

Dissertation

Masters in Business Administration

Oliver Grogan Student ID: 22103597

Artificial Intelligence (AI) Governance in the technology sector: Can we anticipate the future and govern the present?

School of Business National College of Ireland

Supervisor: David Hurley

Abstract

Artificial Intelligence (AI) has experienced a massive growth in development and usage over the last decade, transforming global industries and workplaces. The pace of change caused by AI has been staggering, challenging regulatory bodies to keep pace with its progress. This thesis explores the obstacles facing the relevant regulatory bodies including difficulties defining AI because of technology's broad nature, the governance difficulties caused by its many applications, and its rapid pace of development. The research evaluates the regulatory approach currently adopted by the US, EU, and China, assessing whether industry specific strategies may be more successful than the current systems and whether current enforcement mechanisms will be successful in protecting people's privacy and ensuring the ethical development of the technology. The study also considers the future, seeking ways to overcome theories such as the Collingridge Dilemma and technological determinism that predict the technology is likely to follow its own path and by the time its pitfalls are realised, they will be difficult to govern.

Using a qualitative methodology, the research focuses on understanding the perspective of technology sector professionals to garner insights into the current and future state of AI governance through semi-structured interviews. Key findings reveal several common themes: AI's scope is too broad, the importance of transparency, the role of ordinary people in its governance, and the need for improved AI literacy. The study revealed that the rapid pace of AI's development is likely to challenge legislators, and that transparency will be vital but difficult to obtain due to competition in ensuring responsible and safe AI development. It also revealed that the current role played by the public in its governance is that of a 'Guinea pig' and that education about AI's implications is essential to allow ordinary people to have an informed voice in the AI governance narrative, rather than allowing a small number of technologist and companies dictate AI's future.

The thesis adds to the growing narrative on AI governance by analysing the current regulatory approaches, the roles different stakeholders can play, including the challenges they are likely to face in designing governance models that can balance innovation with sound ethical practices. The findings emphasise the need for improved AI literacy among the public and the development of new governance frameworks that are agile enough to keep pace with AI's rapid development.

Submission of Thesis to Norma Smurfit Library, National College of Ireland

Student name: Oliver Grogan

Student number: 22103597

School: National College of Ireland

Course: Master's in Business Administration

Degree to be awarded: Master's in Business Administration

Title of Thesis: <u>Artificial Intelligence (AI) Governance in the Technology Sector: Can we anticipate the future and govern the present?</u>

An electronic copy of your thesis will be lodged in the Norma Smurfit Library and will be available for consultation. This electronic copy will be accessible in NORMA <u>https://norma.ncirl.ie</u> the National College of Ireland's Institutional Repository. In accordance with normal academic library practice all thesis lodged in the National College of Ireland Institutional Repository (NORMA) are made available on open access.

I agree to an electronic copy of my thesis being available for consultation within the library. I also agree to an electronic copy of my thesis being made publicly available on the National College of Ireland's Institutional Repository NORMA.

Signature of Candidate:

Oliver Grogan

Electronically signed by: Oliver Grogan Reason: I am the author of this document Date: Aug 7, 2024 21:03 GMT+1

For completion by the School:

The aforementioned thesis was received by _____

Date:

This signed form must be appended to all copies of your thesis submitted to your school.



National College of Ireland

Project Submission Sheet

	Oliver Grogan		
Student Name:	22103597		
Student ID:	Masters in business administration		
Programme:	Ye	ar:	2
	Research methods- Dissertation		
Module:			
	David Hurley		
Lecturer:	-		
Submission Due	August 10 th 2024		
Date:	_		
	Artificial Intelligence (AI) governance in the	technolo	gy sector: Can
Project Title:	we anticipate the future and govern the pres	sent?	
	19791		
Word Count:			

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Oliver Grøgan

Electronically signed by: Oliver Grogan Reason: I am the author of this document Date: Aug 7, 2024 21:02 GMT+1

Signature:

07-Aug-2024

Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS:

- Please attach a completed copy of this sheet to each project (including multiple copies).
- Projects should be submitted to your Programme Coordinator.
- You must ensure that you retain a HARD COPY of ALL projects, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
- You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. Late submissions will incur penalties.
- All projects must be submitted and passed in order to successfully complete the year. Any project/assignment not submitted will be marked as a fail.

Office Use Only		
Signature:		
Date:		
Penalty Applied (if applicable):		

AI Acknowledgement Supplement

[MBA Research methods: Dissertation]

[Artificial Intelligence (AI) governance in the technology sector: Can we anticipate the future and govern the present?]

Your Name/Student Number	Course	Date
Oliver Grogan/22103597	MBA	07-Aug-2024

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click <u>here</u>.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool
Chat GPT	My original thesis draft was over the allowed 20,000 words. This tool was used to condense a small number of paragraphs to bring the thesis word count under the allowed 20,000 words	https://openai.com/chatgpt

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. One table should be used for each tool used.

[Insert Tool Name]
[to reduce word count]
Please suggest a way of shortening this[Insert Sample response] piece without affecting its academic integrity

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Acknowledgements

I would like to thank several people for their guidance and assistance writing this dissertation. Firstly, I would like to thank all the faculty at NCI, particularly the research methods team for their guidance in formulating the study's design. I would like to give a special thanks to my thesis supervisor, David Hurley for his encouragement, support, and advise during the process, nothing was ever an issue. Thanks also to Keith Brittle in the library support centre for his assistance in formatting my dissertation.

I would also like to thank all the interviewee's who gave up their valuable time to participate in this study. Their insights added a wealth of knowledge and dept to my learning.

Lastly, I would like to thank my beautiful wife Sinead for all her support over the last two years, particularly over the last three stressful months. Thank you for putting up with me.

Abstract	1
Acknowledgements	5
Abbreviations	8
Chapter 1: Introduction to the Thesis and Research Topic	9
1.1 Background to the Study	9
1.2 Gaps in the Literature	10
1.3 Academic Justification	11
1.4 Research Aim(s)	11
1.5 Research Questions	12
1.6 Methods and Scope	13
1.7 Dissertation Structure	13
Chapter 2 - Literature Review	14
2.1 Introduction to Artificial Intelligence and a Brief History	14
2.2 AI's Recent Evolution	14
Figure 1.0 – Global private investment in AI between 2013 and 2021	15
2.3 Types and Classification	16
2.4 The Reach of AI	16
Figure 2.0 – McKinsey 2023 global survey on AI use by sector	17
2.5 Societal Concerns	17
2.6 Problems defining AI and Governance	18
2.7 Can we govern the Present	19
2.7.1 Different Methods of Governance	19
2.7.2 Incrementalism	20
2.7.3 The US approach to AI Governance	21
2.7.4 The Chinese Approach to AI Governance	22
2.7.5 The European Union's Approach to AI Governance	23
Figure 3.0 Risk Based Approach to Regulation	23
2.8 Can we predict the future?	24
2.8.1 The Collingridge Dilemma	24
2.8.2 Historical examples of Governance Challenges	25
2.8.3 Technological Determinism and Society	26
2.9 Conclusion	27
Chapter 3 - Methodology	28
3.1 Introduction	
3.2 Research Aims and Objectives	28
3.3 Proposed Research Methodology	30
Figure 4.0. The Research Onion	31
3.4 Research Philosophy	32
Figure 5.0 The Regulation Paradigm	33
3.5 Research Approach	33
3.6 Research Strategy	34
3.7 Qualitative Data Primary Collection	35

Contents

3.8 Population	36
Table 1.0 Participant Information	37
3.9 Analysing Qualitative Data	
3.10 Ethical Issues	38
3.11 Limitations to Research	39
Chapter 4 - Research Findings and Discussion	40
4.1 Introduction	40
4.2 Objective 1: Defining the problem	40
4.3 Objective 2: Govern the present. Are the current regulatory approaches adequate?	44
4.4 Objective 3: Anticipate the future	48
Chapter 5 - Conclusion and Recommendations	
Bibliography	54

