted a racist news story with a picture of the wrong mixed-race Little Mix member. A story about singer Jade Thirlwall's reflections on racism was illustrated with a picture of fellow band member Leigh-Anne Pinnock, less than a week after the *Guardian* revealed plans to replace MSN.com's human editors with Microsoft's artificially intelligent agents. Thirlwall, who was in London for a Black Lives Matter protest, criticised MSN, saying she was tired of 'ignorant' media making such mistakes. Thirlwall, according to sources, had no clue that the picture was selected by an artificially intelligent agent, which is already in charge of editing parts of the news website, that has millions of readers worldwide. Thirlwall, in anger, took Instagram to protest. Mentioning MSN she wrote that, if they are going to just copy and paste articles from other news sources, then they will have to make sure to use an image of the correct mixed-race member of the group. In addition, instead of trying to conduct actual journalism, Microsoft hires human editorial staff to pick, modify, and repackage news stories from news outlets or agencies. The articles are then hosted on Microsoft's website, and advertising revenue is split between the original publishers and Microsoft. In the midst of pandemic, Microsoft decided to fire hundreds of journalists and completely replace them with robots.

Concerning the blunder, a Microsoft spokesman stated that as soon as they became aware of the problem, they took immediate action to resolve it and replaced the incorrect

image. One Microsoft employee expressed concern about the reputation of the company's AI product, saying, with all of the anti-racism protests taking place presently, this is not the time to commit mistakes. (Waterson, Jim. 2020.)

There were numerous reports published during the covid pandemic that mentioned that after the pandemic, many companies would advocate for increased use of artificial intelligence. Telstra, for example, intends to reduce customer service calls by two-thirds by 2022. The firm intends to expand its use of Artificially Intelligent Agents. (Waterson, Jim. 2020.)

Google, on the other hand, cautioned that the rise of artificially intelligent agents may lead to trouble for the firm. However, Google has warned that inventions like artificial intelligence could harm its business, leading to penalty fees and ethical problems. According to *The Telegraph*, Alphabet, the search engine's majority shareholder, warns in its recent data that demand for its products and services may suffer as a result of worries about the ethics and legality of machine learning. New products and services, including those incorporating or utilising artificial intelligence and machine learning, may introduce new or aggravate ethical, technological, legal, and other challenges, potentially harming products. (By a staff reporter. 2019)

Following Marx's analysis of the ‘tendency of the rate of profit to fall’, I would like to offer an alternative explanation to this phenomenon. I will not delve deeply into Marx's Labour Theory of Value, but will instead attempt to cite a possible explanation for the aforementioned hypothesis—-Why does the rate of profit of some companies tend to fall? We only intend to use Marx's theory as a case study. Before we proceed, it should be noted that

there is now critical scholarship attempting to find parallels between Actor-Network Theory and Marxism. (Sayes, Edwin. 2017)

It should be noted that Castree (2002) and Gareau (2005) are two Marxist political ecologists. They are the proponents of 'Marxism-Actor Network Theory Synthetic Approach'. Latour (1999) proposes a few 'tests' for an account/analysis to qualify the rigour of the Actor-Network-Theory in *Reassembling the Social*. In one of the tests, non-humans should be acknowledged as actors. In this manner, Karl Marx's texts are rife with these kind of non-human actors, particularly if the bare minimum requirements for an actor is an object's ability to 'make a difference'. As per Sayes (2014), it is hard to find a text of Marx in which nonhumans do not end up making distinct change, ranging from the significance of money to the strike-breaking capacity of equipment, to the more basic role that equipments perform in modifying the work flow, and the role of technology in the manufacturing process. In this spirit, I draw your attention to Chapter 13 of *Capital* Volume 3 (Marx, Karl. 1867 [1981]), where Marx discusses a long-term tendency for profit rates to fall with capitalist development. Marx constructs an intense dialogue between human and non-human agents. For two reasons, this invocation is critical to our quest. For starters, it enables us to make a 'qualitative' distinction between human and non-human agencies, their effects, and their moral and political implications within a network of translation, which is frequently blurred in Latour's framework.

Second, it successfully describes how an AI agent or any constant capital (i.e. plant and building, equipment, physical infrastructures of production, raw materials, auxiliary materials, energy, and so on) could not really substitute human agency, whose distinctive and innovative expression lies in possessing labour power (variable capital), that is the sole source

of value. According to Marx, constant capital can only transfer value to a new commodity that has historically created constant capital (such as AI in our case). According to Marx value can be defined as ‘socially necessary labour time congealed in constant capital’. Artificially intelligent agent, according to Marx, is dead labour. It aids the owner of the means of production (the capitalist) in lowering the value of labour. It can be the source of increased relative surplus value rather than the value per se. Since I am going to explain how a commodity's or company's profit rate declines with the application of technology (in our case, artificially intelligent agents), I should first create some background before discussing this theory. The fear that technological advancements will usher in a future where machines will take our jobs is not new. It dates back to the 19th-century Luddite rebellion and rose to prominence with the introduction of new technological advancements. (Karakilic, Emrah. 2022)

Nevertheless, two dominant schools of thought see a future devoid of jobs for Labourers as a distinct possibility. According to the first group, we are on the verge of a workless upcoming years since new technologies with human-like skills have the ability to eliminate our jobs over time. (Karakilic, Emrah. 2022)

On the other hand, the scholars of the another camp seem to accept that in the not-too-distant future, human labour will be replaced by new technologies and machines. They believe that this will be the termination of capitalism. (Karakilic, Emrah. 2022)

Moreover, two very different schools of thought agree that things have changed. We can assume that in the recent past, the deployment of artificially intelligent devices has speeded up this and has a bearing on productivity. (Karakilic, Emrah. 2022.)

Furthermore, the ‘Grundrisse’ section of *The Fragment on Machines* aids in understanding the relationship between technology (in this case, Artificially Intelligent agent) and capital. Marx also observed that capitalism, as the dominant form of production, ‘works toward its own dissolution’. (Karakilic, Emrah. 2022.)

In addition, he believes that this dissolution will be steered by technological advancements in manufacturing and will occur in the not-too-distant future. (Karakilic, Emrah. 2022.)

Marx observes that with the continuous development of machinery, the production process has ceased to be a labour process in the sense of a process dominated by labour as its governing unity. (Marx, K.1845.) Labor, according to him, has been absorbed under the entire procedure of machinery, and its unification has started to face workers as a simple link of the scheme and their individual, unimportant deeds as a mighty organism. (Marx, K.1845.)

One of the most contentious arguments in the *Fragment* is that general intelligence is not an attribute of labour. It is an attribute of a fixed capital asset, specifically machinery and organisational systems. (Marx, K.1845.) According to Marx, “the general productive forces of the social brains are thus absorbed into capital, as opposed to labour, and thus appear as an attribute of capital, and more specifically of fixed capital.” (Marx, K.1845.) Because the production process will rely progressively less on variable capital as opposed to the general intellect crystallised in fixed capital, capital will inevitably reduce labour time to a bare minimum and, as a result, the worker will step to the side of the production process rather than being its chief actor. (Marx, K.1845.) Such human-like devices are already undermining not only manual labour but also the entire category of work that includes knowledge work.

However, my thesis argues that total knowledge work cannot be acquired by artificially intelligent machines in two respects. Firstly, in this chapter, I will show this taking a cue from Marxist ethics.

Second, in the following chapter, I will show that artificially intelligent agents will find it difficult to participate in the lifeworld like natural agents, which is the source of value production, and thus become successful labour. In this chapter, I've given some examples of how the implementation of artificially intelligent agents tends to reduce profit rates.

To go into the details, let us clarify in brief what Marx understood as labour. Marx defines labour as the capacity to comprehend a problem and implement a remedy. It includes human muscle power, intellectual power and creativity. According to Marx, "Labour is, in the first place, a process in which both man and Nature participate, and in which man of his own accord starts, regulates, and controls the material re-actions between himself and Nature. He opposes himself to Nature as one of her own forces, setting in motion arms and legs, head and hands, the natural forces of his body, in order to appropriate Nature's productions in a form adapted to his own wants. By thus acting on the external world and changing it, he at the same time changes his own nature. He develops his slumbering powers and compels them to act in obedience to his sway. We are not now dealing with those primitive instinctive forms of labour that remind us of the mere animal. An immeasurable interval of time separates the state of things in which a man brings his labour-power to market for sale as a commodity, from that state in which human labour was still in its first instinctive stage. We pre-suppose labour in a form that stamps it as exclusively human." (Marx, K.1845)[11]

[11] Quoted from *Capital* Volume One, Part III: The Production of Absolute Surplus-Value, Chapter Seven: The Labour-Process and the Process of Producing Surplus-Value, see https://www.marxists.org/archive/marx/works/1867-c1/ch07.htm

To him, the fundamental factors of the labour process are as follows:

a) work itself

b) the subject of work

c) instruments.

Thus, in the labour process, man's action, with the aid of labour-tools, causes a change in the material which was intended from the beginning. The process is lost in the creation, which is a use-value, Nature's material is modified to man's needs through a change in form. (Marx, K.1845.)

What is the value, according to Marx? He believes that every product comprises a small portion of the socially required labour time. The degree of value is determined by the amount of socially necessary labour time needed to make a specific product with an assumed use value. Marx claims that when we exchange our various products as values, we also exchange the various types of human labour exerted on them. We are not aware of this, but we do it anyway. (Marx, Karl. 1867 [1887].) So, value sums to the socially necessary labour time embodied in the production of a commodity. (Marx, Karl. 1867 [1887].)

Marx holds that, the basic species-specific feature of human labour is the addition of ‘value’ to what a person makes. Conversely, many AI scientists are talking about a device a very special mechanical standard that claims to substitute human autonomy, intellect, creativity---a machine that devices problems and innovates solutions in some situations.

On the contrary, Marx defined technology/machinery as ‘dead labour’. It assists the owner of the means of production (the capitalist) in lowering the value of labour. This could be the source of increased relative surplus value rather than the value itself. (Marx, Karl. 1867 [1887].)

Marx tells us, the value of a commodity is proportional to the socially necessary labour expended in its making. Marx derives the idea of exploitation from surplus value, i.e. value derived from the uncompensated labour of the workers, 'over and above' the value of their labour-power. The capitalist appropriates this value without compensation and uses a portion of the same to expand the cycle of capital accumulation. Marx further tells us that surplus value is particular to the capitalist mode of production where the surplus product takes the form of surplus value. How to calculate the rate of profit?

Suppose:

C=total capital

c=constant capital (means of production, machinery, infrastructure, etc)

v=variable capital or labour wages

C=c+v

S=surplus-value

S'= rate of surplus-value (surplus-value means the excess of value produced by the workers over and above their wage: unpaid labour that goes to the capitalist for the expanded reproduction of capital>unpaid labour is the source of profit)

P'=rate of profit

Then:

c/v refers to 'organic composition of capital[12]

S'=S/v and hence S=S'v

P'=S/C =S'v/(c+v)

[12] The "organic composition of capital" is the ratio of the value of the materials and fixed costs (constant capital) embodied in production of a commodity to the value of the labour-power (variable capital) used in making it. Marx, Karl, 1867 (1887), ibid.

P’:S’::v:C

Since C>v hence S’>P’

The difference between S’ and P’ is measured by c/v

P’=S/C=S/(v+c)

Now, by dividing both numerator and denominator by v, we get the following equation (Morishima 1973):

P’=S/(v+c)=(S/v)/(1+c/v).

The calculation shows that as relative investment (c/v in the denominator) increases, the rate of profit tends to fall, assuming a constant rate of surplus-value/rate of exploitation (S/v in the numerator). (Bandyopadhyay, Ritajyoti. 2020)[13]

Thus, Marx concludes: “The progressive tendency for the general rate of profit to fall is thus simply the expression, peculiar to the capitalist mode of production, of the progressive development of the social productivity of labour. This does not mean that the rate of profit may not fall temporarily for other reasons as well, but it does prove that it is a self-evident necessity, deriving from the nature of the capitalist mode of production itself, that as it advances the general average rate of surplus-value must be expressed in a falling general rate of profit. Since the mass of living labour applied continuously declines about the mass of objectified labour that it sets in motion, i.e. the productively consumed means of production,

[13] Some Scholars have criticized this theory citing ‘transformation problem’. The transformation problem is the problem of determining a general rule for transforming commodity ‘values’ (according to Karl Marx's labour theory of value, premised on their socially necessary labour content) into market ‘competitive prices.’ Piero Sraffa demonstrated that any theory of surplus production and distribution, however, devised, is logically independent of any theory of labour exploitation. Labor exploitation can occur and be conceptualized in a variety of ways, regardless of which value theory is held to be correct. However, this does not prove Marx's theory of labour exploitation is incorrect. This theory's philosophical significance cannot be overstated. I took it in its philosophical essence for this thesis. (Steedman, Ian. 1977).

the part of this living labour that is unpaid and objectified in surplus-value must also stand in an ever-decreasing ratio to the value of the total capital applied. However, the ratio between the mass of surplus-value and total capital applied constitutes the profit rate, which must therefore fall steadily." (Marx, Karl, 1867 [1981].)[14]

Marx calls this tendency a 'double-edged sword' which produces its countering tendencies. In certain circumstances, these opposing tendencies may cause the profit rate to rise. This, according to my analysis, creates a false impression. I can call this as an illusion, an illusion that compels us to reconsider the artificially intelligent agent's capability to substitute a sentient (Sewell, Robb. 2012). What matters here is that, even after recognising distributed agency across a range of sentient and non-human actors that comprise a network, the foundational distinction between sentient and non-human actors stays in how they have an impact on a third object. (Bandyopadhyay, Ritaprava, et al 2022.)

Marx, according to Latour, believed that his critique of political economy revealed the pretence of the transfer of human agency (labor-power) in the commodity world. By doing so, Marx, according to Latour, undermined the power of non-human actors in the construction of actor-networks. The preceding analysis, on the other hand, demonstrates that Marx developed a historically balanced approach to comprehending the different forms of material connection that, in the first place, contribute to the formation of agency. Marx's analysis of the commodity form as a form of estranged interaction, as White (2013) writes, provides rich resources to that end.

[14] Quoted from Capital Vol. III, Part III. The Law of the Tendency of the Rate of Profit to Fall, Chapter 13. The Law As Such. see https://www.marxists.org/archive/marx/works/1894-c3/ch13.htm

The instances I gave earlier show that the rate of profit increases with the introduction of artificially intelligent agent as labour. According to Marxian evaluation, if I substitute human labourers with artificially intelligent agents as labourers, the profit rate tends to decrease. The rise in profit rate with the implementation of artificially intelligent agents as labour may indicate one of two things:

First, After Marx, I'll call it an ‘illusion’ that will reveal itself over time. In this thesis, I've provided examples of this.

Secondly, why some companies’ profit rate rises with the implementation of AIAs as labour alongside human labour could be that these companies use a labour-capital hybrid model. I have illustrated this too in this dissertation. (Bandyopadhyay, Ritaprava. Et at. 2022).

According to our understanding, I can say that the economic growth of China that I have discussed earlier, can either be an illusion or it is because of the successful deployment of the labour-capital model.

As a result, if we overly depend on ontocentrism, that is, treating ‘nonhuman’ as ‘human’ and attributing human-like agency to them, we may run into some ethical problems. I already discussed in this chapter as well as throughout the dissertation. One problem is regarding the ‘value’, the other is the ‘efficacy’ which will lead to the ‘question of desirability’. I have elaborated on the first argument in this chapter. I will elucidate the second argument in the next chapter. In this chapter, I will show that the problem lies with the ‘replacement’ of human agency by AI agents. This is precisely what our research question is all about--- Can AI Agents as labour replace Humans? If we rely too much on ontocentrism,

then it will eventually lead to the relationship of replacement. When we ascribe human-like agency to AI, we assume that AI agents in due course would replace human labourers.

The question we face at this point is whether we can acquire practical wisdom like value by learning general rules. Don't we need to practise deliberative, emotional, and social skills that allow us to apply our general understanding of wellbeing in ways that are appropriate for each occasion? This is a question of both social ethics and lifeworld. This raises the question of whether a robot can gain a ‘predictive hold’ on its behaviour, allowing it to attribute beliefs, desires, thoughts, and emotions to one another. Can artificial agents participate in human psychological activities and eventually replace humans? This will be covered in the following chapter.

Chapter 5

Can AI agents as labours replace humans in some situations?

I have discussed in the preceding chapter that, according to Karl Marx, the only source of value is human labour. It is human beings, who generate value. Marx would have counted artificially intelligent machines as 'dead labour' because they cannot generate value, they only transfer it.

At this juncture, one may ask: Why do humans generate value and machines cannot?

In this chapter, we will try to solve this problem. Value is an a priori concept that can be derived from higher intuition. It is defined as socially necessary labour time congealed in constant capital. According to Marx, value is an expression of human creativity. Creativity is a species-specific feature. Marx does not assign creativity to anything other than humans. Since value is innate and is expressed through human creativity, it has to be non-algorithmic in nature making it different from mathematics and logic. Liane Gabora and Scott Barry Kaufman have pointed out, it is possible to argue that creative ideas evolve as a result of culture. The adaptive and open-ended manner in which change accumulates distinguishes human creativity. Inventions improve on previous ones by increasing their utility or aesthetic appeal, or by making them applicable in new situations. There is no a priori limit to how a creative idea can evolve over time. For this reason, one has to participate in the 'life world'. It is clear from this discussion that in this chapter we follow a phenomenological approach.[15] Intersubjective experience, according

[15] There may be other explanations to this and philosophers like Dennett may not agree with us and can say that artificially intelligent agents can participate in the life world that we have mentioned, but this debate falls outside our discussion.

to Husserl, is crucial in the formation of both ourselves as objectively existing subjects, other experiencing subjects, and the objective spatiotemporal world. Intersubjective experience, according to Husserl, is empathic experience; it occurs during our conscious attribution of intentional acts to other subjects, during which we place ourselves in the shoes of the other. Amongst basic tenets thus disclosed by Edmund Husserl is the belief (or expectation) that a being who looks and behaves similarly to me, i.e., displays traits similar to my own, will generally perceive things from an egocentric viewpoint similar to my own. This belief enables me to immediately attribute intentional acts to others. It can be thought of in two ways: (1) in terms of belief and (2) in terms of something like a socially, culturally, or evolutionarily established (but nonetheless abstract) sense or meaning. The term 'lifeworld' refers to how members of one or more social groups (cultures, linguistic communities) organize the world into objects.

Let us consider an example consider from Husserl. Husserl says that we find coal as heating material. We recognise it as helpful and as a heating material, as appropriate for and destined to produce warmth. A combustible object can be used as fuel. It has value to us as. It can be a prospective source of heat. That is, it is valuable to us because it allows us to heat a room and thus provide pleasant sensations of warmth to ourselves and others. Others perceive it in the same way, and it acquires an intersubjective use-value and is valued in a social context as serving such and such a purpose, as useful to man, and so on. (Husserliana, vol. IV, pp. 186f; Husserl 1989, pp. 196, cited in Cited in Beyer, Christian. 2020.).

In Husserl's view, it is precisely this 'subjective-relative lifeworld', or environment, that provides the 'grounding soil' of the more objective world of science. In this chapter, we want to show that until and unless artificially intelligent agents as labourers participate in the 'life world' just like humans, it will be difficult for them to 'replace' human labourers in some situations.[16] However, a question that confronts us at this juncture is; can practical wisdom, such as value in our case be acquired by a learning algorithm or emulating our thought process only? Don't we need to learn value through lived experience and practise emotional, and social skills that allow us to apply our general understanding of wellbeing in context-appropriate ways? In other words, can a robot gain 'a predictive hold' over its behaviour so that it could attribute value, beliefs, desires, thoughts, and emotions to one another?

Adam Morton (2003) claims that 'sometimes and in some ways, we understand because we can cooperate rather than the other way around'. That is to say, it is partly because we can engage in cooperative activity that we can predict, explain, and understand action. We know from the everydayness of our experience that every situation is a new situation. Hence question remains, can artificial agents involve in these activities in every new, unknown, complex and emerging situation like human agents and would eventually replace humans in some situations?

[16] I will not go in to the detailed concept of lifeworld and its nuances or how it had been taken further by Heidegger and latter philosophers. I have taken the concept of lifeworld as conceived by Husserl. By lifeworld I mean the phenomenological world of intersubjective experience.

Section 1

Is it possible to compute Situation Ethics?

To elucidate our position, we need to clarify some ideas about Situation Ethics first.[17] Joseph Fletcher (1966) develops an idea of ethical non-system. Fletcher calls this ethical 'non-system', or 'Situationism'. Moreover, a Biblical reference would illustrate his position. There is a story in the Bible about Jesus healing a guy with a withered hand in the Jewish Temple. This was an expression that we thought demonstrated Jesus' fondness for everyone. However, the Pharisees chastise him because he conducted this recovery on the Sabbath, and Jewish law prohibits anyone from working on the Sabbath. According to Fletcher, Jesus's act might be morally acceptable even if it violates Jewish Law.

Fletcher, however, finds a middle way between 'Legalism' and 'Antinomianism'. He calls this 'Situationism'. Furthermore, he claims that in Legalism, people simply trust moral rules without regard for the context. Fletcher criticizes deontologists, who believe that actions are correct or wrong regardless of the consequences. For instance, this doctrine says, one should disclose the truth even if it means killing millions of people. Antinomianism, on the other hand, holds that an agent may do whatever she wants in a given circumstance. Fletcher refers to this as an existential perspective because it propagates the idea that people are independent to do whatever they want. Moreover, if antinomianism is right then morality becomes purely subjective in the sense that if an agent thinks that something is right then it is right. As a result, no laws or core principles exist.

[17] In this chapter I follow a phenomenological argument. But there are researchers, who believe that artificially intelligent agents can participate in the same world just like that of human. However, in this thesis, I will not discuss this.

To develop a middle way, Fletcher relies on Situation Ethics. He appears to believe that if moral laws are rejected, humans are compelled into ungoverned moral chaos. Fletcher, on the other hand, believes in moral law and, as a consequence, dismisses Antinomianism. There is only one moral law in his opinion. He claims that we should always behave in such a way that a large number of individuals are loved.

Fletcher's situationism may appear to be a type of teleological theory because it is concerned with the consequences that will determine if an action is right or wrong. To him, principles are context-sensitive generalisations derived from the one law of maximising love. For example, we could have a moral position that says we should not kill. This is a premise because we may believe that murder is unethical in general. And besides, it fails to elicit the greatest amount of affection. However, it is not a law because, according to Fletcher, killing is not always wrong. For example, a scenario could arise in which the child of a terrorist would have to be assassinated in order to obtain information to prevent a nuclear attack that would devastate a large portion of the globe. Fletcher might argue that there are times when we should disregard the principle and do the most loving thing possible, which in this case happens to be killing. We can only derive principles from universal law, not other universal laws. “We cannot milk a universal from a universal”, Fletcher says. (Fletcher, Joseph. 1966)

He however outlines four working principles of situationalism. These are

1) Pragmatism

2) Relativism

3) Positivism

4) Personalism

In situation ethics, right and wrong are determined by the circumstances. As a result, there are no broadly accepted moral rules or rights. The discipline treats each case is distinct and requires a distinct solution. Furthermore, situation ethics opposes prefabricated decisions and prescriptive rules. It does, however, teach that ethical decisions should be made on a case-by-case basis, using flexible guidelines rather than absolute rules. Because circumstances alter cases, situationism maintains that in practice, what we call right in one time and place may be incorrect in another time and place or in another context. To sum up, the elements of situation ethics as described by Joseph Fletcher are like this:

- Moral judgments are decisions, not conclusions.
- Situation ethics is concerned with circumstances, context, specificity, and cultural traditions.
- Every moral decision is needed to show respect for individuals and communities, as well as the things they value.
- This eschews the logical, detached, impersonal modes of thought that some people believe are overemphasised in other types of ethics.
- It is unique in that moral decisions are handled on a case-by-case basis, with decisions always customized to specific situations. (Fletcher, Joseph. 1966)

Thus, when people make a moral judgment or make a decision, they tend to evaluate an agent's behaviour in the light of a system of norms. Such evaluations of behaviour are fraught with inferences about what was in the agent's mind before, while, and even after performing the behaviour. Moreover, Morton argued that when we arrive at a decision, there are three aspects involved---psychological, physiological and social. Morton, on the other hand, emphasized the mutually beneficial circularities that exist between our abilities to attribute

mental states and our abilities to participate in shared activities. This means that our understanding of mind and action is shaped in part by its need to mediate shared activity, just as the shared activities we engage in are shaped by the need to rely on our capacities to gather and conceptualise information about one another. (Morton, Adam. 2003.)

Moreover, the concept of ethics emerges from society and it is subjective, not excluding inter-subjective. It depends on so many factors. For example, it depends on society, culture, conditions of one's upbringing, emotion, psychology, sexuality, belief, dream, feelings, longing, instinct, non-rational thinking, etc.

We can observe that the horrendous condition in which people of the Gaza strip live may shape their notion of ethics and morality. Similarly, the calm and affluent condition in which most Scandinavians grow up may affect in forming their notion of ethics. Hence, there cannot be a broad and general contour of how ethical notions develop. The realm of ethics, as we can see, is fragmented.

In the fat man example that we have cited previously, we have seen that decision to push fat man may vary and if we problematize and replace the 'fat man' with 'fat woman', 'black fat man', 'white and beautiful looking fat woman' then the result may vary.

It happens, one may argue, because life is multi-dimensional, complex, many-faceted and full of diversities and always flowing like a river. It is difficult to confine it in one form or the other. Hence, often its activity cannot be understood by 'reason' only. Often, we act according to our impulses. Often, we make decisions guided by our impulses and that may prove to be wrong afterwards.

There are, indeed, some, situations where one is drifted along with the inexorable flow of nature. In this context, we can cite an example from the *Gita* (Radhakrishnan, S. 2011). When Lord

Krishna went to Duryodhana and tried to reason with him, Duryodhana said, he knew everything. He, indeed, knew what was wrong and what was right, what was moral and what was immoral. However, he had his notion of right and wrong. And Lord Krishna was unable to stop the impending catastrophe. Duryodhana was, as if, drifting along with the inexorable flow of nature, in Sanskrit this is called *pravitti*. (Radhakrishnan, S. 2011) Since artificially intelligent agent does not belong to our world, it will not be possible for them to understand this.

Heidegger argued that we are able to comprehend the concept of a hammer or a chair since we were born into a culture that allows us to manage these objects. Similarly, Hubert Dreyfus believed that computers could not acquire intelligence because they lacked a body, childhood, and cultural experience. We have previously said that there is a difference between what computers still can't do and what computers will never be able to do. (Fjelland, R. 2020.)

Dreyfus (1992) opined that a significant portion of human knowledge is 'tacit'. As a result, it is impossible to articulate and implement in a computer programme. Michael Polanyi coined the term 'tacit knowledge'. (Polanyi, Michael. 1958) Dreyfus took his idea and ran with it. The majority of the knowledge we use in our daily lives, according to Polanyi, is tacit. In fact, we have no idea which rules we follow when we complete a task. Polanyi cites two examples. These are swimming and bicycle riding. Few swimmers are aware that how they regulate their respiration that keeps them afloat. Consider the sport of bicycling. The bicycle rider maintains his balance by turning the handlebars. She moves the handle to the left to avoid falling to the left, and she needs to turn the handlebar to the right to prevent falling to the right. As a result, she balances herself by moving across a set of small curves. According to Polanyi, a simple analysis demonstrates that for a given angle of unbalance, the curvature of each

twisting is inversely proportional to the square of the bicycle's speed. However, the cyclist is ignorant of this, and it will not help him become a better cyclist. Later, Polanyi stated that we can know more than we are able to say. (Polanyi, Michael. 1958.)

The important aspect of Polanyi's (1958) contribution, however, is that he argued that skills are required for articulate knowledge in general, and scientific knowledge in particular. Physical experiments, for example, necessitate a high level of expertise. These abilities cannot be obtained solely through the study of textbooks. They are learned through instruction from a tradesperson.

According to Hubert Dreyfus (1992), a large percentage of people are strolling specialists. However, attempting to articulate how we walk will almost certainly result in a description that does not catch the skills needed for walking. Likese, according to Dreyfus, artificially intelligent agents cannot grasp tacit knowledge or certain skills. Dreyfus (1992) unquestionably recognised a serious issue in AI. However, since Dreyfus (1992) raised these concerns, the concept of AI has evolved dramatically. During the 1980s, for example, a paradigm had become dominant in AI research. It was based on the neural network concept. It used the processes in our nervous system and brain as a model rather than symbol manipulation.

We can provide additional examples. Watson, IBM's computer, was designed specifically to appear on the game show Jeopardy! This is a competition in which the participants are given the answers and must therefore find the appropriate questions. Fjelland (2020) gives an example. He writes that for example, they might be told that “this ‘Father of

Our Country' didn't really chop down a cherry tree." "Who was George Washington?" is the correct question for the participants to answer. Jeopardy requires a much larger repertoire of knowledge and skills than chess. Science, history, culture, geography, and sports are among the topics covered in the tasks, which may include analogies and puns. It has three contestants competing to see who can answer first. Watson uses natural language to communicate. (Fjelland, Ragnar. 2020.) It was not connected to the Internet when it appeared on Jeopardy, but it did have access to two hundred million pages of information. Despite the fact that Watson was designed to compete in Jeopardy, IBM had other plans.

Watson promptly won Jeopardy! The company announced that it would use computer power in medicine, with the goal of developing an AI medical super-doctor who would revolutionize medicine. They thought, Watson, if given access to all medical literature (patient health records, textbooks, journal articles, drug lists, and so on), should be able to diagnose and treat patients better than any human doctor.

However, in the years since, IBM has been involved in a number of projects, with varying degrees of success. Some have recently been closed, while others have failed spectacularly. Creating an AI doctor has proven to be far more difficult than originally anticipated. Instead of super-doctors, IBM's Watson Health has created AI assistants capable of performing routine tasks. (Fjelland, Ragnar. 2020.)

Another watershed moment in AI research is AlphaGo, because it demonstrated the use of a strategy known as deep reinforcement learning. DeepMind is the company's name, and it reflects this. (Google and DeepMind are now Alphabet subsidiaries following a

reorganisation.) It is an example of an artificial neural network-based artificial intelligence research approach. A neural network serves as the foundation for an artificial neural network. Our brain is made up of approximately 100 billion neurons. Each neuron is connected to approximately 1000 other neurons via a synapse. This equates to approximately 100 trillion connections within the brain. Artificial neurons, which are significantly simpler than organic neurons, comprise an artificial neural network. Nevertheless, this has been established that by connecting a large number of neurons in a network, a sufficiently large network can theoretically perform any computation. Of course, what is practically possible is a different issue. (Fjelland, Ragnar. 2020.)

Another important example is IBM's Deep Blue. It was widely regarded as a breakthrough after defeating the world chess champion, Garri Kasparov, in 1997. Deep Blue was created for a specific purpose. Even if Deep Blue surpassed living beings in a task requiring intellectual ability, no one could assert that it achieved general intelligence. Nonetheless, it is an accomplishment. (Fjelland, Ragnar. 2020.)

Big Data, a recent contribution, is the application of mathematical methods to massive amounts of data to find correlations and infer probabilities. Big Data propagates that it is not necessary to create computers with human-like intelligence. Viktor Mayer-Schönberger and Kenneth Cukier (2013), conveys this message implicitly. Their book is upbeat about what Big Data's potential and the positive effects it will have on people's private lives and society in general. Numerous proponents contend that the conventional scientific process of assumptions, causal models, and experiments is no longer applicable. We all understand that causality is a key component of human thoughts, but this view holds that we don't need it.

Correlations are enough. For example, we can predict where crimes will occur based on criminal data and assign police resources. People could even be capable of anticipating and thus stopping a crime from occurring in the first place. In 2012, the White House, for example, declared a 'Big Data Research and Development Initiative' to address some of the nation's most pressing issues. (Fjelland, Ragnar. 2020.)

Even though Big Data analysis can be introduced as a new epistemological approach, it is more commonly regarded as a supplementary method for massive amounts of data, normally terabytes and petabytes. Viktor Mayer-Schönberger and Kenneth Cukier (2013) begin their book with the example of a 2009 flu outbreak. After combining elements from viruses that transmit bird flu and swine flu, it was given the name H1N1. It rapidly spread, and health officials across the world were concerned about a pandemic within a week. Some anticipated a disease outbreak on the magnitude of the 1918 Spanish flu, which killed millions of individuals. Since there was no vaccine against the viral infection, medical officials could only try to slow it down. Nevertheless, merely before the commencement of the pandemic-like situation, Google researchers created a technique that might anticipate the transmission of the flu far more correctly. Google receives over 3 billion web searches per day and ended up saving those. People who are sick with the flu are more likely to search for flu data on the internet. As an outcome, the investigators were confident to plot the transmission of flu much faster than health officials by looking at search items that are highly correlated with flu. According to Mayer-Schönberger and Cukier (2013), this is a success storey.

Fjelland (2020), on the other hand, thinks that this is an instance of 'the fallacy of initial success'. In 2013, the model reported twice as many visits to doctors for influenza-like diseases. The initial version of the model most likely included seasonal data that were

correlated with the flu but not causally related. As a result, the model served as both a flu detector and a winter detector. Despite getting revised, the model's performance has gone down far short of its preliminary promises. 'From these examples it may appear as if Dreyfus's arguments about what computers couldn't do were out of date. But Ragnar Fjelland, (2020) on the other hand, argued that the disparity between what has been accomplished and what has been promised is striking and Dreyfus's arguments are still valid. According to our hypothesis, Marx would have called this an 'illusion'.

Fjelland (2020) gave a few explanations for these phenomena. One explanation for this disparity could be that profit is the primary motivator for capitalist production, and thus many of the promises should be regarded as a marketing strategy. We know, that marketing strategy involves gimmicks. A marketing gimmick is a tactic used to entice clients to buy something. It's a mix of positioning, distinction, and inventive marketing. A gimmick is a ploy that is used to help you stand out in a crowd and attract attention quickly. However, while commercial interests undoubtedly play a role, Fjelland (2020) believes that this explanation is insufficient. Hence he thinks that there may be other reasons. The first argument is borrowed from Jerone Lanier, one of Silicon Valley's few dissenters. Lanier has argued that belief in scientific immortality, the development of super-intelligent computers, etc. are manifestations of a new religion 'expressed through an engineering culture.' (Fjelland, Ragnar. 2020.)

Secondly, Fjelland (2020) argued that when it is claimed that computers can duplicate human activity, it frequently turns out that the claim is based on a severely simplified and distorted account of that activity. Simply put, overestimation of technology is closely related to the underestimation of humans. (Fjelland, Ragnar. 2020.)

Moreover, Fjelland (2020) thinks that in all the previous cases only correlations were used. However, in both science and everyday life, we seek causal relationships. The nature of causal linkages has been debated for decades, especially since David Hume criticised the traditional notion of a necessary link between cause and effect. We must be content with observing regularities, according to Hume. In contrast, his contemporary Immanuel Kant maintained that causal links are required for knowledge acquisition. Every effect, according to him, must be accompanied by a cause.

Rather than delving into the philosophical debate over causal ties, which has raged on to this day, it might be more beneficial to look at how we recognise a causal relationship. John Stuart Mill, a philosopher, devised a set of rules (which he dubbed "canons") that allow us to recognise causal links. His ‘second canon’ which he also called ‘the method of difference’ is the following:

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon. (Cited in Fjelland, Ragnar. 2020.)