Abbreviations

Artificial Intelligence (AI) Artificial General Intelligence (AGI) Artificial Narrow Intelligence (ANI) Artificial Super Intelligence (ASI) Digital Services Act (DSA) General Data Protection Regulation (GDPR) Large Language Model (LLM) Machine Learning (ML) Organisation for Economic Cooperation and Development (OECD) Public Relations (PR) Responsible Research and Innovation (RRI) Small and Medium sized Enterprise (SME) The European Union (EU) United Kingdom (UK) United States (US)

Chapter 1: Introduction to the Thesis and Research Topic

1.1 Background to the Study

The current trajectory of AI has the potential to change many aspects of our lives and society. Dwivedi et al. (2021a) suggest that these advancements have the capacity to augment our daily existence, much like an industrial revolution. The world has experienced four industrial revolutions, starting in the middle of the eighteen hundreds, characterised by a shift from agrarian economics and hand production to machine-based manufacturing and industry. Each revolution has enacted change, radically altering the social fabric at the time, leading to urbanisation, altered work practices, increased affluence, and economic disparities. With each revolution the quality of life has improved, people live longer, work less, and are generally wealthier than their ancestors, but each revolution has also introduced its own challenges such as air pollution and global warming that require complicated solutions (Groumpos, 2021).

AI as a technology presents a paradoxical scenario where its future opportunities are accompanied by significant trade-offs (Dwivedi et al. 2021b). While AI systems have the capability to optimise power grids and reduce energy consumption, Payton (2023) describes how the large data centres needed to drive AI systems globally, account for the same energy use as the annual consumption of the United Kingdom. Conversely, AI has the potential to harmonise the global workforce, enhancing job satisfaction and equality in both developed and developing countries (Rathi, 2018). It offers the prospect of automating mundane tasks allowing humans to focus on management and creative roles. However, Bughin et al. (2018) warn that AI might widen the gap between developed and developing countries because AI can automate jobs that were traditionally outsourced to developing nations, thus creating economic challenges for these regions. In healthcare, AI promises significant advancements in diagnostics, treatments, and patient outcomes. Tizhoosh and Pantanowitz (2018) acknowledge AI's benefits in healthcare but also highlight the possibility of increased costs associated with its development, which could exacerbate healthcare accessibility for populations in developing countries and lower income groups in developed nations. This could cause a healthcare disparity where only the affluent have access to advanced medical care (Sun and Madaglia., 2019).

The current AI revolution stands apart from previous revolutions due its pace of change. Early revolutions happened over decades and were generally confined to a specific area such as

manufacturing which enabled governance strategies to keep pace with their evolution (Groumpos, 2021). AI presents a unique set of challenges because of its complexity, diverse sectors that it impacts, and ethical and privacy concerns arising from the data required to operate its systems. Also, in an era of globalisation, it requires international cooperation which poses a challenge (Taeihagh, 2021).

The technology industry was selected as the focus of this study as it is the largest user of AI systems and the sector that makes theoretical AI systems a reality. The technical knowledge and practical experiences of employees working in the sector are essential to develop effective governance frameworks (Benbya, Davenport and Pachidi, 2020).

1.2 Gaps in the Literature

AI Governance is an emerging field with grey areas that lack a concise definition and scope for what constitutes an AI system and how it should be governed. Until recently, AI systems came under the umbrella of technology governance, but its recent advances have seen its application across a variety of private and public sectors areas including some high-risk sector's such as finance, automobiles and healthcare (Birkstedt et al., 2023). As a result, concern has grown over the risks of using these systems, including algorithmic bias, transparency, and privacy violations (Birkstedt et al., 2023: Jobin et al, 2019).

AI varies from other technologies in several ways that naturally lead to research gaps. Its rapid evolution means that after a research study is conducted on its latest trend, it is often outdated. The technology is so complex that few academics truly understand its workings and finally, it is expected to one day surpass human intelligence, which creates unique challenges such as responsibility gaps and the potential loss of human agency, both of which are in the early stages of research (Santoni de Sio and Mecacci, 2021).

While there is a wealth of research on the technical aspects of AI and how different systems work, limited research exists on the technology's governance, and its effect on society, possibly due the topics complexity (Papagiannidis et al., 2023). Similarly, there is little research on what role ordinary people can play in its regulation. This study attempts to fill the research gaps regarding what constitutes an AI system, evaluate the current governance strategies in the EU,

China, and the US, and explore what technology industry professionals perceive the future to be, all from the perspective of an ordinary person.

1.3 Academic Justification

The justification for the study lies under two broad themes: A lack of 'AI literacy' among ordinary (non-technology industry) people and the implications of the Collingridge Dilemma in governing new technologies.

Firstly, AI's scope means that it will affect multiple aspects of human life and not just employees working in the technology industry (NG et al., 2021). Technology professionals were chosen as the sample for this study as they are best placed to describe the challenges in governing AI due to their technical knowledge. The goal of this study is to relay these challenges in a way that non-technical person can understand, which may enhance their ability to contribute to the complex narrative.

Secondly, the Collingridge Dilemma (1980) describes the challenges of governing an emerging technology in its infancy and explains that when the pitfalls of a technology become known it can be difficult to enact change. The phenomenon has been experienced over the last thirty years during the development of the internet and later with the explosion of social media. Both technologies have created benefits to society, but they have also posed several challenges in their governance, with leading companies from both industries receiving large fines in recent years for breaches of privacy and data regulations (Ulnicane et al., 2021). The distinction between governing these two industries and AI is the risk associated with errors or breaches of regulation, due to the sectors AI is now embedded in, the scale of the population that it affects and the possible existential threat that future variations of the technology may pose. Also, it is important to harness the obvious benefits of the technology that could improve quality of life and enhance society (Dwivedi et al., 2021a).

1.4 Research Aim(s)

This study has three main aims. The first aim seeks to define and clarify the scale of problem facing regulatory bodies by exploring whether the concept of AI is too broad as well as identifying the challenges posed by its complexity. It will explore the role that 'ordinary people' or those not familiar with the mechanics of AI can play in designing governance strategies.

The second aim is to examine the current regulatory approaches to assess if they are sufficient to govern AI's recent advancements, examining whether industry specific approaches may be more successful, as well as probing how global legislation can balance innovation with sound ethical standards.

Lastly, the study identifies the future challenges in AI governance, looking at strategies that may enable the legislator keep pace with AI's advancement as well as exploring the theory that AI is destined to follow its own evolutionary path due to the sheer complexity of the technology and the slow pace of current governance structures. The study asks whether AI development should proceed at a slower pace to allow legislation to keep up with its advances and whether this is likely to happen and why.

1.5 Research Questions

To achieve the study's aim and add guidance to the study, as well as narrowing the focus to specific areas of interest, the following research questions were used for direction.

- 1) Objective 1: Define the issue.
 - Is the current definition of AI too broad?
 - What are the challenges posed by the scale and complexity of AI in its effective governance?
 - What role can ordinary people play and how can their contribution be improved?
- 2) Objective 2: Evaluate the current global governance strategies.
 - Are existing models such as the EU AI Act, effective?
 - What stakeholders should be involved in designing governance frameworks?
 - Can governance bodies collaborate effectively?
 - Should governance be specific to the sector AI is being deployed?
- 3) Objective 3: Explore the future challenges for effective AI governance.
 - Can past mistakes with similar technologies like social media and the internet be avoided?
 - Is AI destined to follow its own developmental path?
 - Should AI development be slowed to allow legislation keep pace with its advancements?

1.6 Methods and Scope

The study was conducted using a qualitative strategy under an interpretivist philosophy, with inductive reasoning. The qualitative approach was deemed necessary due to a lack of established theories and models about the subject. Also, this methodology is orientated towards establishing a deeper understanding of the research problem using semi-structured interviews that were the primary data collection tool. Seven participants from different demographics and in various hierarchy of the technology industry were chosen for the sample to offer insights from a diverse perspective. The research was limited by the size of the sample available to the study and focused only on the governance strategies of the US, EU, and China only.