Mill’s second cannon is necessary because it throws light to the relationship between cause and effect. However, there are philosophers who do not subscribe to the notion of cause and effect.

There is a critical theoretical discourse on correlation and causation. Some researchers believe that causality can be established by correlation and with the help of statistical evidence. They believe that a strong correlation might indicate causality. Each

public matter of any factual nature now employs statistical methodology. The American Statistical Association's annual meetings cover almost every aspect of public policy, from nuclear reactor safety to census reliability. The efforts to retrieve causal data from statistics with only oblique help from experiments occur in almost all academic pursuits as well as many non-academic endeavours. Social psychologists, political theorists, economic experts, demographers, teachers, psychiatrists, biotechnologists, market analysts, lawmakers, and, on occasion, pharmacists and scientists use such methodologies. (Glymour, Clark. et al. 1986.) This group of researchers even claim that they can explain some natural laws with the help of correlation and statistical generalization.

On the other hand, there is a body of research that believe that "Correlation implies association, but not causation. Conversely, causation implies association, but not correlation." (Naomi, Altman, et al. 2012.) According to this school, mere association cannot be mixed with causation; if X causes Y, therefore the pair have been linked (dependent). However, associations can form between variables in the existence (i.e., X causes Y) or nonappearance (i.e., they share a common cause). The scholars ask us to envision that we notice that individuals who consume more than four mugs of coffee per day have a lesser chance of getting skin cancer. This doesn't necessarily indicate that coffee gives a person cancer resistance. One possible reason is that folks who consume a large amount of coffee work indoors for long working hours and therefore receive little sun exposure, which is a possible risk. If this is true, then the amount of time spent outside is a confounder—-a reason shared by both findings. A direct causal connection could not be deduced in such a case; the connection simply implies a supposition, such as a common cause, but does not offer evidence. Furthermore, when studying numerous factors in

complex systems, dubious connections can emerge. As a result, they hold that association does not mean causation. (Naomi, Altman, et al. 2012.)

According to this school of thought, a correlation between variables does not automatically suggest that an alteration in one variable is the cause of a change in the other variable. Causation indicates that one event occurs as a result of the occurrence of another; that is, there is a causal relationship between the two events. Correlation indicates that variables are statistically related. The process by which a change in one variable causes a change in another is known as causation. Causation always implies correlation because variables with a causal relationship are related. However, correlation does not imply causation because variables can be related without directly influencing each other. David Freedman and Paul Humphreys (1999) in their article conclude that the gap between association and causation is yet to be bridged.

As I have mentioned earlier, in this dissertation, I will not go into the details of the debate as it falls outside my research question. I want to mention that I have followed the arguments from the second school of thought and with the examples I have shown that AIAs in general follow the logic of correlation and not causation, while humans in their everydayness rely more on causation.

My position is that one cannot deny the doctrine of causality entirely as it is one of the key dispositions of humans. Indeed, the question of why Homo sapiens have been so successful in evolution is a complicated one. Numerous aspects have played a role, and one of the most significant is the ability to cooperate. We have already discussed this notion after Adam Morton.

However, between 70,000 and 30,000 years ago, a pivotal event occurred, which historian Harari (2018) refers to as the Cognitive Revolution. According to Harari, one distinguishing feature of the Cognitive Revolution is the ability to envision something that does not exist. Harari uses the 32,000-year-old ivory figurine 'the lion man' (or 'the lioness woman') is covered in the Stadel Cave in Germany as an example. It is made up of a human body and a lion's head.

The development of the lion man, according to Pearl and Mackenzie (2018), is the antecedent of philosophy, scientific discovery, and technological progress. The ability to assume and answer questions like "What happens if I do.....?" is a prerequisite for this creation.

After Fjelland (2020) we can say that in order to replace human labour in some situations, artificially intelligent agents have to pass the mini-Turing test that is based on finding the causal link. This test will be passed if computers can handle causal knowledge. The problem is that computers haven't progressed in this area in decades: "Machine-learning systems (including those with deep neural networks) function nearly entirely in an associative mode, much as they did 30 years ago..." (Pearl, Judea and Mackenzie, Dana. 2018.) This, as we think however, is insufficient. To answer causal questions, we should be capable of intervening in the world. The foundation of the problem, according to Pearl and Mackenzie (2018), is that computers lack a model of reality. The issue is that no one can have a realistic picture of reality. Any model can only show a skewed version of reality.

However, Harari (2018) discusses how behavioural economics and neuroscience have supposedly shown that our decisions are the result of 'millions of neurons calculating probabilities within a split second', rather than 'some mysterious free will' (Harari. 2018. p.

20). As a result, AI can perform many tasks better than humans. He uses driving a car in a crowded street, lending money to strangers, and negotiating commercial deals as examples. These vocations necessitate the ability to ‘properly assess other people's emotions and desires.’ (Harari. 2018.p20-21) To quote him “Yet if these emotions and desires are in fact no more than biochemical algorithms, there is no reason why computers cannot decipher these algorithms—and do so far better than any Homo sapiens.” (Harari. 2018.p20-21)

Fjelland (2020) however argues that there is an ambiguity in this kind of thinking and this is evident from the previous quotation. Fjelland (2020) asks if Harari is correct, why does he use the phrase ‘no more than’ the behaviour of a large group of nerve cells? Fjelland (2020) contends that if we ignore the issue of self-reference considering the perfect world of science to be the only real world, Harari's argument makes sense. But according to Fjelland (2020), the replacement of our everyday world with the world of science, on the other hand, is based on a vital mistake.

Husserl was among the first to point this out, attributing the error to Galileo. Galileo was ‘at once a discoverer and a concealing genius’, according to Husserl. This misunderstanding was dubbed ‘objectivism’ by Husserl. Today, the term ‘scientism’ is more commonly used. Husserl, on the other hand, insisted that science is fundamentally a human endeavour. Even the most abstract theories, Husserl's ‘lifeworld’, are grounded in our everyday world. Husserl mentions Einstein's theory of relativity and claims that it is based on ‘Michelson's experiments and other researchers' corroborations of them’. To conduct these types of experiments, scientists should be able to walk around, manage to interpret scales, and communicate with other scientists. (Fjelland, Ragnar. 2020) There is a much more reliable

narrative of how we understand others than Harari's. (2018). We are corporeal and social species living in a material physical and social world, as Hubert Dreyfus pointed out. Understanding another person and to replace her in certain situations requires being in that person's shoes, not looking into the chemistry of that person's brain or even into that person's soul. It is to comprehend the individual's living circumstances. This is what we mean what artificially intelligent agents will never be able to do as they do not take part in the lifeworld just like humans. (Fjelland, Ragnar. 2020.)

Let us illustrate our position with a thought experiment that, like the Trolley Problem in AI, will open various possibilities in making a decision in 'lifeworld'.

A child-less couple went for a vacation in Darjeeling in December. Every day, however, they would stroll a long way. One day they were returning to their hotel from a distant place. It was quite late. The temperature was almost at a freezing point! They were returning quite briskly. Lights still twinkled in the hills. The shop-fronts were closed. The not-so-wide streets, indeed, were quiet. Doors and windows were shut. The street was so silent, that people living on either side could hear their brisk footsteps. Without any doubt, it was a cold night.

However, the life of the hill town was going on as usual. There were occasional loud remarks, music from the TV, a burst of laughter. The street was unlit. However, there was no problem with visibility. In the clear sky, the three-quarters moon was up. They were alone. But we're aware of the life pulsating around them.

As they walked further along the empty street, all of a sudden they heard a noise, a desperate cry of a child. They started running towards the place from where the sound came.

Moreover, to their utter surprise, they found an almost new-born baby, a girl child. A shawl was wrapped around her shoulder. She was wide awake and shivering.

Immediately, the lady took the child to her lap, wrapped her further with her garments. Then they almost ran to their hotel. Consulted with a physician, gave the new-born her food.

Here comes, however, a question: What is ethically prudent for this couple? This situation has many possibilities.

On one side, there is the question of emotion, the eternal craving for a childless couple to have a child. On the other side, there is the ethico-legal question. Should they adopt the child? Should they inform the police and try to find out her biological parents? The child, being a female one, should the question of gender discrimination anyway influence the decision? Should they hand over the child to an orphanage centre? Is there any probability of a conflict of opinion between husband and wife?

Indeed, the difference and plurality of thought give rise to numerous possibilities which cannot be predicted. It may, however, be possible for human beings to decide in such a situation according to their values and ethics. In this case, the ethical decision should be guided by adaptable guidelines rather than rigid rules. Flexible guidelines, however, depend on so many factors. In our thought experiment, the emotion of the childless couple may differ from the emotion of a couple who have children. Hence, the decision-making process of a childless couple who found the child from the roadside may differ from a child trafficker, the police, the student, a government official, a transgender or a saint. Hence, the narrative of ethics will differ from one to another. Thus, the decision-making process also differs.

The question, however, remains, how does an Artificially Intelligent Agents, which does not belong to this world, react to this type of situation having multiple possibilities?

Society, as we have seen is not a homogenous one. Ethics, as a way of life, emerges from this heterogeneity of society. It is grounded in the fast-changing life world. Hence, it is very tough to generalize a situation beforehand on which a decision is made as every moment is a new moment, every situation is also a new one. We have already shown that in the case of a person, to arrive at a decision depends on so many factors. Hence, one cannot inductively generalize certain situations and try to figure out some formula that will act as a guiding principle for decision-making. Because, the act of generalization may not be an 'objective' one, but 'subjective' and it depends on so many factors which we have discussed earlier. Hence, in our thought experiment, if a robot couple found the baby girl on the desolate road of Darjeeling, what would have been their response? We could speculate but, we can be assured that it may not be as varied as that of a human counterpart as it does not take part in the life-world like the humans.

Moreover, if programmed, ethics defies the very nature of reasoning, which is many-faceted, complex and influenced by so many factors that we have previously seen. Hence, we cannot say that the thought pattern could be generalized and we could arrive at a situation by applying the method of inductive generalization. As a similar situation, does not evoke a similar response and similar approach to arrive at a decision. Moral judgments, indeed, are affected by rights, such as privacy, roles, such as in families and society, past actions, motives, intentions, and other morally relevant features. It may be hard to incorporate these diverse factors into AI systems as they are not bodily and social beings existing in a material and social world.

Fjilland (2020) asks us to consider another thought experiment created by Theodore Roszak, an American author. Assume we're observing a therapist in action. He is a

hardworking as well as an expert psychiatrist with successful experience. The reception area is crammed with sick people who suffer from a diverse range of mental and emotional ailments. A few are nearly hysterical, others have suicidal tendencies, some have hallucinations, still others have the most truly horrific bad dreams, and even more, are feared for the fact that they are being viewed by individuals who will affect them. The psychiatrist devotes particular time to every patient and makes every effort to assist them, but with limited success. Conversely, they all seem to be getting severe, despite the psychiatrist's heroic efforts. (Fjelland, Ragnar. 2020)

Roszak, in Fjilland's (2020) article urges us to consider thinking in a broader background. The psychiatrist's workplace is in a house, which is located in central Germany. The name of the place is Buchenwald. The clients are detainees from a concentration camp. We would not be able to understand the patients using biochemical algorithms. Hence what is required, is knowledge of the greater background. If one doesn't know that the doctor's workplace is in a concentration camp, the example makes no sense. A handful of people can put themselves in the shoes of a prison camp detainee. As a result, we cannot entirely understand persons in circumstances that are vastly dissimilar from our own. But one can understand to some extent because we are also part of the world.

A few people can imagine themselves in the position of a prison camp inmate. As a result, one might not be able to make out how the situations that are vastly different from their own. And yet, to a certain degree, folks could indeed comprehend, since humans, too, have become a part of the universe. In our universe, computer system does not prevail. As we previously stated, neural networks do not need to be programmed and can thus grip tacit knowledge. But handling tacit knowledge is not sufficient to 'understand' the

world. However, it is simply not true, as some Big Data supporters claim, that the data 'speak for themselves'. Typically, the data used is related to one or more models, is chosen by humans, and ultimately consists of numbers. As a result, Fjilland believes that Hubert Dreyfus' arguments against general AI remain valid. (Cited in Fjelland, Ragnar. 2020.)

Now, we will design another thought experiment that will show how the process of generating value from the lifeworld makes the difference between labourers.

Imagine yourself at a restaurant where the waiter is an artificial agent, a robot precisely. Moreover, it is programmed in such a way that it is as helpful as its human counterpart. It knows every nitty-gritty detail of the items in the restaurant. It knows, for example, how to manoeuvre, a dish of chilli chicken tastier by adding extra sauce or making it less spicy as is required sometimes by many. One day, a couple with their five-year-old girl visited the restaurant. Upon reaching there, they were cordially greeted by the robot waiter. It gave the menu card and waited patiently, as had been programmed.

After a brief discussion, the couple settled for a dish of fried rice and Chili Chicken. It is this time the lady asked the waiter to make the chilli chicken less spicy and to add moderate chilli sauce so that her daughter could eat comfortably. The order was served in time by the robot waiter. But having tasted the food, the lady found it to be a bit spicy and felt that her daughter could not enjoy the food. She immediately called the robot waiter and asked why was the chilli chicken cooked spicy even when she mentioned the robot waiter to make it less spicy? How will it negotiate with the situation? Will it show soft-skill to negotiate with the customer, as it did the right thing to make the food less spicy, still the food appears to be spicy to the customer? If the person in the thought experiment becomes furious for not getting specific and customised and goes on a rampage in the restaurant, will the robot waiter stop

him from doing the same? What will it learn from this? Will this situation help other robot waiters to learn? These are the questions that need to be answered by a robot waiter, who, in a sense, is labour.

However, as the research in AI progresses, the use of machine learning is becoming more important. Using this, AI researchers try to identify what the general pattern is and determine the extent to which we could reproduce those kinds of decisions in certain situations. The question arises, how could the AI researchers arrive at a 'general pattern' of human thought process, when there is none? This confronts us with another question: How does a robot with its continuous learning process 'learn' something? Will it be possible for it to unlearn something with the help of the same logic?

Here goes another thought experiment. We know archetypal autonomous machines like driverless cars are learning machines. These machines are programmed in a way that they can collect information, process it, draw conclusions, and change the ways they conduct themselves accordingly, without human intervention. Imagine a situation where such a car may set out with a program that gives an instruction not to exceed the speed limit, only to learn that other cars exceed these limits and concludes that it can and should speed too. To prevent it, you need to incorporate another code and that will be followed by another one. However, this would involve in *infinite regress.*

Nevertheless, this will confront us with another question: is it possible to create code that can handle an infinite number of scenarios? This question seems more relevant as Monica Rozenfield (2018) writes, "Deep learning is a fairly new type of AI which adds a new spin to a former technology named neural network rendering it feasible by big data, supercomputing, and complex algorithms. Each neuron in the network has data lines that interact with each

other. It might be difficult to create code for every conceivable scenario. An AI device would be unable to function without the appropriate code. Deep learning, on the other hand, allows the system to sort things out on its own. The method allows the network to form neural relationships that are most pertinent to each unfolding scenario". (cited in Etzioni, Amitai, and Oren Etzioni. 2018.)

Hence, how will it deal with the fast-changing new situation and the nitty-gritty factors of real-world, is nothing but a mystery and some think that since artificially intelligent agents do not belong to this world, it will not be possible for them to cope with the situation.

Are Artificial Agents going to replace humans?

According to our understanding, thus, to replace human labourers in some situations, one of the main things that artificial agents must acquire is the knowledge of the lifeworld. There is, however, no denying the fact that technology has reached a new height and over time, it will improve further. But we have shown in the previous thought experiments that there will remain a tantalizing and of course very subtle difference between the world of technology and our day-to-day natural world.

It is a fact that our day-to-day world or lifeworld plays a very important role in understanding others. We hope we shall be able to clear it by providing a thought experiment.

Let us imagine, in a given situation a mother feigns anger to her child who is very reluctant to eat her meal. However, the mother rebukes her and says, she would get cross with her and never take her for a joy ride! This is a kind of acting on the part of the mother which the baby takes in its face value and thinks that her mother is very angry, which, actually, is

not the case. Nevertheless, sometimes this pays. The child thinks that her mother is really very angry and she would not allow her to a joy ride. So, she eats her meal quickly.

In another situation, however, the mother rebukes her child because her pestering crosses all limits. This time mother is really angry and is not, indeed, acting.

Now the question is: can AI agents distinguish between these two situations i.e. situation where the mother feigns anger and the situation where she is really angry? However, technology has developed facial recognition devices. No doubt many more sophisticated devices will be developed in near future. Will these be sufficient to differentiate between pretension and intention in a given situation just like human agents in the natural world?

Now, let us replace the mother in this thought experiment with an artificial agent. The question will arise; will it handle the two situations like that of the real mother? If it plays the role of the said mother, could the child's reaction be similar?

Through statistical generalization and programming, artificially intelligent agents can do a miracle! But will this be sufficient to learn how to participate in the life-world just like the ways humans do and interact similarly to a human? Another question is, will the child in our thought experiment react similarly to an artificially intelligent agent mother as she reacted with her biological mother? Will these two situations be the same?

Our hypothesis is that; since overt physical activity is similar in the case of 'pretension' and 'intention', it will be difficult to differentiate between two situations for Artificially Intelligent Agents. Though the supporter of affective computing[18] may not agree with us.

[18] In this context, we need to discuss some ideas regarding affective computing proposed by Rosalind Picard in brief. In the year 1995 Rosalind Picard introduced the concept of Affective computing. According to Picard, affective computing is such computing that associates to emotions. Picard thinks that closing the gap between humans and machines is one of the prime objectives of affective computing. Building artificial agents (such as robots) to interact with humans naturally and emotionally is another objective. However, at present research in this area has acquired immense importance. Disciplines such as neuroscience, psychology, education, medicine, sociology, computer science, and cognitive science are contributing to this area of research.