1.7 Dissertation Structure

Chapter 1 – Background and Justification for the Study

This chapter introduced the topic of AI governance, explaining why its regulation is an important factor in its development. The justification for the study based on research gaps is also explained as well as an outline of the objectives of the study and the methods and the rationale behind them that were used to achieve its goal.

Chapter 2 – Literature Review

This chapter consists of a comprehensive review and analysis of the current literature available on AI governance and an outline of the different approaches currently used by the EU, China, and the US. It also describes the challenges faced by legislators in defining the problem and the obstacles they are likely to face in the future.

Chapter 3 – Methodology

The methodology chapter outlines the objectives of the study and the strategies and approaches that were adopted to achieve the study's aims. It describes how the primary data was collected and analysed, the population that was selected and why, and the philosophy that was used to underpin the study.

Chapter 4 – Discussion and Findings

In this chapter the findings of the study are presented and critically analysed. It highlights the significance of the results and discusses their implications as well as their limitations.

Chapter 5 – Conclusion and Recommendations

The final chapter offers a summary of the study outlining the research gaps that were discovered based on the primary data and the literature review. It also describes the research goals that were reached and makes recommendations for future research.

Chapter 2 - Literature Review

2.1 Introduction to Artificial Intelligence and a Brief History.

The term Artificial Intelligence (AI) was first used at a gathering of academics at Dartmouth University in 1956 where scientist aimed to design a machine that could mimic human intelligence, capable of learning and making decisions (Hopgood, 2005). The twenty years following the Dartmouth Conference, saw considerable progress in the development of theoretical AI concepts but funding was significantly cut in the 1970's due to a lack of tangible applications for the technology. This era was called the 'AI Winter' which lasted until 2014, when investment in AI began to increase as the more advanced computational power made it possible to utilise some of the concepts developed in the 1970's (Wamba et al., 2021). In the current era, AI encompasses a broad range of systems, including simple rule-based algorithms to complex self-learning machines rendering it challenging to generate a single definition. AI also involves several diverse fields of study including computer science, mathematics, psychology, and philosophy, which often sees it described by experts in terms that are most relevant to the field that they work in (Gbadegeshin, et al., 2021). For this review, AI is defined as the use of machines to replicate human intelligence by training computers using algorithms and data processing to exhibit traits such as learning, judgement and decision making, allowing a computer to perform tasks that could only be performed by humans previously (Zhang and Lu, 2021). This definition differs from the aim of the scientists at the Dartmouth Conference, in that AI relies on algorithms and data rather than mimicking the biological workings of the human brain and its cognitive abilities.

2.2 AI's Recent Evolution

In its early stages, AI developers using algorithms such as neural networks and decision trees were restricted by the need for structured and organised data to work effectively, limiting its capabilities to supervised learning. In essence, AI required clear instruction from humans to execute a task (Hilb, 2020). These constraints combined with a lack of computational power and available data when compared to the current era curtailed its uses and prevalence in everyday life (Kar, 2016). AI algorithms have evolved since this time with deep learning and reinforcement learning algorithms which are able to overcome these limitations, now able to process data in its natural form even if unstructured. This progress has been assisted by the advances in computational power and an abundance of available data (Dwivedi et al., 2021b). With this improvement has come the emergence of machine learning and deep learning techniques that allow systems to learn from vast amounts of data, enabling them to uncover patterns in the data to make accurate predictions. This has caused a rapid growth in investment in the sector over the last decade (Figure 1.0). Societies understanding of AI has also changed with applications such as Open AI's chat GPT generative AI changing people's perception of the technology by showcasing everyday uses rather than abstract concepts that large parts of society do not understand (Baldassarre et al., 2023). However, there is still a gap in understanding AI's broader implications and for many it remains a conversation piece, unaware of its various applications.



Source: https://iapp.org/

Figure 1.0 – Global private investment in AI between 2013 and 2021

2.3 Types and Classification

AI can be divided into three distinct types: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Super Intelligence (ASI). ANI, also known as 'weak AI', operates with machine learning and is commonly used in everyday life through applications such as Amazons Alexa and Open AI's Chat GPT. ANI can only perform a specific task and do not possess the ability to transfer knowledge between different tasks or systems. However, ANI systems are expected to automate and replace many roles within society due to their ability to complete tasks at a higher speed and greater accuracy than their human counterparts (Taeihagh, 2021). ANI is also expected to form the foundations of AGI which is still theoretical but may be capable of achieving goals independently within a wide range of scenarios beyond human capabilities. The theoretical underpinning of AGI is that it will be able to adapt to performing a task without the intervention of a human, whereas an ANI system needs reprogramming by a human to adopt to a new task if the context or required behaviour is changed, even if only a minor or subtle change was needed (Goertzel, 2014).

Finally, ASI which is also still hypothetical is predicted to one day be the most accurate form of AI and for some scholars an existential threat because it may be able to function beyond human abilities capable of advancements in fields such as science, medicine, and academia that are beyond human intelligence, potentially leading to the redundancy of large parts of society (Kelly, Kaye and Oviedo-Trespalacios, 2023). Kurzweil (2024) describes this event as 'singularity' where machines will be capable of improving their own capabilities and intelligence exponentially. This 'self-improvement' will allow ASI to rapidly surpass human intelligence, potentially solving some of the complex global issues such as disease and climate change. However, the author also notes the potential difficulties in controlling such system to ensure that they align with societal values such as the right to privacy or preventing their malicious use.

2.4 The Reach of AI

The McKinsey Global Survey (2023) found that one third of the organisations surveyed are using generative AI in at least one business function and over half of the companies surveyed are using AI for business analytics or automation of some kind. The adoption of the technology has spiked over the last five years, and AI is playing a pivotal role in diverse industries such as finance, energy, healthcare, retail, manufacturing, education, and technology (Figure 2.0).



Figure 2.0 – McKinsey 2023 global survey on AI use by sector

Source: mckinsey.com

AI systems are supporting medical professionals in diagnosing diseases and identifying early signs of illness, and helping energy companies to predict consumption across grids, reducing the amount of wasted, unused energy. They are also improving communication around the globe through real time speech translation (Maslei et al., 2023). Contrary to some dystopian portrayals of AI, it has the potential to make driving a car safer, improve education and enhance quality of life. Universities, governments, and technology companies are currently investing heavily in AI research to improve AI systems interaction with humans so that they will become more adaptive to human personalities, capable of monitoring their well-being and alerting them to risk (Stone et al., 2022). Governments are investing in AI to improve their performance in areas such as the environment by identifying high contributors to air pollution to inform policy interventions or to improve defence by predicting vulnerable areas within their cybersecurity systems to prevent attacks (Leslie et al., 2023).

2.5 Societal Concerns

While the technology has the potential to provide significant benefits to society, there still exists concerns about the possible negative's aspects of AI. This fear is often fuelled by a lack of understanding of how AI operates, and prominent figures expressing concerns about societies ability to control the technology, or the potential job loses that may be caused by its

development (Beauchamp, 2018). Nassar and Kamal (2021) note the significant ethical concerns that are ingrained in the technology including algorithmic bias and a lack of transparency and accountability making it difficult to hold corporations responsible for an AI's actions. A 'Responsibility gap' may occur because as AI systems evolve and learn from interactions with other agents, it will make human control of their actions difficult causing a dilemma for society as to whether to continue developing the technology and risk losing control or maintaining control and forgoing the potential uses of these systems (Santoni de Sio and Meccaci, 2021; Matthias, 2004).

Further challenges include the collection and use of data by AI systems which affects people's right to privacy, and informed consent. AI can also be misused to generate 'fake news' or to impersonate another person using 'machine speech' to manipulate public opinion (Wang and Siau, 2018). Corrêa et al (2023) see a difficulty in creating a globally accepted set of ethical principles for AI, because of numerous national interests and cultural perspectives leading to different ethical standards.

These concerns are further magnified by the companies that are driving AI development such as Microsoft and Alphabet with annual revenues greater than many countries economy. With a small cohort of powerful companies holding a considerable influence on the technology's direction, the potential for abuse of power is significant. Several technology giants have faced government fines for privacy breaches, antitrust violations, and misuse of their market power in recent years (Tirole, 2021). However, these fines often fail to act as a deterrent for improper behaviour due the financial resources of the corporations. (Vallance, 2024).

2.6 Problems defining AI and Governance

In the context of a technology, Kuhlmann, Stegmaier and Konrad (2019) describe governance as a set of frameworks, policies and processes that direct the administration and regulation of a sector to ensure that it functions in an organised an ethical way. In the context of AI, this involves setting standards, rules, and procedures to manage the development and deployment of the technology to reduce negative impacts. According to de Almeida, dos Santos and Farias (2021) effective governance needs to balance innovation with ethical and compliance considerations to ensure that AI benefits society while minimising risks which involves applying various methods and approaches each suited to different contexts and objectives rather than a single mechanism. This becomes more prevalent as Burt (2021) describes how law makers have not agreed on a consensus of what constitutes an AI system rendering it difficult to develop a common standard to govern the technology. Without a clear definition of what is considered an AI system, it creates ambiguity as to what legislation is governing a particular application.

The high degree of uncertainty and complexity around AI presents challenges for governments in designing and implementing effective policies. AI's non-linear nature makes it difficult to establish objectives and even with high levels of transparency, it can be difficult to explain how some outputs are reached from algorithmic inputs (Taeihagh, 2021). The complexity in some AI system results in only their designers understanding the potential outputs. Therefore, potentially the only people with the knowledge to suggest correct governance strategies are the same people designing the systems reinforcing the need for transparency. It is also difficult to design governance strategies for systems and uses that have not yet been created (Roberts et al., 2023).

In crafting a governance strategy, legislators have several methods available to them which will be discussed in the next part of this chapter.

2.7 Can we govern the Present

2.7.1 Different Methods of Governance

There are three recognised governance strategies available to legislators to design AI regulatory frameworks, with each having their strengths and weaknesses. This section will outline these methods and how they could be applied to AI.

2.7.1.1 Democratic Deliberation

Democratic deliberation is a governance approach that promotes participation from diverse stakeholders in the decision-making process. It prioritises transparency, inclusivity and open dialogue which ensures that diverse perspectives are considered which may help prevent biases in AI systems enhancing fairness and equity (Gherghina, Mokre and Miscoiu, 2021). This type of approach could also foster a framework that reflects the values of a diverse spectrum of society rather than those of powerful corporations. However, a negative of this type of approach is that it may not be agile enough to deal with the rapidly changing nature of AI. Also, the

complexity of many AI systems means most of society lack the technical knowledge to make an informed contribution. (Buhmann and Fieseler, 2023).

2.7.1.2 Multistakeholderism.

Multistakeholderism is a governance model that aligns various stakeholders including governments, businesses, civil society, academia, and technical experts to collaborate on policy making and implementation. The concept grew out of traditional frameworks inability to deal with globalisation and its challenges which included climate change, the internet and more recently vaccine research. The approach recognises that no single entity has all the answers which is apt when dealing with the complexities and rapid evolution of a technology like AI (Raymond and DeNardis, 2015). According to Gleckman (2018), proponents of the theory, argue that the model is well suited to solving complex governance gaps, but often risks can be overlooked or inadequately addressed due a belief that major international stakeholders will deal with any issues. Opponents of the method argue that despite its inclusivity, there is a risk of power imbalances where dominant stakeholders such as larger nations or corporations heavily influence discussions and the decision-making process, marginalising less influential stakeholders (van Klyton et al., 2023). Also, due to its nature, with diverse stakeholders, it can take time to reach decisions, and establishing clear responsibility for decisions and actions becomes difficult with so many parties involved. Therefore, while multistakeholderism has many benefits, its lack of agility may prevent its use in governing AI effectively as a single method.

2.7.2 Incrementalism

Incrementalism is a pragmatic approach to governance that involves making small, manageable adjustments to policies and regulations over time based on feedback and ongoing assessment. A core belief of the theory is that rational decision making is not possible for most policy issues due to disagreements over objectives from the individual stakeholders involved in the process. The theory also recognises that not all the stakeholders have access to the necessary information needed to offer valid inputs, and the natural limitations of human foresight (Hayes, 2022). The approach allows for flexibility and adaptability which are essential to a technology like AI which is continuously evolving. However, while the method provides a structured way to navigate uncertainty, and regarded as conductive to innovation, it may also result in slow progress and thus may not be agile enough to address fast-paced changes that are predicted to occur in the future (Gikay, 2024).

According to Jones et al. (2024), governing bodies will need to employ components of all three of these models to design a new 'hybrid' approach to govern AI effectively. The open dialogue associated with democratic deliberation is key to promoting transparency. The inclusivity of multistakeholderism allows governing bodies, technical experts, and corporations to come together to discuss policy, while the theory of incrementalism allows for small manageable changes to be made to regulations over time. The challenge facing governing bodies is how to design regulatory frameworks that can keep pace with AI's development, identify its risks, and mitigate against its potential negative impacts in a timely manner (Jones, 2023).

The next part of this chapter will analyse the strategies currently being used by the three territories driving AI growth.

2.7.3 The US approach to AI Governance

The United States has adopted a sectoral approach to AI governance, where regulations are developed for specific industries such as healthcare, finance, aviation, and defence. The complex nature of the US legislature that involves the creation of state and federal laws means that the US currently lacks a comprehensive federal law specific to AI governance. However, there have been several executive orders and regulations passed in recent times. One such law saw the establishment of an AI Safety Institution which is supported by a consortium of stakeholders that helped create the National AI research and development strategic plan which has a key role in directing federal investment in AI research and development. The US is also investing heavily in evaluating how existing laws apply to AI (Fazlioglu, 2023).

This investment may be due to the US policy of allowing the private sector to self-regulate to accelerate advancements and deployment of the technology. This approach has fostered innovation and sees several US companies at the forefront of AI development, but it also poses the possibility of risky behaviour from these companies, potentially compromising safety, and ethics. A difficulty in the US approach will be attaining the correct balance between oversight and innovation to align the private sectors interests with the public good (Dafoe, 2018). According to Daly et al. (2019) the US policy of prioritising its competitive position could lead to a lack of comprehensive governance resulting in gaps in consumer protection, ethical standards, and privacy. While the US has employed the principles of multistakeholderism with various contributors to the narrative, its policy is seen to lack inclusivity and diversity as policy tends to be driven by powerful corporations (Shams, Zowghi and Bano, 2023). The lack of a

comprehensive federal law and the current system adopted by the US could see complex and inconsistent regulation across different states and sectors that may be difficult to change in the future. When combined with the complexity of the US legislative system involving state and federal laws, it may be problematic to develop the agile governance methods necessary for effective governance of the technology. Huttenlocher, Ozdalgar and Goldston (2023) highlight three key areas that are essential to the success of the current US strategy. Firstly, they argue that to regulate a technology, its scope and definition must be agreed. Lawmakers in the US then need to update existing laws to ensure they adequately cover AI technologies once defined, to prevent misuse. Lastly, they recommend developing strict industry auditing regimes to ensure the safe and ethical use of AI.