However, we have designed a thought experiment to illustrate our position. Mr X is very depressed. Two incidents have almost simultaneously occurred. First, the result of his daughter is out and it is not at all up to the mark. At the same time, he received a mail from the HR department of his office. He came to know that this year also he has been denied much-coveted promotion citing the pandemic situation. In such a case what would be the primary reason for his depression? Is it her daughter's result, or is it for not getting a promotion, or because of both? However, our question is, can the content of thought be measured? If the new-age machines are endowed with thought, then they will be able to recognize human thought and respond accordingly. Theoretically, indeed, it is accepted. However, a question

Hence, it is an interdisciplinary field of research. The recent trend in Affective Computing research, moreover, pinpoints its focus in the matter of estimating human emotions taking a cue from different forms of signals. For example, 'face recognition', 'EEG', 'Speech Perception', 'PET scans' or 'fMRI' help in providing the required data to unveil the human emotion. Inferring the emotion of humans accurately, indeed, is difficult, as emotion is subjective. Additionally, psycho-physiological expressions and biological reactions give birth to some unconscious experience. Hormones and neurotransmitters such as dopamine, serotonin and oxytocin play a pivotal role in the psycho-physiological response. Moreover, various mental states are closely related to the arousal of the nervous system which plays a big role in eliciting emotion. Picard, however, thinks that only machine learning or big data analysis is not enough to make out emotion or affect. In understanding emotion, the help of neuroscience is necessary. Recent research, however, shows that computers can achieve very near-accuracy levels of emotion recognition. Thus, visual, textual, and auditory sources help computers in this regard. Furthermore, new findings show that artificial intelligence can recognize emotions using facial gestures and voice recognition techniques.

Thus, Picard believes that really 'intelligent' computers must 'recognize', 'understand' emotion. In this way, it can interact normally with us. She moreover opines that not only recognizing emotion is essential, but computers must also express it.

Research, in particular, shows that emotion plays a great role in cognitive processes like perception, learning, decision making. Hence, every aspect of rational thinking is influenced by emotion. The researchers, however, think that decision-making is greatly influenced by emotion. It should be noted that, too much emotion, as well as too little emotion, can be detrimental in the decision-making process. Hence, there must be a perfect balance.

According to Picard computational devices are useful to detect emotions. It can detect a subset of emotions that can show characteristic patterns in measurable physiological states. Nevertheless, to paraphrase Picard, the motor system serves as a vehicle for expressing emotional state. Sendic modulation refers to the influence of emotion on bodily expression. However, Picard borrowed the adjective 'sentic' from the Latin word 'sentire', which is the mother word of 'sentiment' and 'sensation'. By this word she emphasizes 'physical mechanisms of emotional expression.'

However, it may seem that not all emotions generate measurable physiological responses or 'sentic modulation' as Picard has envisioned. On the contrary, some emotions involve thought contents. These emotions can be as diverse as thought. Of course, there are some 'universal categories' of emotion like sadness, anger, happiness etc. but there may be a difference in their contents.

Hence Picard attempts to build the foundations for providing computer technologies with emotion. But she is realistic in her attempt. She, however, feels no need for her printer to be emotional. She also acknowledges that building affective computer is a challenge. From the previous discussion, we find that human emotion, both primary and secondary can be decoded. Artificially intelligent agents can be trained in this. The question however remains; can artificially intelligent agents understand the content of emotion? Human also sometimes make error to understand other person's emotion. However, this is a different domain of research and for our present purpose, We will not go into detail about this because it is outside the scope of our research. [Picard, Rosalind. 1997. *Affective Computing*. *Cambridge:* MIT Press.]

arises, what would be the content of the thought that our engineer friend wants to measure? We popularly believe, that a particular type of thought can be identified but we cannot know its content. We propose that, since humans have an 'inner life' that artificially intelligent agents don't have, it will be hard to measure the content of thought.

We know that human beings have an evolutionary history. Humans, indeed, consist of so many evolutionarily old components. One of these is intelligence. If an artificial agent has every input regarding a person, it can do many things. For example, it can even do counselling for its human counterpart. But if it has to simulate its counterpart then it will not succeed as it does not participate in the life world.

In this context a question arises: without understanding the content of thought can an artificial agent take part in this 'shared' activity? Through the definition of labour, we know that labourers participate in a shared activity.

Adam Morton, however, believe that there is a "beneficial circularities between our capacities to attribute states of mind and our capacities to engage in shared activities."[19] This means that our understanding of mind and action is formed in part by the need to mediate shared activity, like the shared activities we engage in are moulded by the need to rely on our capabilities to gather and conceptualise information about one another. (Morton, Adam. 2003. p 149.)

According to Morton, the idea of 'shared activity' is that we reach a judgment about what we ought to do first, and only then form an expectation of what the other person will intend/desire to do we act. In this context, we can cite an example after Morton. Suppose that a person is helping another person to move a table through a narrow doorway. Will the first

[19] Morton, Adam. 2003. P 148.

person turn the table to the left or the right? If she turns it to the left, she might get some advantage and similarly, if she turns it to the right, then also she will have some advantage. But their direction will have to be identical when the decisive moment comes. When the first person moves to the right, the hand of the second person will follow suit.

In this way, one makes it clear that this is the only way to solve the problem. The second person then understands what would be the action of the first person. And he will act accordingly. (Morton, Adam. 2003. pp 14-15.)

Morton's opinion is that his propensity enables us to know, anticipate, and demonstrate the activities of someone else, which on the other hand helps the former to choose her own course of action. The author demands that cooperative activities of some type are based on everyday psychological understanding conversely. Following this, we behave to make ourselves intelligible to others. Likewise, one gains from being comprehended. This concept of 'beneficial circularities' is central to Morton's research. According to him, we comprehend each other because we have learned to make ourselves comprehensible. Adam Morton examines the notions of believing and simulation, the idea of understanding by intent, and the causal force of psychological explanation using examples from cooperative activities such as driving a cab and playing table tennis.

Hence, there are two issues here: if an artificial agent has to qualify as an 'agent' it has to make itself understandable to others. It will have to act in such a way as to make its actions easily intelligible to others so that it can be benefited from being understood. In so doing, it will have to participate in the life world and understand among many other things the semantic content of thought and has to be easily accepted by others. Otherwise, it cannot 'replace' human labour in certain situations. This paves the way for our second section.

Are they going to supplement human capacities in some respect?

In some cases, indeed, the answer is yes. We have plenty of examples that show in some cases artificial agents may be helpful and do better than humans. If it is repetitive work, an artificial agent excels. We can take the example of Amazon. The company has employed more than ten thousand robots in its warehouses to efficiently move things around. It has increased its warehouse workforce by more than eighty thousand. (Swapna. G and Nivashiniya. R. 2021.) It is learned that in Amazon's warehouse humans do the picking and packing of goods. On the other hand, Robots move orders around the giant warehouses, essentially cutting 'down on the walking required of workers, making Amazon pickers more efficient and less tired.'(Lokitz, Justin. 2021.) In addition to this, robots do many other things to facilitate the process much more smoothly. (Lokitz, Justin. 2021.)

Sentic computing (Cambria, Erik, Hussain, Amir, 2021.) requires a multidisciplinary approach to solving problems in the framework of natural language processing. Sentic is derived from the Latin 'sentire' and 'sensus'. In sentic computing, however, the analysis of natural language is based on general knowledge reasoning tools. These analyse text at the document, page, or paragraph level. At the same time, it also analyses the sentence, clause, and concept levels. (Lokitz, Justin. 2021.) It can be useful to combat trolls on social media as well. Currently, the method for identifying anti-trolling consists of the discovery of other accounts which follow the same IP address. Thus, this method helps block fake accounts if it finds some anomalies. Recently Meta launched an application. This application provides users with a link that will help them in reporting cyber problems like child exploitation. Moreover, it also provides an online protection centre (CEOP). (PTI. 2010.) Hence, in these cases, artificial agents can substitute human labour. But in the case of differentiating between 'pretension'

and ‘intention’ or ‘generating value’, this model may not work. Because ‘pretension’ or ‘acting’ or ‘generating value’ can only be understood in proper context. For that, they will have to participate in the ‘lifeworld’. In order to understand artificially intelligent agents must learn the act of simulation, which they cannot.

We have seen in the thought experiment what pretension is. Now we will give an example of ‘acting’. The most powerful actor is one who bridges the gap between the real and the reel. Take for example the last scene of the 2016 American musical romantic comedy-drama film *La La Land* (Chazelle, Damien. Director & Writer. 2016). In the final scene, Emma Stone (Mia) and Ryan Gosling (Sebastian) gaze at one another for a few seconds then both of them smiled and Mia left. They did not say any words but that gazing and the smile eloquently speak about their past, their friendship, and the relationship they had. Some years passed. A sea change, indeed, had happened to the life of both of them. Mia excelled as an actress. She became famous. In course of time, she married a person David. She had a daughter. All the signs of a happy family, however, were there. One day, amid her busy schedule, she wanted a break. However, she came to her old place along with her family. On a moonlit night, Mia and David had gone for a long drive. They came to a jazz bar. Suddenly Mia notices a logo in the bar. Seeing the logo, she understood that bar might belong to Seb. They entered into it. When Seb saw Mia in the crowd, he begins to play their love theme on the piano. Amid the programme, she was at a reverie in which she could visualize what would have happened had their relationship materialised. Meanwhile, Seb completes his playing. After that, a round of tremendous applause filled the bar. Now Mia had to leave. Before she left the place she looked back to Seb and they had a mute exchange of smiles, full of melancholy. This is the most dramatic scene in the movie. Now if we replace Seb with an Artificial Agent, the question

arises, will it act like Seb? For the sake of argument if we consider an Artificial Agent will play the role of Seb, will it assimilate the long history and evolution of Seb? If we take it for granted too, then the next question would be, will its behaviour generate the same reaction as that of Mia in the film? The whole episode revolves around the semantic content of emotion where their past is involved and there are lots of things implicit in this saga.

So the question remains, can an artificial agent understand the implicit content of human emotion? By implicit content of emotion, we mean the past life with its entire struggle, their mutual departure for the pursuit of excellence, and their acceptance of the course of life. This is a situation where both of them engage in cooperative activity so that we can predict, explain, and understand action. Devoid of the content of a particular emotion, however, can an artificial agent enter in such a cooperative activity? If not, then can it be responsible for generating the reaction of its fellow human being? Taking part in the lifeworld plays a crucial role here as both of the agents are involved in a 'cooperative activity'. In the backdrop of Mia and Seb's reaction on the screen, there are lots of things that happened. It is, indeed, participation in a lifeworld that makes all the difference. In order to understand others, they relied on the simulation theory, which is nothing but putting oneself in another person's shoos.

In this context, we will give an example of another situation. It was the early 1990s. After prolonged chaos and mayhem Darjeeling was relatively calm. The agitation was called off; a peace accord was signed. Mr X, (Bhattacharya, Parimal. 2017.) after the completion of his post-graduation degree, got a fellowship. But the tenure of the fellowship was nearing its end. Hence, employment was a crying need for him. However, at that time the job scenario was not very good because of a controversy over the mandatory National Eligibility Test for college teachers. In such a situation Mr X got an appointment letter from the office of the

Public Service Commission. His posting was in Darjeeling Government College. He was, indeed, ecstatic and excited! In his words, "a lectureship in a Government College was like the last metro in a midnight city." However, his mother was very apprehensive because from the newspaper report it seemed to her that all was not well on the hill. However, the fire was smouldering. Son tried his best to make her understand the reality. Her answer was, 'but you wouldn't understand what passes through the mind of a parent when her son decides to go to work in such a dangerous place!' So he tried to pursue her mother, gave her logic, and took the job. In this situation also, we can see a higher level of cooperative activity is involved between a son and her mother in reality. No one is pretending or acting. This is a real-life scenario where a mother and her son are involved in a dialogue that would shape the course of action in the future. If we replace either the mother or the son with an AI Agent, will it be possible to generate the same reaction from their counterpart? Will natural language understanding and neuro-physiological data explain 'agony' and 'ecstasy'? This will not be the case, as the AI agents do not belong to the lifeworld, or can simulate.

In this context, we would like to mention that Rabindranath (1926) wrote a song '*anek katha jaao je bole kono katha na boli/ tomar bhasha bojhar asha diyechi jalanjoli*'(You tell many things without telling anything/ I have failed to understand your language.) Pupe, adopted grandchild of Rabindranath was the inspiration behind this song. So even apparently the meaningless strings of words find meaning in the mind of a poet. From the meaningless strings of words, Rabindranath could formulate a poem that has meaning. Can a robot do the same?

From the previous discussion, we come to the conclusion that, in some cases, artificial agents would perform better than humans in certain situations, but there are situations where

they would be outplayed by humans. Hence, in those situations, they cannot replace humans, they can at best supplement humans' capacities in some respect. This confronts us with a question--- are they going to be there along with humans? If so, what would be the relation? This will be discussed in the last chapter, which will be the conclusion of the thesis.

Chapter 6

Conclusion

Relationship

I proposed in the introduction that our research question was: would Artificially Intelligent Agents as labour replace human labour in some situations? Furthermore, I have framed our research question within the larger context of soft morality and artificiality. In so doing, I defined some key concepts (such as soft ethics, hard ethics, soft-morality, artificially intelligent agents, labourers, value, and so on) that would be important throughout my thesis. However, in the introduction (first chapter), I attempted to situate our problem within the larger contexts of Ethics, Artificial Intelligence, and lifeworld, which serve as the 'foundation' for all shared human experience.

In the second chapter, I discussed how artificially intelligent machines as autonomous agents failed in certain situations, leaving us in moral quandaries of various sources. To problematize further, I mentioned some well-known thought experiments (e.g., the Trolley Problem in AI, Fat Man Thought Experiment, etc.) and reiterated some of the famous ethical dilemmas in AI in this chapter.

The following chapter is primarily a review of the literature. I stated that my research problem stems from Digital Ethics. Digital Ethics, on the other hand, emerges at specific points in time. Similarly, Information Ethics (defined as ontocentric, patient-

oriented, ecological macro ethics) has emerged as a critique of anthropocentric ethics. Luciano Floridi, a pioneer of information ethics, sees it as the 'fourth revolution,' following the Copernican Revolution, Darwinism, and Freudianism. Human beings have been pushed from the centre to the periphery in each previous 'revolution.' In this chapter, I attempted to trace the path from anthropocentrism to ontocentrism in Ethics. We've seen in this chapter that Artificially Intelligent Agents are regarded as 'agents' by Information Ethics. These agents, however, are the ones who pose a threat to human labour in some situations. From the previous discussion, a question comes to my mind--- Could ontocentrism be the end of the road? Thus, I answered this question in the next chapter.

In Chapter 5, I began by stating that a few decades ago in Sociology, we saw a familiar turn that we have observed in ethics. Furthermore, the main theme in this theory is that humans, along with other things (non-humans), occupy the centre of discussion. In this chapter, I discussed Bruno Latour's Actor-Network Theory (ANT) and how the concept of ANT has since been expanded to include AI agents. Furthermore, the incorporation of AI in actor networks raises some legitimate concerns about its ethical implications, such as the replacement of human labour in certain situations. Moreover, we cited some mainstream media reporting that demonstrated that the use of artificial agents in some cases affects the rate of profit of some companies. Towards the close of the chapter, I explained these phenomena using Marxist ethics of value as a case study.

Why AI agents will not replace human labour in certain situations has been answered using two arguments. One has been explained using Marxist ethics of value, which states that only humans create value, and machines, such as AI agents, pass it on.

This is one of the reasons why, after the implementation of AI Agents as labourers, the profit rate of some companies tends to fall. Another question that has been raised at this point is why Artificially Intelligent Machines do not generate value. In this chapter, I demonstrated that value, understood as socially necessary labour time congealed in constant capital, is a concept derived from lifeworld. I have shown that in order to generate value one has to participate in the lifeworld. By lifeworld, I mean the phenomenological world of intersubjective experience. In Phenomenology, the lifeworld is the world as it is immediately or directly experienced in the subjectivity of everyday life, as opposed to the objective 'worlds' of the sciences, which employ the methods of mathematical sciences of nature.

The lifeworld involves personal, social, perceptual, and practical experiences. Through experience, humans learn whatever they could learn. Furthermore, human experience tells her that reality is too diverse to be explained algorithmically. To establish my position, I designed some thought experiments. I have shown that understanding others through mental simulation occurs by participating in the lifeworld.

Moreover, if opponents argue that they will create an algorithm for every changing situation, the question of whether it will be desirable remains unanswered. As we know, at a point of time creating an atomic bomb was plausible but history teaches us that it was never desirable. I tried to show that Artificially Intelligent Agents can imitate humans' emotions and their linguistic ability (in this respect we have discussed Rosalind Picard's view on affective computing in brief), but it is impossible to imitate humans by their subjective experiences. We can say after Nagel that 'What is it like to be a bat' can

only be understood from a bat's point of view. Similarly, 'What is it like to be human' can only be understood from a human's point of view. Hence it is impossible for an artificially intelligent agent to 'understand' a human's point of view.

We want to add further that lifeworld is the grand narrative that provides the content for smaller narratives that constitute individual selves which is the repository of value. This makes human labour different from Artificially Intelligent Labour.

However, it is the lifeworld where agents interact with each other and come to decisions about each other's thoughts, plans, actions, feelings and emotions. In some cases, human labour does precisely these things to do the job. Throughout the thesis, we have cited plenty of thought experiments to show this. Moreover, by participating in the lifeworld we understand each other as like-minded, autonomous, and efficient beings like human labourers. It is the lifeworld which forms our concepts of personhood, mind, and action through which humans conceptualize a problem and devise its solution. In the process, she generates value with time. Lifeworld, however, is the source of human creativity and value too.