2.7.4 The Chinese Approach to AI Governance

Originally, the Chinese approach to AI governance differed to the US using a top-down method where the Chinese government approved all AI algorithms before their deployment to confirm that they adhered to core socialist values (Whyman, 2023). The method was based on guidelines set out by the Chinese government in 2017 that formed part of a goal of becoming an AI superpower by 2030 with a targeted output from the Chinese AI industry of 1 trillion yuan (State Council of the Peoples Republic of China, 2017). China updated this strategy in 2021 to a more fluid and interactive process driven by a mix of stakeholders from diverse backgrounds including regional governments, academia, journalism, technologists, and employees at large technology companies. The evolution of their policy has accelerated China to become a leader in AI governance that has enabled the state roll out targeted and binding regulations where other territories have been slow to act including regulations to control recommendation algorithms, generative AI, and facial recognition (Sheehan, 2024). China is also working on legislation to address AI as whole rather than its subsections. This legislation is expected to take some time to complete and in advance of its publication, the Chinese Academy of Social Sciences drafted an advisory version of the future law highlighting areas of AI that Chinese companies should avoid unless they have the approval of the government (Yang, 2024). While China has been able to overcome one of the drawbacks of a Multistakeholder approach to governance in enacting swift legislation, Yang (2024) highlights how Chinese courts are set to take a business-friendly approach to disputes over ownership of data and intellectual property which could mean innovation and economic growth are prioritised over ethics and privacy. Also, China's focus on controlling public content which is regarded negatively due the possible human rights violations may damage China's global

reputation and create a reluctance for collaboration with them by other territories, which is likely to be needed to govern AI effectively (Gong, Qu and Dorwart, 2024).

2.7.5 The European Union's Approach to AI Governance

The European Union (EU) is the first world superpower to draft a comprehensive AI governance framework. The EU AI Act was passed by the European Parliament in June 2024 and will become law by 2026 (European Parliament, 2024). The strategy was developed using inputs from multiple stakeholders such as academics, industry experts, and public consultation. The policy has been crafted to allow incremental changes to the legislation to ensure continuous improvements that may be required in response to societal changes or technology advances. The policies of the act will be monitored by a new body called the European AI office using a risk-based approach to governance. The Act categorises AI systems into four risk categories as set out in Figure 3.0 with technologies considered unacceptable risk prohibited and those with minimal or no risk subject to no restrictions. The policy does not yet clearly define all AI technologies, but its broad description of AI is set out to be as neutral as possible to cover technologies which are not yet known or developed (Sioli, 2021).



Figure 3.0 Risk Based Approach to Regulation

Source: <u>https://www.ceps.eu/</u>

The Act is designed to promote innovation by allowing developers to experiment with AI systems in a controlled environment whilst safeguarding an individual's rights by prohibiting high risk applications that could manipulate behaviours or exploit the vulnerable. As part of its scope, the EU has placed the regulatory obligations on the developer and has the power to impose fines up to 7% of a company's annual turnover for breaches of the legislation (Carter, Hyman and Silver, 2024). The regulation imposes legally binding rules to promote

transparency that requires AI developers to inform users when they are interacting with an AI system such as a chatbot. The same companies are also obligated to identify AI generated content or deepfakes in such a way that they can be detected. The bill also requires companies using AI to mine data for business decisions in industries such as insurance and banking to assess how using AI will affect people's rights (Heikkila, 2023).

However, critics of the act believe that it is overly restrictive and complex and may stifle innovation by small and medium sized enterprises (SME) due to the possible fines for breaches of rules, putting European companies at a disadvantage to their international counterparts. The same fines may not act as a proper deterrent for larger technology companies due to their size. The Act has been criticised for the ambiguous definition and risk level of AI, resulting in simple software systems falling under the regulations that do not pose risk, further slowing innovation (Higgins, Jackson and Korten, 2023). Novelli et al. (2024) also argue that the new legislation fails to address the liability concerns about responsibility for the autonomous actions of systems such as large language models (LLM's) should they be seen to cause damage or breach the rules of the Act.

2.8 Can we predict the future?

The current challenges facing regulatory bodies which include defining the scope of AI, and refining current strategies to keep pace with the technology's innovations are likely to continue to be an issue in the future. This rationale assumes that we have a limited understanding of the potential challenges that a new technology poses (Veale, Matus and Gorwa, 2023). The theories behind this thought process will be explained in the next part of this chapter.

2.8.1 The Collingridge Dilemma

The Collingridge Dilemma (Collingridge, 1980) explores the relationship between society and technological development and the challenges associated with governing new innovations. The fundamental of the theory is that while it may be easier to enact governance guidelines for a technology such as AI when in its infancy, there is limited information about the technology's pitfalls, and once the impact of the technology and its regulatory gaps become known, the technologies integration into society can lead to difficulties controlling or modifying it. Collingridge (1980) saw control as an essential part of responsible research and innovation (RRI). This notion is important considering the size and influence of the corporations driving

AI's growth. Part of the dilemma is how to manage the conflicting interests and power imbalances that the technology may cause. He saw democracy as the best way to govern innovation which may be difficult due to the competing interests and the need for global cooperation (Genus and Stirling, 2018).

This obstacle may be compounded by the 'pacing problem' which challenges existing models of governance to keep up with rapid advancements in a technology's development due to their cumbersome and bureaucratic nature. As a result, regulations often fail to enact sufficient control by the time technology has an impact on society and in many instances, they are outdated when they become law (Wallach and Merchant, 2019). For this reason, traditional government regulation is no longer seen as adequate to oversee AI's development. Wheeler (2023) argues that while lawmakers still have a role to play, AI governance needs to be expanded to include new institutions that use more agile and reflexive methods to govern the technology effectively. Allowing companies to self-regulate AI may repeat the previous faults from self-regulation on social media and the internet.

2.8.2 Historical examples of Governance Challenges

The Collingridge dilemma may be easily illustrated by the development of a technology that most people have lived through. Thirty years ago, it would have been difficult to predict the ways the internet would integrate into our everyday lives. Originally a US military project, the internet developed into a global tool for sharing and accessing information easily amongst its users. The internet supports a large quantity of legitimate businesses, but overtime, the internet has also been exploited for malicious purposes such as crime, unauthorised surveillance, and disinformation. Privacy has also been a concern. Search engines like Google do not charge users for their products so in this instance many users information becomes googles product as they share this information with other companies to enhance their business objectives (Di Pietro and Cresci, 2021). For lengthy periods, internet companies were able to use people's data without their knowledge for their own business concerns. Upon discovering these gaps, governments took actions and legislation was subsequently introduced and updated to drive governance including Europe's General Data Protection Regulation (GDPR) in 2018 and Digital Services Act (DSA) in 2022.

The growth of social media over the last twenty years also exemplifies the dilemma as it has evolved from a platform created for personal communication, to share stories and images, into a powerful tool with political, social, and economic impact. Social media platforms have become entrenched in society where changes in their operations and regulation faces resistance from both the companies and its users despite regular breaches of people's privacy, misuse of their data and the spread of misinformation or 'fake news' (Ortiz-Ospina, 2019). Cammaerts and Mansell (2018) describe data as being more valuable than oil and of significant economic importance to social media companies in a comparable way to internet companies. Social media companies have lacked transparency in their operation and faced penalties for breaches of data laws and governments have struggled to govern them effectively. Just as social media companies have struggled with governance and regulation, AI development poses major challenges due to its reliance on data and the pace at which it is evolving, which highlights the need for agile governance frameworks to counteract the theory of the Collingridge dilemma.

2.8.3 Technological Determinism and Society

Technological determinism is the concept that technology shapes the structure and culture of our society. In essence, technology may determine the course of history at any given time and social progress follows an inevitable course that is driven by technological innovation (Hallstrom, 2022). Critics argue that the concept oversimplifies the causes of societal change because the theory ignores significant factors such as economics, and politics. The development of the internet and social media illustrate the concept as these technologies have changed the way we communicate, access information, and conduct business, although opponents of the theory would argue that other factors were involved in these changes.

While these technologies have created change, the main difference between them, and AI is the concept of agency. Choosing not to use social media is unlikely to have a direct impact on one's life, whereas AI's revolutionary potential could cost a person their job. As individuals, people rely on proper governance to protect their best interest and maintain their individual agency. An example of where governance has failed society is climate change. Many people choose responsible actions to reduce their environmental impact while some large industries continue to burn fossil fuels for energy and transport enhancing climate change (Salsone et al., 2022). Human agency is also a factor in the technology's development and the economic interests of individuals and corporations can often supersede those of society in general.

According to Roser (2022), the wider public needs to become informed and engaged with the development of AI, or a small number of technologists will decide how AI changes our world

affecting everyone, even those who choose not to use the technology. This is increasingly important as AI systems are developing their own agency, capable of acting autonomously when interacting with humans. Sundar (2020) describes how this possibility poses significant challenges, particularly concerning privacy and human autonomy. AI systems abilities to collect vast amount of personal data raises privacy concerns and its increasing ability to make autonomous decisions enhances a fear of losing human control leading to a diminished sense of agency. An ironic outcome to this dilemma is that organisations are using AI machines to reclaim human agency. AI's ability to prevent data breaches enhances human agency by safeguarding personal information. Likewise, AI's ability to analyse data at an elevated level enables individuals to make more informed decisions (Fanni et al., 2023).

The key to governing AI in the future will be finding the balance between leveraging the benefits of AI and protecting individual's rights and freedom. This involves implementing robust data protections measures, transparency in the development of the technology and maintaining oversight to prevent the loss of agency and control (de Almeida et al., 2021). The complexity of the task outlined in these theories underscores the need for effective governance frameworks. The autonomous capabilities of AI and its potential to disrupt employment and privacy requires robust oversight to balance innovation and AI's potential with societies needs and safety.

2.9 Conclusion

AI has the potential to improve the world we live in and innovate new technologies yet unimagined. However, it also poses several risks that need to be managed, including ethical breaches, privacy loss and potential negative disruptions to the global workforce. There are evident gaps in research regarding the effectiveness of the current governance models, debate over what constitutes an AI system, and little known about what the future holds. This is largely due to the evolving nature of the technology. The purpose of this research project is to add to the growing narrative on how to manage AI both now and in the future. The next chapter will outline the methodology used in doing so.

Chapter 3 - Methodology

3.1 Introduction

Saunders, Lewis and Thornhill (2019) describe research methodology as a systematic approach undertaken to discover new findings. They emphasise the importance of selecting appropriate data collection and analysis techniques to ensure clarity and success in achieving a research paper's objective. Effective research is constructed by removing inner beliefs or biases about the information gathered during the process to discover logical relationships that lead to new findings. This involves explaining the methods used to collect and process the data and the approach and philosophy used during the process, all of which should be based on a clear purpose to answer a question or find a solution to a problem (Ghauri and Gronhaug, 2010).

This chapter outlines the stages followed to complete this research project using a systematic approach to data collection and interpretation. It outlines the aims of the research, the reason for using a qualitative strategy, as well as explaining why an inductive approach was chosen to conduct the study. It also describes the method used to collect the data, how the sample was selected, and discusses the study's limitations and ethical concerns.

The study is likely to form part of an ongoing narrative about AI governance for some time to come due to the pace of the technology's development. Every effort has been made to ensure the validity and reliability of the research to support increasing knowledge in the field of AI governance which is still in its infancy.

3.2 Research Aims and Objectives

According to Fossey et al. (2002), the aim of a study or research paper needs to be clearly defined to help readers understand the authors intentions and the choices made in selecting the sample participants and analysis methods used. A clearly defined aim also provides the reader with a better understanding of the author's starting point, allowing them to see how their perspective has evolved through the research process which indicates that the researcher has learned from the sample participants, adding authenticity to the study.

AI is creating a high degree of uncertainty, and due to its complexity, poses considerable challenges for regulators in designing effective governance policies that will keep pace with its development. Even reaching a consensus on what constitutes an AI system is proving problematic (Taeihagh, 2021).

This study had three aims. The first aim sought to define and clarify the problem by exploring whether the concept of AI is too broad as well as identifying the challenges governing bodies face due to its complexity. It also explored the role that 'ordinary people' or those not familiar with the mechanics of AI can play in designing governance strategies as well as assessing the need for greater transparency in achieving successful AI governance.

The second objective of the study was to examine the current regulatory approaches to assess if existing regulations are sufficient to deal with AI's recent advancements, exploring whether industry specific approaches may be more successful, as well as probing how global legislation can balance innovation with sound ethical standards.

Lastly, the study attempted to identify the future challenges in AI governance, looking at strategies that may enable the legislator keep pace with AI's advancement as well as exploring the theory that AI is destined to follow its own evolutionary path due to the sheer complexity of the technology and the slow pace of current governance structures. The study also asked whether AI development should proceed at a slower pace to allow legislation to keep up with its advances and whether this is likely to happen and why.

Saunders et al. (2019) describe how research can begin with a broad question that generates more detailed questions, helping to achieve the researcher's objectives. While questions and objectives may be similar, they differ in that the researcher uses questions to develop a theory or solution to the issue discussed in the study as part of its objectives. The following research questions were asked during semi-structured interviews as part of this study's objective of defining the problem of AI governance, evaluating the effectiveness of current governance strategies, and predicting the future problems facing regulatory bodies.

- 1) Is it possible to clearly define the problem at hand?
 - 1.1 Is the term AI too broad and ambiguous?
 - 1.2 What challenges does the scale and complexity of AI bring to its effective governance?
 - 1.3 Is greater transparency a key parameter to achieving success in AI governance?
 - 1.4 What role can ordinary people play in designing an effective governance strategy, if any?

2) Govern the present.

2.1 Are current regulatory approaches adequate to address the recent level of AI development?

2.2 Would Industry specific policies be more effective in governing AI development and deployment?

2.3 How can the global governing bodies design legislation that fosters innovation with the requirement for regulations to prevent bias and protect privacy

2.4 How effective are the current enforcement mechanisms and will they be successful in ensuring the ethical development of AI?

3) Anticipate the future.

3.1 What are the future challenges for governing bodies? Is it possible to anticipate and prepare for them?

3.2 What strategies can be employed to ensure that legislation keeps pace with the technology's rapid development.

3.3 Is it possible to design governance models that guide the direction of AI development in a way that enhances society or is the technology destined to follow its own course?

3.4 Should development of the technology happen at a slower pace to allow legislation keep up with advances?

3.3 Proposed Research Methodology

Mukherjee (2019) describes how a good research methodology enhances the reliability and validity of research findings, making it an essential part of any study. The methodology provides a structured approach to conducting research, ensuring that the process is systematic, structured and credible by guiding the researcher in the collection, analysis and interpretation of data.

The research methodology chosen for this study was created using Saunders, Lewis and Thornhill's (2007) 'research onion'. The 'onion' framework breaks the methodology into six layers, with each stage having a fundamental purpose in delivering a successful project (Tshitadi, 2024). Originally created to structure social science research, some critics argue that using the 'onion' for guidance oversimplifies the process by presenting the methodology as

linear with a step-by-step progression that neglects certain factors and has limited philosophical scope. Also, as it was originally designed for business studies and social sciences, it is regarded as better suited to quantitative studies that rely on hard facts rather than philosophical opinions (Melnikovas, 2018). However, even though this study does not use quantitative methods such as surveys, it is still considered a social science study as it focuses on technology professionals' attitudes to ethics, bias, and governance. Another advantage of the 'research onion' is the flexibility it offers to assess and combine different approaches and methods that best fit the research. It also offered a clear guide to the stages that must be completed as part of the methodology, providing clarity, and encouraging critical reflection on each stage of the process (Saunders et al., 2007). A similar methodology was successfully used by (Aunger et al., 2022) in their peer paper analysing the regulatory reforms in the UK's National Health Service during the Covid-19 pandemic.

The next part of this chapter outlines the methods and approaches that were adopted for this study, explaining the reasons behind each choice, as well as offering an analysis of the alternatives that were available to the study but were not utilised.