This confronts us with a question, can I put myself in an Artificial Agents (AI Labour in our case) shoes and simulate what would be its course of action? The question is a tricky one as those who have made these artificial agents can predict their actions in some cases (even they cannot predict their actions in some cases as well), but what then for the common people? Will they simulate their states and understand them? Will they be ready to ascribe agency to them?

In this context, we want to understand the role of the lifeworld in the ascription of the agency. However, it may be viewed from the perspective of the first, second and third-person narrative.

Sometimes we engage in the act of third-person psychological understanding and use some concepts to explain and predict the thoughts and actions of others. In these cases, we use a third-person narrative (or theory) to ascribe selfhood and agency to them. Moreover, this can be understood with an example. If we see someone sitting at the corner of a resting room in a railway station, with her head bent and eyes wet, we reckon that she is upset about something. Though she is a stranger, we may feel the urge to console her and relieve her of her misery. We believe she is capable of doing and thinking things, as well as reacting to situations. She is someone who can be spoken to and consoled. She is someone whose actions and behaviour can be altered. This is a case of the third-person ascription of the agency.

Moreover, I may modify this example slightly to see how a narrative may be used to ascribe agency in the second-person case. Suppose we see our brother, *Rohit*, a middle-school-goer, sitting in a corner of a gym, head bent and eyes wet, we imagine that he has been hurt or perhaps punished for some misdemeanour. We may rush to him, gather him in our arms and try to find out what happened. We may know what the best means of consoling him are, as he is our brother. We may, for instance, know that an hour spent with our pet parrot, *Bonnie*, will make him forget his unpleasant experience and later, a visit to the local Children's Park, where he can play a game of cricket with his friends, may cheer him up further. *Rohit*, we believe, is an agent. His thoughts, feelings and

actions matter; they make sense. They are caused by factors in the immediate and not-so-immediate vicinity (in this case lifeworld), and if the factors are changed, a change may come about in his thoughts and actions.

What then would the first-person ascription of the agency be like? Suppose one of us, *RB*, goes to buy fresh fish at the evening market where he meets his arch professional enemy. *RB* is jittery but he conceals his emotions well. This is what he tells himself: 'I know you will expect me to react and misbehave so that my misbehaviour becomes a story for you to narrate at the University Commons tomorrow. But I know better. Here I am, giving you one of my best smiles. What do you say, huh?' When *RB*'s professional enemy saw him, he was looking for some other people nearby. *RB* understood that he was looking for others who would support his foul play. Putting himself in his enemy's shoes *RB* reacted fast. He smiles and greets his arch-enemy as though he is his best friend. Arch-enemy looks flustered and rushes out rather awkwardly. *RB* pats himself silently on his back and then smiles a real smile and as though to reward himself he purchases the largest hilsa in the market in the evening, momentarily thinking of himself as the Nawab who vanquished his arch enemy! This is *RB*'s narrative about himself. *RB* seems to know that he is an agent capable of producing certain effects by his gestures. He can think about what someone else might be thinking of him and act accordingly. His agency certainly depends upon his relationships with the lifeworld but this is a story he is telling about himself. This is his self-narrative about his agency.

Now, consider the following scenario: you are playing a football match and your opponent passes the ball in the opposite direction of the goal, and you are looking for an

explanation as to why he did so. While searching, you notice that none of your opponent's team members is anywhere near the goal. Simulating what you would have done in such a situation leads to the conclusion that it would have been safer to pass the ball in the opposite direction, which corresponded with the performed act. In this example, you use your own mental resources to understand the intentions of others, assuming that they are similar to yours. You are not using any theory to account for your opponent's action. In individual partnership sports like Tennis, Badminton or table tennis this kind of reading of opponents' mind is involved. In these sports, one tries to read the opponent's mind and the opponent also tries to read the mind of her counterpart and develops her game plan accordingly to deceive others. Reading of the mind does not exclude the visual perception of physical movement. So mind in this sense is an embodied mind. This is a Cooperative-competitive situation.

In both scenarios, understanding others take place by participating in the lifeworld. Humans are naturally occurring agents. They are evolution's by-products. The subtle nuances of human beings' behaviour at a particular situation are beyond the comprehension, representative ability and capture of algorithms. Using mental Simulation (Barlassina, Luca and Robert M. Gordon), they can predict or explain others' behaviour.[20] At this juncture we can ask, can an artificial agent do the same? Here we are talking about the empirical and contingent matter which is subject to testing. Given the conception of simulation as embedded in phenomenological notion of lifeworld it does

[20] One may argue, however, how closely are agency and predictability liked in the human context? It is, indeed, true that with the help of predictive tools we, as agents, sometimes fail to predict others. We have seen this in the example of *RB* we have cited above. *RB*'s colleague, perhaps could not predict *RB*'s behaviour. If, in the human context in some cases agency and predictability could not be linked, how could one link these two in the artificially intelligent agents context?

not seem appropriate to think of an artificial agent as a participant in the lifeworld and certainly not an agent of mental simulation.

This is one side of the story. Another thing is that sometimes the creators of Artificial Agents cannot explain their behaviour. We have cited an example of this previously (i.e. Autonomous self-driving car defies certain rules and meets with an accident, the second example was that of two facebook bots who began chatting with each other).

However, from this perspective, as common or folk people it will not be prudent to ascribe agency to them, because their functioning is unknown to the common people and in some cases, it is strange to their creators as well. Technically knowledgeable individuals from specialised disciplines may claim to have special knowledge of 'the mental life of machines' but the matter must be seen through the eyes of the roadside commoner, the employed, the teacher, the parent, and the labour. The question of autonomy for such agents is thus even more far-fetched than their mental lives.

Hence it is hard to accept artificial agents as autonomous agents just like their human counterparts who participate in the lifeworld. If humans do not consider artificial agents to be autonomous agents, then how does the question of replacing human labourers in some situations arise?

Now we want to look at the issue from a different perspective. Someone may argue, however, if AI is intelligent enough to create Artificial Agents, then can these created agents surpass humans? This is the central question of the problem of singularity (Müller, Vincent C. 2021). As stated in the Stanford encyclopaedia of Philosophy, "The

idea of *singularity* is that if the trajectory of artificial intelligence reaches up to systems that have a human level of intelligence, then these systems would themselves have the ability to develop AI systems that surpass the human level of intelligence, i.e., they are 'super intelligent'. Such super-intelligent AI systems would quickly self-improve or develop even more intelligent systems. This sharp turn of events after reaching super intelligent AI is the 'singularity' from which the development of AI is out of human control and hard to predict." (Müller, Vincent C. 2021)

However, Ray Kurzweil (2005) prophesied that in a "Post-Singularity world", homo sapiens could spend much of their time in augmented worlds, which would be nearly identical to reality as we know it. Through statistical equations, Kurzweil anticipates that the Singularity will occur in twenty to twenty-five years. Furthermore, Borna Jalsenjak (2020) mentions super-intelligent AI and the commonalities among the natural and artificial living. In his article, he says, "… once there is an AI which is at the level of human beings and that AI can create a slightly more intelligent AI, and then that one can create an even more intelligent AI, and then the next one creates even more intelligent one and it continues like that until there is an AI which is remarkably more advanced than what humans can achieve." (Jalšenjak B, 2020.) According to Jalsenjak (2020), its algorithm is such that it makes "AI that is not subject-specific, or for the lack of a better word, it is domain less, and as such, it is capable of acting in any domain." (Jalšenjak B, 2020).

Machine learning algorithms, as the author argues, are now written in such a way that they can adapt their behaviour to their surroundings. They evolve as a result of

constant input from their surroundings. With this input, they can alter the algorithm too. But what if we have problems which are non-algorithmic in nature? We know that most psychological and moral problems are non-algorithmic in nature. We have previously discussed this (see Fat-man thought experiment and pretension thought experiment). In order to make a moral decision one needs to develop moral intuition. We have seen that we often cannot reach a decision with the help of our critical reason. In these cases, we appeal to our moral intuition and this paves the way for empathy, fellow feeling and respect for others. Participating in the lifeworld makes it possible for humans.

Now we want to design another thought experiment that will show, how it is difficult to cope with the ever-changing natural world. Imagine a situation where *A*'s laptop is only opened with the help of face recognition technology. To open the laptop, *A* has to sit before it. Moreover, in the post-covid era wearing musk is a must. If the laptop's algorithms are not trained beforehand (Suppose that it was written pre-covid era.), can it recognize the face with musk on it? Can it alter its algorithm itself? How often will it be able to adapt to this ever-changing situation? For example, musk also has different shapes, sizes, designs and dimensions. However, a small alteration in the shape, size, and design is enough to hoodwink the laptop. Hence, the question remains, is it possible for an artificially intelligent agent with facial recognition technology to identify faces with different musk and continuously change its algorithm?

To complicate it further, let us imagine, that person *A* in our thought experiment has a twin sister *B*. They are not only look alike; their physical features too are similar. Even their relatives sometimes cannot differentiate between the two. Now, if the

computer belongs to *A*'s and her twin sister *B* with musk on her face tries to open it, then can the Artificially Intelligent Agent (computer having face recognition techniques in our case) distinguish them?

Our engineer friend, on the other hand, believes that an ideal self-improving AI would be one that could create new algorithms that would result in fundamental improvements. This is known as recursive self-improvement, and it would result in an infinite and accelerating cycle of ever-smarter AI. He can tell us that it could be the digital equivalent of the genetic mutations that organisms undergo over many generations (though highly accelerated, which in itself should implant some seeds of doubts in a discerning mind). In the case of AI the pace is much faster.

If this is the case, we can also ask whether those algorithms can grasp the ever-changing, ever-evolving contents of the lifeworld to keep pace with its own accelerated evolution or will it be more like the blurred vision of a passenger in a superfast train watching the world whizz past the train window? Isn't there always a compromise between pace and cognizance? The lifeworld is the storehouse of value-laden content. For the sake of argument, if we accept that Artificially Intelligent machines will still alter their algorithms in accordance with the natural dynamics of the lifeworld without the assistance of a human person, we must ask whether this is desirable, just as we know that creating an atomic bomb was once plausible, but history teaches us that it was not desirable.

However, it is impossible to build an agent which would be the same as a natural agent in respect of participation in the lifeworld as we have seen that lifeworld, in the Husserlian sense, cannot be fully computed or explained in terms of algorithm.

It could be argued that if the lifeworld has a high degree of algorithmicity, then it can be captured by an explanatory and predictive device like a theory. Then it must be rule-bound in (at least) the most basic sense. If this is the case, it must be highly predictable and programmable. If it is programmable, it can be programmed into a robot or any intelligent artefact. That is correct.

However, if we consider the lifeworld in its essential dynamicity then the only way to capture its sometimes fleeting and sometimes lasting clusters of meaning would be through some kind of participatory process like mental 'simulation'. The computability of the lifeworld is put to question due to its essential subjectivity and uniqueness. One may ask if this does not put into question the predictability of activities in the lifeworld altogether and whether that would be desirable. The absolute unpredictability of the lifeworld as such and the directions in which it may progress or venture in the future is perhaps unquestionable. That does not, however, completely jeopardise the possibilities of statistical prediction in identifiable sections of the lifeworld.

Every moment is a fresh start. Because we are in the present, we cannot predict the next moment because it is unknown to us. We can hope for, expect, and even predict the next moment, but in reality, our prediction, our hope, may remain elusive. Our quest for a new adventure begins as soon as we discover that the reality is quite different than we thought. However, lifeworld plays an important role in shaping our thoughts as we

embark on this new journey and it would be wrong to equate this lifeworld with the physical world. Our experience and thought are the value reservoirs that make human labour a labour.

What if, it is suggested that, the artificial agents might as well join in the lifeworld and enjoy an acceptable degree of inter-subjectivity with the natural agents therein? This too is not to be so.

We can argue that even though the theoretical structure may be built into artificial agents to capture different kinds of regularities in the lifeworld, to attain any degree of inter-subjectivity the artificial agents must possess the capacity to simulate.[21] To qualify as a mental simulator like a natural agent, artificial agents must acquire the ability to generate, grasp and comprehend the content of the lifeworld**.** In the same way, the natural agent, the human, cannot meaningfully interact with the artificial agent through simulation. The bar lies in attempting to put herself in the digital shoes of the artificial agent. This is a serious problem with regard to any interpersonal exchange built upon first-person experience.

Even if the 'agency' of an artificial agent is established by some ethical doctrines, it is clear from this discussion that we cannot ascribe agency to them like the natural agent. There will be an unbridgeable gap between human agents and artificial agents.

[21] One may object that we have examples that Artificial Neural Network proceeds by pattern recognition and simulation and these kinds of computation are not rule-governed. Our answer would be: there are situations in the reality of which no pattern can be generated.

This gap is hard to bridge. However, we can proceed further by accepting this gap and actualizing our goals in more realistic terms.

We have already discussed that artificially intelligent agents too have a history. After Affective computing and Sentic computing, Augmented intelligence (Boschert, Stefan, et al. 2019.) is a new phenomenon to look out for in AI research. In a nutshell, it is a type of artificial intelligence that is a step forward toward a 'more human like' intelligent machine. Augmented intelligence uses machine learning and deep learning to provide humans with actionable data. (Boschert, Stefan, et al. 2019.)

The traditional view of AI, however, conceives Artificially intelligent agents as autonomous systems, which can be operated without human involvement. On the contrary, the recent trend in Artificial Intelligence research relies heavily on human involvement. (Boschert, Stefan, et al. 2019.) Moreover, we have the example of Alexa and Siri, which use augmented intelligence. We have the example of wearable technology like smartwatches which can analyze the data from our body and in that way empower us to take a minute to minute decisions. Where then has the confidence in standalone artificial agency vanished? The question is whether the parasitic dependence of AI technology on human agency can translate into something more symbiotic. So even if we accept that Artificially Intelligent Agents can complement humans, we must watch out for the possible parasitic relation. If we are not that pessimistic, we may at best admit that Artificial Agents can never replace natural agents. Hence we propose that the relationship will be to complement each other.

Science fiction has given us HAL 9000, Borg of *Star Trek*, the robotic assassins of the *Terminator* series or Robo of Satyajit Ray's Professor Shanku. We have a series of stories by Sir Isaac Asimov. Today we even can fancy more and extend our imagination and can dream of more interactive, more intelligent Artificial Agents. The more intelligent machines will require more human involvement otherwise the gap that we have talked about will not be narrowed.

Throughout the thesis, we saw a brief history of technology's journey. We have tried to sketch the journey from Turing's computer to a wearable smartwatch or a driverless automatic car or an Amazon Alexa Voice AI device, a translation bot or even an implant device like a brain chip. There is, of course, no denying the fact that Artificial Agents enabled human advancement and changed our life. They have a deep impact on our workspace as well. But as we have seen, without human involvement it will not prosper. Perhaps that's why recent development shows that there is a turn towards 'embodiment' in artificial intelligent literature. Artificial agents' potentiality will complement human labour, rather than replicate them. To borrow the term from bi-valued logic, we propose that the relationship would be of 'conjunction' (Shramko, Yaroslav and Heinrich Wansing) where if both the operands are true then the entire system will be true and if any one of the operands malfunctions, the entire edifice will collapse like a house of cards.

Additionally, without the co-existence of humans and artificially intelligent agents value generation and value transfer will not be possible. However, in our third chapter, we have explained taking a cue from Marxist literature that human (labour)

generates values and the machine (in my thesis Artificially Intelligent Agents as labourers) transfer them. Hence, one will be paralyzed without the other.

It is a fact that machines cannot generate values; similarly, human labour in some situations can't transfer it like Artificially intelligent agents' pace.

Since the small hyphen between 'life' and 'online life' is getting blurred, our dependence on artificially intelligent agents is skyrocketing. The pandemic situation has exposed very vividly this phenomenon.

However, the covid has exposed that the concept of a monolithic workspace is no more. Outside the periphery of office and home there emerges a third workspace. In the post-covid world, the notion of 'space' is getting constructed and the notion of 'time' has become an all-important factor. Moreover, we have already seen that there is a close relationship between 'time' and productivity.

It is a fact that artificially Intelligent Agents like Alexa or Siri contribute immensely to the management of time. In the post-covid neo-normal 'online' world time has become cynosure rather than space. More dependence may lead to more interaction between humans and artificially intelligent agents, as their relationship is closely connected.

It may be due to this reason, that there is a clamour for the inclusion of humans in the domain of artificially intelligent agents. Hence, the new relationship between human labour and artificial labour is of inclusion and not exclusion.

Thus our research question was, would Artificially Intelligent Agents as labourers replace human labourers in some situations? I have arrived at a point where I can answer this question and our answer is a big no. Artificial agents as labour cannot replace human labourers as their relationship is one of interdependence.

Reference

Abramovitz, Moses. 1993. "The Search for the Sources of Growth: Areas of Ignorance, Old and New." *The Journal of Economic History 53*. no. 2, 217-43. http://www.jstor.org/stable/2122991.

Adams, R.L. 2017. '10 Powerful Examples Of Artificial Intelligence In Use Today', New Jersey: *Forbes*. https://www.forbes.com/sites/robertadams/2017/01/10/10-powerful-examples-of-artificial-intelligence-in-use-today/?sh=3a0daafa420d.

Adams, Tim. 2016 "Artificial intelligence: 'We're like children playing with a bomb'", London: *The Guardian*, https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine.

Aguayo, Carlos. 2020. 'AlphaZero, a novel Reinforcement Learning Algorithm, in JavaScript', *Towards Data Science*, https://medium.com/towards-data-science/alphazero-a-novel-reinforcement-learning-algorithm-deployed-in-javascript-56018503ad18.

Allen, Colin et al. 2000. 'Prolegomena to any future artificial moral agent', *Journal of Experimental and Theoretical Artificial Intelligence, http://commonsenseatheism.com/wp-content/uploads/2009/08/Allen-Prolegomena-to-any-future-artificial-moral-agent.pdf*,

Amoth, D. 2014. 'Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test', *Time*, http://time.com/2847900/eugene-goostman-turing-test/.