Figure 4.0. The Research Onion

Source: Saunders et al., 2007

3.4 Research Philosophy

The first layer of the 'onion' details the research philosophy adopted for the study. Bleiker et al. (2019) emphasise how maintaining consistency between the philosophical position taken and the methods used during a study play an important role in the quality of the research. The adopted philosophy also gives a perspective on the researcher's beliefs, offering insights into their chosen strategy. Some philosophical choices may be forced due to the nature of the study, while others explain how a researcher views the relationship between knowledge and the research process. Researchers focusing on factual data will have different methods and priorities than those studying people's feelings and attitudes, which affects what they consider to be important (Saunders et al., 2007).

Saunders et al., (2007) describe three ways of thinking about research philosophy: epistemology, ontology, and axiology. For this study, epistemology, and its interpretivist approach was chosen, as it focuses on the subjective nature of human experiences and understanding, providing a perception of the world from the research participants viewpoint. Epistemology, also known as 'the theory of knowledge', is regarded as a sound philosophical guide for researchers conducting qualitative research, as it adds to the understanding of how knowledge is obtained, while recognising some of the limitations of the knowledge acquired such as scope and validity (Myers, 2008).

Saunders, Lewis and Thornhill (2019) describe three main approaches under the epistemology philosophy: Positivism, Realism and Interpretivism.

Positivism focuses on objective and testable data without allowing for speculation or opinions that may enhance the narrative around AI governance (Myers, 2008). Realism is based on the concept that reality exists independently of human perceptions, and our understanding of reality is influenced by social and cultural factors. Research of this kind involves the analysis of historical data and how it has changed over time which may be applicable to the early evolution of AI but is not suitable to studying the current and future governance strategies required for the technology (Saunders et al., 2019).

Ontology which is associated with the nature of reality and its subjectivist approach was also considered as it helps researchers understand the nature, structures, and relationships within the social and political world (Hall, 2003). However, Grass (2024) argues that this approach may lack clarity for a qualitative study on AI governance as it deals with assumptions about

reality and their causes that are not always clear. This makes it difficult to explain and develop theories about AI's social and regulatory impacts. Also, its dependence on structure deemed it unsuitable to this study, which better fits an epistemological approach that offers flexibility and adaptability to garner a deeper understanding of the topic. Tallburg et al. (2023) also note the importance of an epistemological approach to studying AI governance as it allows for the combination of empirical and normative perspectives which increases the flexibility of the study to adapt to the evolving nature of AI governance.

Also, although this study does not sit directly under the heading of organisational analysis, it addresses similar challenges to those covered in the four paradigms for organisational analysis (Figure 5.0 – Burrell and Morgan cited in Saunders et al., 2019), which under the regulation paradigm recommends an interpretive approach to comprehend the challenges associated with AI governance. The functionalist paradigm which also sits under the regulation paradigm was not considered because it is better suited to quantitative research.



Figure 5.0 The Regulation Paradigm

Source: Four paradigms for organisational analysis (Burrell and Morgan cited in Saunders et al., 2019)

3.5 Research Approach

Qualitative research involves the creation of hypotheses rather than the testing of hypotheses as done in quantitative research. It gathers qualitative data such as ideas, statements, and reason to answer questions like what, why and how. To change qualitative data to theory development, inductive reasoning is required (Barroga et al., 2023). Saunders et al. (2019) recommend when

researching a new topic with limited existing literature to work inductively by gathering data initially, then analysing the data to see what theoretical themes emerge from the research findings. An inductive approach was used due to a lack of available data about AI governance to analyse, especially on its future direction. The inductive approach allowed data to be created on the subject, derived entirely from the experiences of the interview participants to create a hypothesis. According to Azungah (2018), although the objectives of the study influence the findings, by utilising inductive analysis they arise from the analysis of the raw data collected and not from pre-existing models or expectations which can allow new insights to be discovered, reducing researcher bias.

The lack of established theories and models about AI governance deemed a deductive approach unsuitable because this type of approach depends on well-defined theoretical frameworks to guide the research process (Stone et al., 2022). Similarly, an abductive approach was rejected because it involves moving from deduction to induction to form a theory, although this approach may suit later studies analysing the effects of current governance strategies (Saunders et al., 2019).

3.6 Research Strategy

An interpretivist philosophy using inductive reasoning forms the basis of this qualitative research strategy where the focus of the study is on understanding the qualities of the phenomena being studied rather than measuring them numerically. This strategy allows a researcher to produce data that cannot be generalised or quantified easily, offering the flexibility to explore unexpected patterns and develop new theories that may not have been anticipated at the study's beginning (Sutrisna, 2009). Emerging technologies like AI are often engulfed by hype and expectations which shape their development and influence their regulation. Using a qualitative strategy helps provide an understanding of the context of this hype by capturing people's experiences within the technology industry to produce more informed regulatory guidance (Ulnicane et al., 2020).

While it may be interesting to conduct a survey on whether technology industry employees view the current regulations as being adequate to control the direction of the technology, it would do little to add to the narrative about what approaches are needed to ensure its ethical development. The inductive approach associated with this qualitative strategy is designed to ensure that the research is orientated towards establishing a deeper understanding of the research problem and its challenges rather than finding a consensus on whether current AI

governance strategies will be effective or whether people statistically agree that we will be able to govern AI effectively in the future (Savai, 2016).

Also, grounded theory which aligns with the exploratory nature of qualitative research and is underpinned by the assumption that progress and social change are best understood through the eyes of the stakeholders involved is a suitable and effective strategy for writing a study on AI as it allows for the development of theory directly from empirical data. Pragmatism which underpins grounded theory, notes that research answers are better achieved inductively with constant empirical verification (Wuest, 2012). Semi-structured interviews which are a core qualitative method are a valuable tool for collecting empirical data, as they allow the researcher to focus interviews on specific topics while concurrently enabling them to explore ideas that arise during the interview which enhances the depth of the data available for analysis (Adeoye-Olatunde and Olenik, 2021).

3.7 Qualitative Data Primary Collection

The primary data collection method adopted for this study was a series of semi-structured interviews due to the subjective nature of the study and its qualitative methods. Easton, McComish and Greenberg (2000) outline several challenges faced by qualitative researchers when conducting interviews that can threaten the quality of the data such as stereotyping the results, rearranging it to suit a bias, or allowing interviews to become too unstructured, losing their focus. They also note the importance of using the correct equipment for the interviews, in the correct environment to allow easy transcription for later analysis.

A pilot study as part of an iterative process was used to pre-test the interview questions for bias and validity and ensured the technology used in the interviews, Microsoft Teams was adequate to allow for easy transcription afterwards. One advantage of using pilot studies is that it can allow a researcher to predict any potential project failures (Van Teijlingen and Hundley, 2002). Semi-structured interviews are the most common tool used in qualitative research as they outline the topics and research objectives of the study (Alshenqeeti, 2014). Although the questions function as a guide, the semi-structured nature of the interview allows the researcher to pose more enhanced questions based on the responses of the subject, further enhancing the validity of the study (Adhabi and Anozie, 2017).

The interview questions were designed as open-ended to allow no definitive answer from the interviewee. The interviews consisted of four questions in three key areas to offer insights into

the challenges faced in defining the problem at hand, evaluating the current AI governance strategies and predicting the future direction of the technology. While the secondary data collected during the study proved useful in gaining an understanding of the research problem, the semi-structured interviews allowed a deeper dive into the issue of AI governance, fostering a better understanding of the challenges it poses (Edwards and Holland, 2013).

3.8 Population

Khan (2020) asserts that the sampling technique used in research plays a pivotal role in the validity of a study's results. In order to find an answer to the research problem, it is important to select a sample of the population that is best suited to enhancing the accuracy of the study within its timeframe. Ideally, a researcher would ask every interested party or the full population their opinion on a research topic. Unfortunately, this is not feasible due to the costs and timeframes associated with doing so. Also, in the case of this study where one of its objectives is to discover what role 'ordinary people' can play in AI governance, it is not warranted. Correct sampling helps to avoid selection bias where certain groups are overrepresented or underrepresented which can taint the results (Babbie, 2020). The complex nature of the subject matter of this study required a specific focus on certain areas that the general population would have little knowledge of. A purposeful sampling method was used to capture the perspectives of stakeholders who understand the current and emerging issues.