Applied Machine Learning at Facebook a Datacenter Infrastructure'. IEEE International Symposium on High Performance Computer. Architecture, https://research.fb.com/wp-content/uploads/2017/12/hpca-2018-facebook.pdf

Asimov, I. 1990. Robot Visions. New York: Penguin Books.

Asimov, Isaac. 1950. *I, Robot*. Greenwich, Connn: Fawcett Publications.

Awad, E. Dsouza, S. Kim, R. *et al.* 2018. 'The Moral Machine experiment'. *Nature* 563, 59–64. https://doi.org/10.1038/s41586-018-0637-6.

Bandyopadhyay, R. and Guha, M. 2008, 'Contextualizing Ethics in the Realm of Robotics', accepted for presentation at the Wesleyan Philosophical Society Conference on 'Philosophy and Science: Contemporary Explorations', held on Thursday, March 13, 2008 at the Duke University Divinity School, USA.
https://www.academia.edu/82396783/Contextualizing_Ethics_in_the_realm_of_Robotics

Bandyopadhyay, Ritajyoti. 2020. 'Migrant Labour, Informal Economy, and Logistics Sector in a Covid-19 World' in Ranabir Samaddar eds *Borders of an epidemic: covid-19 and migrant workers*, Kolkata: CRG Publications.

Bandyopadhyay, Ritaprava, Maushumi Guha and Amita Chjatterjee. 2022. 'The correlation between fall in profit and the deployment of AI labour: A Marxian Analysis', *Antrocom Online Journal of Anthropology* vol. 18. n. 1 (2022) 313-325 – ISSN 1973 – 2880, http://www.antrocom.net/current-issue.html.

Barlassina, Luca and Robert M. Gordon, 'Folk Psychology as Mental Simulation', *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/.

Bhbattacharya, Parimal. 2017. *No Path in Darjeeling is Straight: Memories of a Hill Town*, New Delhi: Speaking Tiger.

Blackcoffer Insights, https://insights.blackcoffer.com/man-and-machines-together-machines-are-more-diligent-than-humans-blackcoffe/.

Boschert, Stefan, et al. 2019. Symbiotic Autonomous Systems, Canada: White Paper III, *IEEE*. https://digitalreality.ieee.org/images/files/pdf/1SAS_WP3_Nov2019.pdf

Brownlee, Jason. 2019. 'A Gentle Introduction to Generative Adversarial Networks (GANs)'. *Machine Learning Mastery*. https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/.

Brynjolfsson, Erik et al. 2017. 'Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics', Chicago: NBER Working Paper No. 24001.
https://siepr.stanford.edu/system/files/Artificial%20Intelligence%20and%20the%20Modern%20Productivity%20Paradox-%20A%20Clash%20of%20Expectations%20and%20Statistics.pdf

By a staff reporter. 2019. 'Google warns rise of AI may backfire on company', London: *The Telegraph*, https://www.telegraph.co.uk/technology/2019/02/11/google-warns-rise-ai-may-backfire-company/

Bynum, Terrell. 2016. Computer and Information Ethics, *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) https://plato.stanford.edu/archives/win2016/entries/ethics-computer/.

Cadwalladr, Carole, 2014, 'Are the robots about to rise? Google's new director of engineering thinks so…', London: *The Guardian*, https://www.theguardian.com/technology/2014/feb/22/robots-google-ray-kurzweil-terminator-singularity-artificial-intelligence

Cambria, Erik. Hussain, Amir, 2021, 'Sentic Computing: A Common-Sense-Based Framework for Concept-Level

Cambridge Analytica, a British consulting firm, collected personal data from millions of Facebook users without their consent, primarily for political advertising. The information was gathered using an app called "This Is Your Digital Life," which was created in 2013 by data scientist Aleksandr Kogan and his company Global Science Research. See Confessore, Nicholas. 2018. 'Cambridge Analytica and Facebook: The Scandal and the Fallout So Far', New York: *The New York Times*.

Capek, Karel. Paul Selver and Nigel Playfair, 1923, *R.U.R. (Rossum's universal robots): a fantastic melodrama in three acts and an epilogue*, French. New York: Harvard (18th ed.).
Case, Anne and Angus, Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity*, Spring 2017. http://www.princeton.edu/~accase/downloads/Mortality_and_Morbidity_in_21st_Century_Case-Deaton-BPEA-published.pdf.
Castree, N. 2002. 'False antitheses? Marxism, nature and actor-networks'. *Antipode* 34, no. 1: 111–46.
Cellan-Jones, Rory, 'Stephen Hawking warns artificial intelligence could end mankind', London: *BBC*. https://www.bbc.com/n
Cellan-Jones, Rory. 2014. 'Stephen Hawking warns artificial intelligence could end mankind', London: *BBC*, https://www.bbc.com/news/technology-30290540.

Chaplin, Charlie. director. 1926. *Modern Times*, United Artists. 87 minutes. https://www.dailymotion.com/video/x3mhpli
Chatterjee, A. Basu, P. and Guha M. 2017. A discussion on Information Ethics at Department of Philosophy, Jadavpur: Jadavpur University,
Chazelle, Damien. Director & Writer. 2016, *La La Land*. Lionsgate.
Churchland, P. 1994, *Matter and Consciousness*, A Bradford Book. p 36. Cambridge: The MIT Press.
Cited in Beyer, Christian. 2020. 'Edmund Husserl', *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/win2020/entries/husserl/.
Coeckelbergh, M. 2009. 'Virtual Moral Agency, Virtual Moral Responsibility: On the Significance of the Appearance, Perception and Performance of Artificial Agents'. *AI and Society* 24 (2), 181–189.
https://link.springer.com/content/pdf/10.1007/s00146-009-0208-3.pdf.
Colestock, H. 2005. *Industrial robotics: selection, design, and maintenance*. New York: McGraw-Hill.
See also, Cubero, S. (Eds). 2007. *Industrial robotics: theory, modelling and control*. Mammendorf: pro literature Verlag Robert Mayer-Scholz.
Committee on Legal Affairs European Parliament. 2016. 'Draft Report with recommendations to the Commission on Civil Law Rules on Robotics', https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.
Committee on Legal Affairs. 2017. Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). Brussels: European Parliament Press, https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.
Crootof, R. 2015. War, 'Responsibility, and Killer Robots'. *North Carolina Journal of International Law and Commercial Regulation*, 40(4), 909-932. https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=2606&context=law-faculty-publications.
Curtis, S. 2016. 'Google driverless car involved in "worst crash yet" after van runs a red light', *Mirror*, http://www.mirror.co.uk/tech/google-driverless-car-involved-worst-8917388

Curtis, S. 2016. 'Google driverless car involved in 'worst crash yet' after van runs a red light', *Mirror*, http://www.mirror.co.uk/tech/google-driverless-car-involved-worst-8917388, last accessed on April, 2017.

Daley, Sam. 2021. '23 Examples of artificial intelligence shaking up business as usual', New York: *Built*.
https://builtin.com/artificial-intelligence/examples-ai-in-industry
Damioli, G. Van Roy, V. & Vertesy, D. 2021. 'The impact of artificial intelligence on labor productivity', *Eurasian Bus Rev* 11, 1–25. https://doi.org/10.1007/s40821-020-00172-8.
Dennett, D. C. 1984. Cognitive Wheels: The Frame problem of AI. In C. Hookway (eds.). *Minds, Machines and Evolution*. , Cambridge: Cambridge University Press.
Dennett, Daniel, C. 1998, *Brainchildren: Essays on Designing Mind*, London: Penguin Books.
Dennett, Daniel. 1995. *Darwin's Dangerous Idea: Evolution And The Meaning of Life*. pp 422-426. New York: Simon & Schuster.
Dennett, Daniel. C. 1976. Conditions of Personhood. Richard Rorty eds, *The Identities of Persons*. Berkeley: University of California Press.
Dennett, Daniel. C. 1992. 'The Self as a Center of Narrative Gravity'. In: F. Kessel, P. Cole and D. Johnson (eds.) *Self and Consciousness: Multiple Perspectives.* Hillsdale, NJ: Erlbaum. Danish translation, "Selvet som fortællingens tyngdepunkt," *Philosophia* 15 275-88, 1986.
URL=http://isites.harvard.edu/fs/docs/icb.topic565657.files/9/Dennett%20self%20as%20center%20of%20gravity.pdf.
Descartes, Rene'. (1637) 1960, *Discourse on Method*, translated by Lawrence LaFleur, NJ: Bobbs Merrill.
Delvaux, Mady. 2016. 'Draft Report with recommendations to the Commission on Civil Law Rules on Robotics', Committee on Legal Affairs European Parliament. Strasbourg: European Parliament Publication,
https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.
Dewhurst, Martin and Willmott, Paul. 2014. 'Manager and machine: The new leadership equation,' *The McKinsey Quarterly*, New York: McKinsey & Co.
Distefano, Joseph n. 2021. 'Inside Amazon's largest warehouse — where you'll find more robots than people', *The Star*,
https://www.thestar.com.my/tech/tech-news/2021/10/19/inside-amazon039s-largest-warehouse-where-you039ll-find-10-robots-for-every-human.

Donagan, A.1977. *The theory of morality*, Chicago: University of Chicago Press.
Driessen, C. & Heutinck, L. F. M. 2015. 'Cows desiring to be milked? Milking robots and the co-evolution of ethics and technology on Dutch dairy farms'. *Agriculture and Human Values*, 32(1), 3-20.
Dunis, C. L. Middleton, P. W., Karathanasopolous, A., & Theofilatos, K. A. (Eds.). 2017. *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*. London: Palgrave Macmillan.
Edmonds, David. 2014. *Would You Kill the Fat Man?: The Trolley Problem and What Your Answer Tells Us About Right and Wrong*, Princeton. New Jersey: Princeton University Press.
Etzioni, Amitai, and Oren Etzioni. 2017. "Incorporating Ethics into Artificial Intelligence." *The Journal of Ethics* 21, no. 4, 403-18. http://www.jstor.org/stable/45204573.
EU. 2016. 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive' 95/46/EC (General Data Protection Regulation), *Official Journal of European Union*, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.
European group on ethics in science and new technologies. 2018. 'Statement on artificial intelligence, robotics and "autonomous" systems', Brussels: EU publications. https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1.
Fjelland, R. 2020. 'Why general artificial intelligence will not be realized'. *Humanit Soc Sci Commun* 7, 10. https://doi.org/10.1057/s41599-020-0494-4.
Fletcher, Joseph. 1966. *Situation ethics: the new morality*. Louisville: Westminster John Knox Press.
Floridi, L. 1999. 'Information ethics: On the philosophical foundation of computer ethics', *Ethics and Information Technology* 1: 33. https://link.springer.com/article/10.1023/A:1010018611096.
Floridi, L. 2006. Information Ethics: Its nature and Scope, *SIGCAS, Computers and Society*, Volume 36. No. 3. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.499.6848&rep=rep1&type=pdf..
Floridi, L. 2008. Foundations of Information Ethics, in Kenneth Einar Himma and Herman T. Tavani eds. *The handbook of Information and Computer Ethics*, New Jersey: Wiley.
Floridi, L. 2010. *The Cambridge handbook of Information and Computer Ethics*. Cambridge: Cambridge University Press.
Floridi, L. and Sanders, J. W., 2004, 'On the Morality of Artificial Agents', *Minds and Machines*. 14(3): 349-379.
Floridi, L. Forthcoming, 'Global Information Ethics: The Importance of Being Environmentally Earnest', accepted for publication in *International Journal of Technology and Human Interaction*, IGI Publishing. http://www.philosophyofinformation.net/wp-content/uploads/sites/67/2014/05/gie.pdf.
Floridi, Luciano, 2014, *The 4th Revolution: How the infosphere is Reshaping Human Reality*, London: OUP.
Floridi, Luciano. 2014. *The 4th Revolution: How the infosphere is Reshaping Human Reality*. London: OUP.
Floridi, Luciano. 2018. 'Soft ethics, the governance of the digital and the General Data Protection Regulation'. *Phil. Trans. R.* Soc. A.3762018008120180081, http://doi.org/10.1098/rsta.2018.0081.
Floridi, Luciano. 2018. 'Soft ethics, the governance of the digital and the General Data Protection Regulation', *Phil. Trans.* R. Soc.A.3762018008120180081http://doi.org/10.1098/rsta.2018.0081.
Floridi, Luciano. 2018. 'Soft ethics, the governance of the digital and the General Data Protection Regulation' *Phil. Trans. R.* Soc. A.3762018008120180081. http://doi.org/10.1098/rsta.2018.0081.
Floridi. L. 2004. Information, in Luciano Floridi eds, *The Blackwell* guide to the *Philosophy of Computing and Information*, pp 40-61. Malden: Blackwell Publishing.
Floridi. L. and Sanders, J. W. 2004. The Method of Abstraction. In *Yearbook of the Artificial. Nature, Culture and Technology. Models in Contemporary Sciences*. Negrotti, M. (ed.), PP 177-220. Bern: Peter Lang.
Floridi. L. 2021, TEDxMaastricht. "The fourth technological revolution", a talk, https://www.oii.ox.ac.uk/news-events/videos/tedxmaastricht-luciano-floridi-the-fourth-technological-revolution/
Foot, Philippa. 1967. 'The Problem of Abortion and the Doctrine of the Double Effect', *Oxford Review*, No. 5. Included in Foot. 1977/2002 *Virtues and Vices and Other Essays in Moral Philosophy*. https://philpapers.org/archive/FOOTPO-2.pdf
Frankena, William, K. 1973. *Ethics*, 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.
Frankena, William, K. 1973. *Ethics*, 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.
Freedman, David, and Paul Humphreys. 1999. "Are There Algorithms That Discover Causal Structure?" *Synthese* 121, no. 1/2 (1999): 29–54. http://www.jstor.org/stable/20118220.
Gareau, B.J. 2005. 'We have never been human: Agential nature, ANT, and Marxist political economy', *Capitalism, Nature, Socialism*, no 16. 4: 127–40
Garisto, Dan. 2019. 'Google AI beats top human players at strategy game StarCraft II'. London: *Nature*. doi: https://doi.org/10.1038/d41586-019-03298-6.

Gert, Bernard and Joshua, Gert. 2020. 'The Definition of Morality'. *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2020/entries/morality-definition/>.
Gibbs, S. 2016. 'Google's self-driving car in broadside collision after other car jumps red light'. London: *The Guardian*. https://www.theguardian.com/technology/2016/sep/26/google-self-driving-car-in-broadside-collision-after-other-car-jumps-red-light-lexus-suv.

Gibbs, Samuel. 2014. 'Elon Musk: artificial intelligence is our biggest existential threat', London: *The Guardian*, https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat.
Glymour, Clark. et al. 1986. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science*. Cambridge. Massachusetts: Academic Press.
Greco, G. M. and Floridi, L. 2004. The Tragedy of the Digital Commons. *Ethics and Information Technology*. 6(2): 73-82.
Griffin, Andrew. 31 July, 2017. 'Facebook's artificial intelligence robots shut down after they start talking to each other in their own language', London: *Independent*, https://www.independent.co.uk/life-style/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html.

Harari, Yuval. 2019. *21 Lessons for the 21st Century*. London: Vintage.
Hazelwood, Kim. 2018. 'Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective
Hepburn, R., 1984, *Wonder and Other Essays*, , Edinburgh: Edinburgh University Press.
Hill, J. Ford. W. R. & Farreras, I. G. 2015. 'Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations'. *Computers in Human Behavior*, Pages 245-250, https://doi.org/10.1016/j.chb.2015.02.026.
Himma, K.E. 2009. 'Artificial Agency, Consciousness and the criteria for moral agency: what properties must an artificial agent have in order to be a moral agent?' *Ethics and Information Technology* 11(1), 19–29.
Homburg, Vincent. 2008, *Understanding E-Government: Information Systems in Public Administration*. p22. Oxon: Routledge.
Hornby, A, S. 2010. *Oxford Advanced Learner's Dictionary of Current English.* Oxford: Oxford University Press.
https://www.brookings.edu/research/the-slowdown-in-manufacturing-productivity-growth/.
https://www.marxists.org/glossary/terms/v/a.htm.
Hubert L. Dreyfus. 1992. *What computers still can't do: a critique of artificial reason*, Cambridge MA: MIT Press.
Ihde, D., 1990, *Technology and the lifeworld*, Bloomington: Indiana University Press.
Illies, C.F.R. & Meijers, A.W.M. 2009. Artefacts without Agency. *The Monist* 92(3), 420–440.
Institute of Technology 37. 52 + vii pp. Stockholm. ISBN 978-91-7415-898-4. https://www.diva-portal.org/smash/get/diva2:410512/FULLTEXT02.pdf.
Ishiguro, Kazuo, 2021, *Klara and the sun*, New York: Alfred A. Knopf.
Ivanov, S. & Webster, C. 2018. 'Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies – a cost-benefit analysis'. In Marinov, V. Vodenska, M. Assenova, M. & Dogramadjieva E. (Eds) *Traditions and Innovations in Contemporary Tourism*, 190-203. Cambridge: Cambridge Scholars Publishing.
Ivanov, S. 2016. 'Will robots substitute teachers?' Paper presented at the 12th International Conference "Modern science, business and education", 27-29 June 2016, Varna University of Management, Vol. 9. pp. 42-47, Bulgaria: *Yearbook of Varna University of Management*.
https://deliverypdf.ssrn.com/delivery.php?ID=435099002092125083123084098078113107105086054036036018076005003025122029006108123023119126122100023056020089109122097017101012006074049005029098079080104074018106073002048123089087007126109006019070125065080086070126027112103108113122023094000094068&EXT=pdf&INDEX=TRUE.
Jalšenjak B, 2020. 'The Artificial Intelligence Singularity: What It Is and What It Is Not', In: Skansi S. (eds) *Guide to Deep Learning Basics*. Cham: Springer. https://doi.org/10.1007/978-3-030-37591-1_10.
Johansson, L. 2011. Robots and Moral Agency. *Theses in Philosophy from the Royal*
Johnson, D. G. Computer Ethics, in Floridi, L., 2004, eds *The Blackwell Guide to Philosophy of Computing and Information*, Oxford: Blackwell Publishing Ltd.
Johnson, Robert and Cureton, Adam. 2017. Kant's Moral Philosophy. *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta eds. https://plato.stanford.edu/archives/spr2017/entries/kant-moral/.
Joseph, Sam and Kawamura, Takahiro. 2001. Why Autonomy Makes the Agent. in Jiming Liu. Ning Zhong. Yuan Y Tang and Patrick S P Wang eds. *Agent Engineering*: *Series in Machine Perception and Artificial Intelligence-Vol. 43*, London: World Scientific.
K. Shahriari and M. Shahriari. 2017. 'IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems,' *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, 2017, pp. 197-201, doi: 10.1109/IHTC.2017.8058187.Wakabayashi, Daisuke. March 19, 2018. 'Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam'. Arizona: *The New York Times*. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html.

Karakilic, Emrah. 2022. 'Why Do Humans Remain Central to the Knowledge Work in the Age of Robots? Marx's *Fragment* on Machines and Beyond', *Work, Employment and Society*, Vol. 36(1) 179–189, California: SAGE Publications. https://journals.sagepub.com/doi/10.1177/0950017020958901
Keski-Äijö, Outi etal. 2021. 'Artificial intelligence from Finland', Helsinki: *Business Finland*. https://toolbox.finland.fi/wp-content/uploads/sites/2/2021/01/bf_ai_from_finland_web.pdf.
Kim, J. 2006. *Philosophy of Mind*. p. 144. Boulder and Oxford: Westview Press.
Konar, Amit. 1999. *Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of Human brain*. Boca Raton. Florida: CRC Press.

Konar, Amit. 23 March, 2016. An interview regarding Agent and Application of Ethical principle in Engineering, by Ritaprava Bandyopadhyay. Jadavpur: Jadavpur University.
Koukku-Rondem, Ritva and Ramos Gabriela. 2022. 'A new global standard for AI ethics', *The Hindu*, June 22, 2022.
Kraut, Richard. 2018. 'Aristotle's Ethics', *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition). Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2018/entries/aristotle-ethics/.
Kröger F. 2016, Automated Driving in Its Social, Historical and Cultural Contexts. In: Maurer M., Gerdes J., Lenz B., Winner H. (eds) *Autonomous Drihving. Springer*, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-48847-8_3
Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology, New York: Viking.*
Latour, Bruno. 1996. "On Actor-network Theory: A Few Clarifications." *Soziale Welt* 47, no. 4: 369-81. http://www.jstor.org/stable/40878163.
Latour, Bruno. 1999. *Pandora's Hope. Essays on the Reality of Science Studies*, Cambridge, MA: Harvard University Press.
Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford: Oxford University Press.
Leonelli, Sabina. 2020. "Scientific Research and Big Data", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2020/entries/science-big-data/.
Li G. Zhang D. 2017, 'Brain-Computer Interface Controlling Cyborg: A Functional Brain-to-Brain Interface Between Human and Cockroach'. In Guger C. Allison B. Ushiba J. (eds) *Brain-Computer Interface Research*, SpringerBriefs in Electrical and Computer Engineering. Cham: Springer. https://doi.org/10.1007/978-3-319-57132-4_6.
Lokitz, Justin. 2021. 'The future of work: How humans and machines are evolving to work together', *BMI*, San Francisco, https://www.businessmodelsinc.com/machines/
Lyu, Yitian & Zhang, Chenrui. 2019, 'Slowing Economic Growth around the World in the 21st Century'. *Open Journal of Business and Management*. 07. 1926-1935. 10.4236/ojbm.2019.74131.
Magnani, L. & Bardone, E. 2008. 'Distributed morality: Externalizing ethical knowledge in technological artifacts', *Foundations of Science*, 13(1), 99–108. URL =https://link.springer.com/article/10.1007/s10699-007-9116-5 (Last seen 30.04.2017 at 2.56 pm).
Makridakis, S. 2017. The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. *Futures*, Pages 46-60, https://doi.org/10.1016/j.futures.2017.03.006. (https://www.sciencedirect.com/science/article/pii/S0016328717300046).
Mani, Chithrai. 2020, 'How Is Big Data Analytics Using Machine Learning?', New Jersey: *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2020/10/20/how-is-big-data-analytics-using-machine-learning/?sh=312c08a271d2.
Marx, K.1845. *Theses on Feuerbach*, Marxists Internet Archive Encyclopedia http://www.marxists.org/archive/marx/works/1845/theses/index.htm,
Marx, K.1845. *Theses on Feuerbach*, Marxists Internet Archive Encyclopedia. http://www.marxists.org/archive/marx/works/1845/theses/index.htm
Marx, K.1845. *Theses on Feuerbach*. http://www.marxists.org/archive/marx/works/1845/theses/index.htm, accessed February 2022.
Marx, Karl, 1867 (1981), *Capital: A Critique of Political Economy*, Penguin
Marx, Karl. 1867 (1887). *Capital*, Volume I, Chpt. 1: Section 4, cited in, Marxists Internet Archive Encyclopedia.
Marx, Karl. 1867 (1981). *Capital: A Critique of Political Economy*. London: Penguin. p. 356. https://www.marxists.org/archive/marx/works/1894-c3/ch13.htm
Marx, Karl. 1867 (1981). Wage Labour and Capital, What are Wages? How are they Determined? *Capital: A Critique of Political Economy, Penguin, https://www.marxists.org/archive/marx/works/1847/wage-labour/ch02.htm*
Marx, Karl. 1996. *Das Kapital*. Edited by Friedrich Engels. Washington, D.C. DC: Regnery Publishing. https://www.marxists.org/archive/marx/works/1865/value-price-profit/ch02.htm#c6.

Maurer, M. Gerdes, J. C. Lenz, B. & Winner, H. (Eds.). 2016. *Autonomous driving: technical, legal and social aspects*. Berlin, Heidelberg: Springer Open. https://unglueit-files.s3.amazonaws.com/ebf/883c597a81c34af6ad130c08ead0c4d6.pdf.
Mayer-Schnberger. Viktor, Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray Publishers.
McIntyre, Alison. 2019. 'Doctrine of Double Effect', *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition). Edward N. Zalta (ed.). https://plato.stanford.edu/archives/spr2019/entries/double-effect.
Mechatronics is a multidisciplinary branch of engineering that focuses on a combination of AI and deep learning, data science, sensor technology, Internet of Things, mechanical and electrical engineering. See, Winterstein, Dave, 2022, 'Innovative projects prepare students for 'real' engineering'. *Cornell Chronicle*. NY: Ithaca. https://news.cornell.edu/stories/2022/02/innovative-projects-prepare-students-real-engineering.
Min, H. 2010. 'Artificial intelligence in supply chain management: theory and applications'. *International Journal of Logistics Research and Applications*, 13(1), 13-39.
Moor, J. 1985, 'What is computer Ethics?' *Metaphilosophy*, 16 (4): 266-75. https://web.cs.ucdavis.edu/~koehl/Teaching/ECS188_F20/PDF_files/MOOR-1985-Metaphilosophy.pdf
Moor, J. 2006. 'The Nature, Importance and Difficulty of Machine Ethics'. *Intelligent Systems, IEEE* 21(4). pp. 18–21.
Morton, Adam. 2003, *The Importance of Being Understood: Folk Psychology as Ethics*, New York: Routledge.

Müller, Vincent C. 2021, 'Ethics of Artificial Intelligence and Robotics', *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/.

Nagel, T. .1974. 'What is it like to be a bat?' *The Philosophical Review*, Vol. 83, http://www2.warwick.ac.uk/fac/cross_fac/iatl/activities/modules/ugmodules/humananimalstudies/lectures/32/nagel_bat.pdf.
Naomi, Altman, et al. 2012. 'Association, correlation, and causation', *Nature*, https://www.nature.com/articles/nmeth.3587.pdf
On 12th January 2017, the European Parliament Committee on Legal Affairs moved by 17 votes to 2 to approve a draft report
Pearl, Judea and Mackenzie, Dana. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books, Inc. Division of Harper Collins.
Peterson, M. & Spahn, A. 2011. Can Technological Artefacts Be Moral Agents? *Science and Engineering Ethics*, 17 (3), 411-424, http://download.springer.com/static/pdf/814/art%253A10.1007%252Fs11948-010-9241-3.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs11948-010-9241-3&token2=exp=1493791100~acl=%2Fstatic%2Fpdf%2F814%2Fart%25253A10.1007%25252Fs11948-010-9241-3.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.springer.com%252Farticle%252F10.1007%252Fs11948-010-9241-3*~hmac=d109cb1983cc858216ee457c6ec3413ff9daf773790344a31a42de8cb5a72f14 (last seen on 03.05.2017 at 11.11 am)
Polanyi, Michael. 1958. *Personal knowledge, towards a post-critical philosophy*. Chicago: University of Chicago Press.
Nathalie Nevejans. 2018. AI/robotics researchers and industry leaders, Physical and Mental Health specialists, Law and Ethics experts gathered to voice our concern about the negative consequences of a legal status approach for robots in the European Union, 2018, 'Open Letter To The European Commission Artificial Intelligence And Robotics', *Politico*, https://www.politico.eu/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf.
Popper, K. 1966. *The open society and its enemies* (5th ed.). Oxfordshire: Routledge and Kegan Paul.
Powers, T.M. 2006. Prospects for a Kantian Machine. *IEEE Intelligent Systems* 21(4), 46–51. See also Grau, C. 2006. 'There is no I in Robot: Robots and Utilitarianism'. *IEEE Intelligent Systems* 21(4), 52–55.
PTI. 2010. 'Facebook promises new measures to combat 'trolling'', London*: The Hindu.*
Purdy, Mark and Daugherty, Paul. 2017. 'How AI boosts industry spotlights industry profits and innovation', Dublin: *Accenture*, https://www.accenture.com/fr-fr/_acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf
Radhakrishnan, S. 2011. *The Bhagavadgita*. Noida: HarperCollins.
Raredon, J and Blais. M. 1998.The Frame Problem in Artificial Intelligence. http://groups.engin.umd.umich.edu/CIS/course.des/cis479/projects/frame/welcome.html last seen on 01.05.2017.
Reed, Mallory. 2018. "The Classification of Artificial Intelligence as 'Social Actors'." Thesis. Georgia: Georgia State University. Accessed June 30, 2021. https://scholarworks.gsu.edu/rs_theses/58.
Remus, D. & Levy, F. 2015. 'Can robots be lawyers? Computers, lawyers, and the practice of law'. SSRN Working paper: http://ssrn.com/abstract=2701092.
Roy, Anna. 2021. 'Responsible AI, #AIFORALL, Approach Document for India, Part 1 – Principles for Responsible AI', New Delhi: Niti Aayog. https://www.readkong.com/page/responsible-ai-aiforall-approach-document-for-india-part-2179495.
Roy, Priyanka. 2021. 'Preparing for India's new personal data protection law', New Delhi: *The Times of India*.
Russell, S. J., & Norvig, P. 1995. Artificial intelligence: A modern approach. Englewood Cliffs, NJ: Prentice Hall.
Florian, R˘azvan V. (2003). 'Autonomous Artificially Intelligent Agents', Technical Report Coneural, Center for Cognitive and Neural Studies, https://neuro.bstu.by/my/Tmp/Florean/Coneural-03-01.pdf.

Rabindranath Tagore. 1932. *Gitabitan*, Kolkata: Viswabharati.

Sakovich, Natallia. 2020. 'The Ultimate Guide to Enterprise AI', Texas: SaM Solutions, https://www.sam-solutions.com/blog/enterprise-artificial-intelligence/.
Sam, Kriegman, Douglas Blackiston, Michael Levin and Josh Bongard, 2021, 'Kinematic self-replication in reconfigurable organisms', *Proceedings of the National Academy of Sciences*; 118 (49): e2112672118 DOI: 10.1073/pnas.2112672118.
Satwant, Kaur. 2012. 'How Medical Robots are going to Affect Our Lives', *IETE Technical Review*, 29:3, 184-187, DOI: 10.4103/0256-4602.98859.
Sayes, Edwin. 2017. 'Marx and the critique of Actor-Network Theory: mediation, translation, and explanation', Distinktion. *Journal of Social Theory*. http://dx.doi.org/10.1080/1600910X.2017.1390481.
Schlosser, Markus. 2019. 'Agency'. *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2019/entries/agency/.

Schlosser, Markus. 2019. winter edition. 'Agency'. Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2019/entries/agency/>.
Searle, John. R. 1980. Minds, brains, and programs, *Behavioral and Brain Sciences* 3 (3): 417-457. http://cogprints.org/7150/1/10.1.1.83.5248.pdf
Sentiment Analysis'. Amir Hussain Eds *Socio-Affective Computing* Volume 1, Springer, London, DOI 10.1007/978-3-319-23654-4.

Sewell, Robb. 2012. 'The Capitalist Crisis and the Tendency of the Rate of Profit to Fall'. *In Defence of Marxism*. https://www.marxist.com/the-capitalist-crisis-and-the-tendency-of-the-rate-of-profit-to-fall-full.htm.
Shanahan. Murray. 2016. 'The Frame Problem', *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.). https://plato.stanford.edu/archives/spr2016/entries/frame-problem/.
Shramko, Yaroslav and Heinrich Wansing, 'Truth Values', *The Stanford Encyclopedia of Philosophy* (Win0ter 2020 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/win2020/entries/truth-values/.
Shramko, Yaroslav and Heinrich Wansing. 2020. 'Truth Values', *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/truth-values/>.
Silver, David. et al. 2018. 'A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play', *Science*, Vol. 362, Issue 6419, pp. 1140-1144, DOI: 10.1126/science.aar6404.
Steedman, Ian. 1977. *Marx after Sraffa*. London: Verso Books.

Stern, R. 2004, 'Does 'Ought' Imply 'Can'? And Did Kant Think It Does?', *Utilitas*, 16 (1). pp. 42-61. ISSN 0953-8208, https://doi.org/10.1017/S0953820803001055.
Stuart, Mill John. 1957 [1861]. *Utilitarianism*, Indianapolis: The Liberal Arts press.
Sullins, J. P. 2006. 'When is a robot a moral agent?', *International Review of Information Ethics* 6 (12):23-30, URL= http://www.i-r-i-e.net/inhalt/006/006_Sullins.pdf.
Swapna. G and Nivashiniya. R. 2021. 'Will machine replace the human in the future of work?'
Syverson, Chad. 2016. 'The slowdown in manufacturing productivity growth', Washington, DC: Brookings.
Taylor, Josua and Bringsjord, Selmer. 2012. The Divine-Command Approach to Robot Ethics in Lin. Patrick. et al. eds *Robot Ethics: The Ethical And Social Implication of Robotics.* Cambridge: MIT Press.
Thomson, Judith Jarvis. 1985. 'The Trolley Problem'. *The Yale Law Journal* 94, no. 6 (1985): 1395–1415. https://doi.org/10.2307/796133.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 49: 433-460. https://www.csee.umbc.edu/courses/471/papers/turing.pdf.
UNESCO and COMSET. 2015. 'Preliminary Work Programme of COMEST for 2014-2015: Potential topics for reflection', Paris: UNESCO. https://unesdoc.unesco.org/search/N-EXPLORE-51febc6a-534e-436c-acd5-c3791570dd19.
UNESCO. 2022. 'Recommendation on the Ethics of Artificial Intelligence'. Adopted on 23 November 2021. Paris: The United Nations Educational, Scientific and Cultural Organization. https://en.unesco.org/artificial-intelligence/ethics.
Unudurti, Jaideep. 2021. 'Journals of the plague year: The books of 2021 straddle two worlds, pre-and post-pandemic'. *The Hindu*. 24 December, 2021. https://www.thehindu.com/books/journals-of-the-plague-year-the-books-of-2021-straddle-two-worlds-pre-and-post-pandemic/article38026530.ece.
Verbeek, P. P. 2006, 'Materializing morality: Design ethics and technological mediation', *Science, Technology & Human Values*, 31(3), 361–380. https://academics.design.ncsu.edu/student-publication/wp-content/uploads/2016/11/Verbeek_DesignEthics.pdf.
Verbeek, P. P. 2006, 'Materializing morality: Design ethics and Technological Mediation'. *Science Technology & Human Values* 31, 361–380.
Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency and Design*, University Park, Pennsylvania: The Pennsylvania State University Press.
Voort, V. D. Pieters, W. and Consoli L. 2015. 'Refining the ethics of computer-made decisions: a classification of moral mediation by ubiquitous machines', *Ethics and Information Technology*, Volume 17, Issue 1, pp 41–56, https://link.springer.com/article/10.1007/s10676-015-9360-2 (last seen 30.4.2017 at 2.34 pm)
Wallach, W. & Allen, C.2009. *Moral Machines – Teaching Robots Right from Wrong*. Oxford: Oxford University Press.
Waterson, Jim. 2020. 'Microsoft's robot editor confuses mixed-race Little Mix singers', London: *The Guardian*. https://www.theguardian.com/technology/2020/jun/09/microsofts-robot-journalist-confused-by-mixed-race-little-mix-singers
Webster, C. & Ivanov, S. 2019. 'Robotics, Artificial Intelligence, and the Evolving Nature of Work'. George, B. & Paul, J. (Eds.) *Digital Transformation in Business and Society*, 127–143. Palgrave-MacMillan, London, https://doi.org/10.1007/978-3-030-08277-2_8.
Webster, C., Ivanov, S. 2020. 'Artificial Intelligence, and the Evolving Nature of Work'. In: George, B., Paul, J. (eds) *Digital Transformation in Business and Socie*ty. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-08277-2_8.
Weiß, Gerhard. 2012. Agent Orientation in Software Engineering. Cambridge: *Knowledge Engineering Review*. https://pdfs.semanticscholar.org/d6fe/cb59952c5593c310e230b87e04cadeba57ef.pdf.
White, Hylton. 2013. 'Materiality, Form, and Context: Marx Contra Latour'. *Victorian Studies* 55. no. 4: 667-82. Accessed July 1, 2021. doi:10.2979/victorianstudies.55.4.667.
Woollard, Fiona. Frances, Howard-Snyder. Fall 2021 Edition, 'Doing vs. Allowing Harm', *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.). https://plato.stanford.edu/archives/fall2021/entries/doing-allowing/.
Young, Fredric C. 1979. On Dennett's Conditions of Personhood, *Auslegung*, https://kuscholarworks.ku.edu/bitstream/handle/1808/8944/auslegung.v06.n03.161-177.pdf;sequence=1.

Soft-morality and artificiality

By: Ritaprava Bandyopadhyay

As of: Nov 3, 2022 8:34:47 AM

65,033 words - 346 matches - 140 sources

Similarity Index

7%

sources:

768 words / 1% - Internet
Fjelland, Ragnar. "Why general artificial intelligence will not be realized", 'Springer Science and Business Media LLC', 2020
9 words / < 1% match - Internet
index:. "NOTES TOWARDS A SPATIAL READING OF MARX'S FRAGMENT ON MACHINES", University of North Carolina at Chapel Hill Graduate School, 2017
8 words / < 1% match - Internet
MacInnis, Luke. "The Unity of Political Principle", 'Columbia University Libraries/Information Services', 2014
164 words / < 1% match - Internet from 29-Aug-2020 12:00AM
mafiadoc.com
85 words / < 1% match - Internet from 20-Aug-2020 12:00AM
mafiadoc.com
231 words / < 1% match - Internet from 15-Aug-2022 12:00AM
www.nature.com
15 words / < 1% match - Internet from 08-Oct-2020 12:00AM
www.nature.com
218 words / < 1% match - Internet
Illari, PK, Allo, P et al. "An introduction to the Philosophy of Information", The Society for the Philosophy of Information, 2013
215 words / < 1% match - Internet from 20-Aug-2018 12:00AM
www.nber.org
203 words / < 1% match - Internet from 13-Mar-2019 12:00AM
ar.scribd.com
147 words / < 1% match - Crossref
B. Czarniawska. "Book Review: Bruno Latour: Reassembling the Social: An Introduction to Actor-Network Theory", Organization Studies, 07/19/2006
137 words / < 1% match - Crossref
Emrah Karakilic. " Why Do Humans Remain Central to the Knowledge Work in the Age of Robots? Marx's and Beyond ", Work, Employment and Society, 2020
126 words / < 1% match - Internet from 03-Dec-2020 12:00AM
www.theguardian.com
115 words / < 1% match - Internet from 29-Oct-2022 12:00AM
seop.illc.uva.nl
52 words / < 1% match - Internet from 21-May-2020 12:00AM
link.springer.com
11 words / < 1% match - Internet from 21-Nov-2017 12:00AM
link.springer.com
9 words / < 1% match - Internet from 24-Dec-2019 12:00AM

link.springer.com
9 words / < 1% match - Internet from 10-May-2022 12:00AM
link.springer.com
61 words / < 1% match - Internet
Floridi, Luciano. "Information ethics, its nature and scope", 2006
10 words / < 1% match - Internet from 27-Nov-2020 12:00AM
philpapers.org
46 words / < 1% match - Internet from 04-Oct-2021 12:00AM
ndl.ethernet.edu.et
17 words / < 1% match - Internet from 23-Sep-2022 12:00AM
ndl.ethernet.edu.et
56 words / < 1% match - Crossref
Altman, Naomi, and Martin Krzywinski. "Points of Significance: Association, correlation and causation", Nature Methods, 2015.
52 words / < 1% match - Internet from 24-Oct-2022 12:00AM
github.com
52 words / < 1% match - Internet from 01-Aug-2022 12:00AM
kapua.gilead.org.il
48 words / < 1% match - Internet from 04-Dec-2020 12:00AM
www.forbes.com
48 words / < 1% match - Internet from 16-May-2020 12:00AM
www.mcrg.ac.in
47 words / < 1% match - Internet from 14-Dec-2020 12:00AM
bdtechtalks.com
44 words / < 1% match - Crossref
Erin Koch. "Negotiating “The Social” and Managing Tuberculosis in Georgia", Journal of Bioethical Inquiry, 2016
43 words / < 1% match - Internet from 24-Oct-2022 12:00AM
www.antrocom.net
42 words / < 1% match - Internet from 24-Nov-2020 12:00AM
futurism.com
38 words / < 1% match - Crossref
Steve Guglielmo, Andrew E. Monroe, Bertram F. Malle. "At the Heart of Morality Lies Folk Psychology", Inquiry, 2009
27 words / < 1% match - Internet from 01-Oct-2020 12:00AM
archive.org
11 words / < 1% match - Internet from 16-Jan-2020 12:00AM
archive.org
38 words / < 1% match - Internet from 21-Oct-2022 12:00AM
www.coursehero.com
37 words / < 1% match - Internet from 21-Jan-2020 12:00AM
www.tandfonline.com
11 words / < 1% match - Internet from 18-Mar-2019 12:00AM
epdf.tips
9 words / < 1% match - Internet from 24-Jan-2019 12:00AM
epdf.tips
8 words / < 1% match - Internet from 28-Feb-2019 12:00AM
epdf.tips
8 words / < 1% match - Internet from 21-Feb-2019 12:00AM
epdf.tips
36 words / < 1% match - Internet from 29-Mar-2020 12:00AM
www.sam-solutions.com
33 words / < 1% match - Internet from 20-Jun-2015 12:00AM
surplusvalue.org.au

32 words / < 1% match - Internet from 08-Oct-2022 12:00AM
dspace.bracu.ac.bd
29 words / < 1% match - Internet from 27-Jun-2019 12:00AM
andrelemos.info
29 words / < 1% match - Internet from 30-Jan-2016 12:00AM
dspace.lboro.ac.uk
28 words / < 1% match - ProQuest
Gillman, Lin. "The psychospiritual dimensions of living with life-threatening illness", Proquest, 2014.
28 words / < 1% match - Internet from 11-Aug-2019 12:00AM
pagetectonics.com
26 words / < 1% match - Internet from 19-Mar-2021 12:00AM
lea.vitis.uspnet.usp.br
23 words / < 1% match - Internet from 30-May-2021 12:00AM
www.slideshare.net
22 words / < 1% match - Internet from 12-Dec-2021 12:00AM
www.themusicallyrics.com
20 words / < 1% match - Crossref
Murray Smith. "Invisible Leviathan", Brill, 2019
20 words / < 1% match - Internet from 28-Apr-2016 12:00AM
issuu.com
19 words / < 1% match - Internet from 23-Aug-2019 12:00AM
oecdinsights.org
19 words / < 1% match - Internet from 24-Oct-2014 12:00AM
www.philosophyofinformation.net
18 words / < 1% match - Crossref
Elena Louisa Lange. "Value without Fetish", Brill, 2021
18 words / < 1% match - Internet from 25-May-2019 12:00AM
index-of.co.uk
18 words / < 1% match - Internet from 04-Jan-2021 12:00AM
termer.net
9 words / < 1% match - Internet from 26-Jan-2022 12:00AM
ebin.pub
8 words / < 1% match - Internet from 11-Oct-2022 12:00AM
ebin.pub
16 words / < 1% match - Publications
Accounting Research Journal, Volume 27, Issue 1
16 words / < 1% match - Crossref
Satinder P. Gill. "Tacit Engagement", Springer Nature, 2015
16 words / < 1% match - Internet from 06-Oct-2022 12:00AM
resistir.info
15 words / < 1% match - Crossref
Collective Agency and Cooperation in Natural and Artificial Systems, 2015.
15 words / < 1% match - ProQuest
Davies, Jeannine A. "An experiential inquiry into the intersubjective principles of Relational Dharma as exemplified in Aung San Suu Kyi's spiritual revolution", Proquest, 20111109
15 words / < 1% match - Internet from 18-Oct-2022 12:00AM
www.docslides.com
13 words / < 1% match - Crossref
Abdalsamad Keramatfar, Hossein Amirkhani. "Bibliometrics of sentiment analysis literature", Journal of Information Science, 2018
13 words / < 1% match - Internet from 17-May-2003 12:00AM
www.worldalternative.org
12 words / < 1% match - Crossref
Andrew Feenberg. "Peter-Paul Verbeek: Review of What Things Do", Human Studies, 2009

12 words / < 1% match - Crossref
Ben Fowkes. "Marx's Economic Manuscript of 1864-1865", Brill, 2015
12 words / < 1% match - ProQuest
Freeman, Lauren. "Ethical dimensions in Martin Heidegger's early thinking", Proquest, 20111003
12 words / < 1% match - Crossref
Henryk Grossman. "Henryk Grossman Works, Volume 1", Brill, 2019
12 words / < 1% match - Crossref
IFIP Advances in Information and Communication Technology, 2009.
12 words / < 1% match - Crossref
Jim S. Dolwick. "'The Social' and Beyond: Introducing Actor-Network Theory", Journal of Maritime Archaeology, 04/28/2009
12 words / < 1% match - ProQuest
Ranade, Nupoor. "Re-Contextualizing Audiences: New Conceptualizations of User Interactions in Product Documentation Spaces", North Carolina State University, 2021
12 words / < 1% match - Internet from 01-Feb-2020 12:00AM
www.oapen.org
11 words / < 1% match - Crossref
"Advances in Water Resources Management for Sustainable Use", Springer Science and Business Media LLC, 2021
11 words / < 1% match - Crossref
George Henderson. "Marxist Political Economy and the Environment", A Companion to Environmental Geography, 01/30/2009
11 words / < 1% match - Crossref
James Steinhoff. "Automation and Autonomy", Springer Science and Business Media LLC, 2021
11 words / < 1% match - Internet from 05-Mar-2019 12:00AM
www.geetabitan.com
10 words / < 1% match - Crossref
Subramaniam Meenakshi Sundaram, Tejaswini R. Murgod, Sowmya M.. "chapter 1 Artificial Intelligence and Data Analytics-Based Emotional Intelligence", IGI Global, 2022
10 words / < 1% match - Internet from 15-Mar-2018 12:00AM
www.scribd.com
9 words / < 1% match - Crossref
"Social Informatics", Springer Science and Business Media LLC, 2017
9 words / < 1% match - Crossref
Beyond Capital, 1992.
9 words / < 1% match - Crossref
Carsten Herrmann-Pillath, Juha Hiedanpää, Katriina Soini. "The co-evolutionary approach to nature-based solutions: A conceptual framework", Nature-Based Solutions, 2022
9 words / < 1% match - ProQuest
Daou, Jean Georges. "Personhood: An ethical understanding", Proquest, 20111003
9 words / < 1% match - ProQuest
King, Kathryn Real. "Corporations as group agents/responsible collectives in theory and in practice", Proquest, 20111003
9 words / < 1% match - Internet from 14-Jul-2020 12:00AM
citeseerx.ist.psu.edu
9 words / < 1% match - Internet from 08-Mar-2016 12:00AM
community.seidenberg.pace.edu
9 words / < 1% match - Internet from 30-Aug-2020 12:00AM
content.sciendo.com
9 words / < 1% match - Internet from 02-Nov-2022 12:00AM
libcom.org
9 words / < 1% match - Internet from 01-Mar-2022 12:00AM
opus.uni-hohenheim.de
9 words / < 1% match - Internet from 05-Jun-2018 12:00AM

research.gold.ac.uk
9 words / < 1% match - Internet from 28-Jun-2022 12:00AM
storage.googleapis.com
8 words / < 1% match - Crossref
"David Harvey", Wiley, 2006
8 words / < 1% match - Crossref
"Ethics and Policies for Cyber Operations", Springer Science and Business Media LLC, 2017
8 words / < 1% match - Crossref
"Human and Machine Learning", Springer Science and Business Media LLC, 2018
8 words / < 1% match - Crossref
"Marx Today", Springer Science and Business Media LLC, 2010
8 words / < 1% match - Crossref
"Service Excellence in Tourism and Hospitality", Springer Science and Business Media LLC, 2021
8 words / < 1% match - Crossref
"The Hegel-Marx Connection", Springer Science and Business Media LLC, 2000
8 words / < 1% match - Crossref
Bob Milward. "Marxian Political Economy", Springer Science and Business Media LLC, 2000
8 words / < 1% match - Publications
Bowling, Ann, Ebrahim, Shah. "EBOOK: Handbook of Health Research Methods: Investigation, Measurement and Analysis", EBOOK: Handbook of Health Research Methods: Investigation, Measurement and Analysis, 2005
8 words / < 1% match - Crossref
Brasilina Passarelli, Fabiana Grieco Cabral de Mello Vetritti. "chapter 11 #ConnectedYouthBrazil Research: Emerging Literacies in a Hyperconnected Society", IGI Global, 2016
8 words / < 1% match - Crossref
Christian Fuchs, Vincent Mosco. "Marx in the Age of Digital Capitalism", Brill, 2016
8 words / < 1% match - ProQuest
Fisher, Joseph Andrew. "Enhancing 'Human Nature': The Human Enhancement Debate in U.S. Bioethics", Columbia University, 2021
8 words / < 1% match - Crossref
Luca M. Possati. "Freud and the algorithm: neuropsychoanalysis as a framework to understand artificial general intelligence", Humanities and Social Sciences Communications, 2021
8 words / < 1% match - Crossref
Mark Skilton, Felix Hovsepian. "The 4th Industrial Revolution", Springer Science and Business Media LLC, 2018
8 words / < 1% match - Crossref
Martin Mullins, Christopher P. Holland, Martin Cunneen. "Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market", Patterns, 2021
8 words / < 1% match - ProQuest
Roncancio, Ivan Dario Vargas. "The Legal Lives of Forests: Law and the Other-Than-Human in the Andes-Amazon, Colombia (An Anthropological and Legal Theory Approach)", McGill University (Canada), 2021
8 words / < 1% match - Crossref
Sean Sayers. "Marx and Alienation", Springer Science and Business Media LLC, 2011
8 words / < 1% match - ProQuest
Showler, Paul. "Pragmatism, Genealogy, and Moral Status", University of Oregon, 2022
8 words / < 1% match - Crossref
Steve Torrance. "Ethics and consciousness in artificial agents", AI & SOCIETY, 2007
8 words / < 1% match - ProQuest
Thomas, Lisa. "Towards "After-Modern" Design: A Practice-Based Inquiry", Lancaster University (United Kingdom), 2020
8 words / < 1% match - Crossref
Upamanyu Pablo Mukherjee. "Postcolonial Environments", Springer Science and Business Media LLC, 2010
8 words / < 1% match - Internet from 03-May-2022 12:00AM
aceh.b-cdn.net

8 words / < 1% match - Internet from 11-Jul-2013 12:00AM
communism.blogsport.eu
8 words / < 1% match - Internet from 02-Dec-2021 12:00AM
cris.maastrichtuniversity.nl
8 words / < 1% match - Internet from 22-Nov-2018 12:00AM
ebooks.bharathuniv.ac.in
8 words / < 1% match - Internet from 11-Apr-2022 12:00AM
epdf.pub
8 words / < 1% match - Internet
Baygeldi, Murat. "Social construction of IS evaluation: a case study of IT investment appraisal", 2011
8 words / < 1% match - Internet from 28-Jun-2022 12:00AM
nanopdf.com
8 words / < 1% match - Internet from 06-Aug-2022 12:00AM
participedia.net
8 words / < 1% match - Internet from 01-Apr-2019 12:00AM
pure.uvt.nl
8 words / < 1% match - Internet from 02-Oct-2022 12:00AM
research.vu.nl
8 words / < 1% match - Internet from 20-Oct-2022 12:00AM
researcharchive.vuw.ac.nz
8 words / < 1% match - Internet from 30-Oct-2022 12:00AM
www.aeaweb.org
8 words / < 1% match - Internet
Jemsek, Misha Frame. "Heat Pumps and Household Energy Consumption in Norway: An actor-network and practice theory approach"
8 words / < 1% match - Internet from 28-Sep-2022 12:00AM
www.frontiersin.org
8 words / < 1% match - Internet from 19-Nov-2020 12:00AM
www.marxist.com
8 words / < 1% match - Internet from 14-May-2019 12:00AM
www.parliament.uk
8 words / < 1% match - Internet from 29-Oct-2022 12:00AM
www.secureworld.io
8 words / < 1% match - Internet from 21-Feb-2015 12:00AM
z-ecx.images-triumphmotorcycle.nickelgarage.com
7 words / < 1% match - Crossref
"Handbook of Technology Application in Tourism in Asia", Springer Science and Business Media LLC, 2022
7 words / < 1% match - Crossref
Massimo Durante. "Ethics, Law and the Politics of Information", Springer Science and Business Media LLC, 2017
7 words / < 1% match - Internet from 09-Dec-2020 12:00AM
newsroom.accenture.com
6 words / < 1% match - Crossref
"The Blackwell Guide to the Philosophy of the Social Sciences", Wiley, 2003
6 words / < 1% match - Crossref
John Weeks. "Capital and Exploitation", Walter de Gruyter GmbH, 1982
6 words / < 1% match - Crossref
Philosophy of Engineering and Technology, 2012.
6 words / < 1% match - ProQuest
Salour, Sam. "Marx's Critique of Bourgeois World.", University of California, Santa Barbara, 2021
6 words / < 1% match - Internet from 30-Nov-2020 12:00AM
thenextrecession.wordpress.com
4 words / < 1% match - Internet from 25-Jul-2020 12:00AM
cdn.ymaws.com