The purposeful population was designed using information obtained from an OECD report by Lane, Williams and Broecke (2023) which identified a different perception to AI technology adoption within the labour market based on age, sex, and seniority within an organisation. Seven candidates from different segments of the technology industry, at different levels in their careers and of different ages were selected and interviewed as demonstrated in Table 1.0. The Lane et al. (2023) report provides empirical evidence on varying perceptions of AI based on demographic factors. By using the report and its trends as a guide, it assisted in creating a population that would offer diverse opinions on the subject matter that best informed the research (Creswell and Poth, 2016).

Participant No.	Gender	Age Range	Seniority Level
1	Male	50-60	Programme Manager
2	Male	40-50	Contractor
3	Female	30-40	Middle Management
4	Male	20-30	Graduate / Intern
5	Male	40-50	Company Director
6	Male	30-40	Senior Data Analyst
7	Male	40-50	Team Manager

Table 1.0 Participant Information

3.9 Analysing Qualitative Data

Data analysis is the process that enables a researcher to find meaningful information from the data gathered during the research process. In this study, data analysis began in the early stages of the research project as the initial analysis was used to create a framework for the subsequent work (Saunders et al., 2009). An inductive approach with thematic analysis using a grounded framework was used to identify patterns or themes to develop theory. There were five steps followed in developing theory from the dataset: compiling, disassembling, reassembling, interpreting, and concluding (Castleberry and Nolen, 2018).

Compiling:

The data was compiled through semi-structured interviews conducted on Microsoft Teams. Each interview was recorded and transcribed verbatim using the platform's software. After each interview, the transcripts were formatted and reviewed to create a familiarity with the data.

Disassembling:

Coding of data is an essential step in data analysis. It encompasses organising and describing data to make it more manageable. This involved highlighting key lines of text from the transcripts and assigning labels. The interview transcripts were reread several times, refining

the labels until a set of codes that clearly captured the main ideas in the data were realised (Knott et al., 2022).

Reassembling:

The codes were then used to create themes within the data. A theme highlights important aspects of the data relating to the research question that can be used to generate a theory (Bruan and Clarke, 2006). The themes were then analysed to identify common trends that related to the research objectives.

Interpreting:

Themes were used to make analytical assumptions that added insight into the research question. Even though the analytical steps are listed in a linear format, the grounded framework of the study allowed for an evolution of the interview questions to cover new areas of interest that arose during the study's early interviews, enhancing the validity of the study (Castleberry and Nolen, 2018).

Concluding:

Conclusions offer an answer to the initial research objective and help create recommendations that will add to the narrative in solving the research problem (Castleberry and Nolen, 2018). The goal of this study was to add knowledge to the area of AI governance in a manner that is open to scrutiny of the analytical process.

3.10 Ethical Issues

Saunders et al. (2007) describe the importance of three key areas to avoid ethical issues when conducting research. It is important that a researcher maintains objectivity during data analysis to avoid misrepresentation of the data collected. They need to ensure the confidentiality and anonymity of the participants to protect them from unintended harm or potential embarrassment. Additionally, research questions should be designed in a way that does not disadvantage a participant in the future. This study sought to uphold the highest ethical standards and adopted the following process to ensure objectivity and privacy.

Prior to the interviews, participants received an information sheet and consent form detailing the study's scope, and objectives, with a copy of the semi-structured research questions that were asked during the interview, emphasising the voluntary nature of their participation and the stringent confidentiality measures in place to protect their anonymity. They were required to sign the document digitally and return it via email to confirm their consent. The interviews were recorded for transcription purposes to facilitate later analysis. The data collected is securely stored on a password protected external hard drive and will be held for a period of five years, in compliance with NCI policy and observing GDPR rules, and will be destroyed thereafter. As the research did not encompass sensitive topics, moral dilemmas were not anticipated. However, should any issues arise, it will be addressed with the guidance of the research supervisor. There are no identified conflicts of interest that could potentially bias the research or its findings.

3.11 Limitations to Research

According to Saunders et al. (2019) proper evaluation during a study not only involves an analysis of the data but also an awareness of what did or did not work to understand the limitations of the study. Transparency about a study's limitations also enhances the credibility of the research and identifies direction for future studies (Greener, 2018).

This study was based on a small sample size of seven participants which was limited in comparison to the population of the global technology workforce. A larger sample size may have added further insights into the topic. It was initially hoped that the sample population would include an equal number of males and females as well as personnel who were involved in the design and monitoring of global AI governance strategies, but this was not possible due to time and access issues. Finally, the evolving nature of AI and its governance policies mean that this study's findings and recommendations may quickly become outdated as new technologies and regulations emerge. It is likely that this study will form part of an increasing narrative about the subject for a considerable time to come.

Chapter 4 - Research Findings and Discussion

4.1 Introduction

This chapter discusses the findings from the thematic analysis conducted on the qualitative data collected in semi-structured interviews with professionals from the technology sector. The inductive reasoning adopted for the study allowed for analysis of data to begin at the pilot interview that was conducted at the start of the process. This iterative strategy helped refine the research process by focusing the study on themes that emerged in the first semi-structured interview that were central to achieving the study's objectives (Shaw, 1999).

The qualitative approach allowed for theories to be developed about an area with significant research gaps. The seven participants from various demographic, hierarchy and aspects of the technology industry were selected to offer insights from a diverse perspective. Despite the diversity of the population, several common themes emerged through the analysis of the data that are key to governing AI now and in the future. These include: the complexity of the challenge ahead, the need for transparency and cooperation, increased AI literacy through education and the requirement for new forms of agile governance models to ensure the safe and ethical development of AI.

This chapter applies these themes and others to answer the research questions outlined in Chapter 2 which were essential to the study's objectives. It critically analyses the findings to add to the narrative around AI governance, discussing how they compare to current literature, as well as identifying where further research is warranted.

4.2 Objective 1: Defining the problem.

One of the key objectives of this study is to define the challenges facing society in governing AI's development effectively. Agreeing on a universal definition of AI presents challenges, and with no global standard, there exists a variety of definitions in circulation making it difficult to define the problems ahead (Sheikh, Prins, and Schrijvers, 2023). The problem is enhanced by, the EU, US and China having different interpretations of what constitutes an AI system. AI has been a concept for over sixty years now, but its usage has surged over the last decade. This rapid growth combined with its impact on various aspects of society including political, economic, and cultural make it challenging but imperative to decide what kind of future society we want as AI continues to shape it (Ulnicane et al, 2021).

Effective governance is crucial to ensure AI's positive impact on society. In order to govern the technology effectively, there needs to be a clear and concise definition of what constitutes AI to identify what system should be included under its governance to prevent regulatory problems now and in the future (Roberts et al, 2023).

4.2.1 Is the term AI too broad and ambiguous?

The qualitative analysis conducted for this study revealed a consensus that the term AI is too broad and ambiguous. 'The term is definitely very broad, which leads to misunderstandings from people of what AI is. This leads to a lot of oversimplification, because there is not enough education about AI and I think that causes a big problem in terms of, how are you going to govern something if you don't even know what it is' (Participant 3).

The participants cited the diversity of its applications as a central cause of the difficulty in defining AI. Another common theme that emerged as a cause of AI's ambiguous nature was a lack of education, causing misconceptions about the technology's capabilities. The progress from basic algorithms to LLM's such as copilot has altered peoples understanding of AI which has exacerbated fears and expectations about its capabilities. The pace of change also caused confusion in defining what sits under its umbrella, with a large portion of the sample interviewed exhibiting a limited understanding of AI system that were not relevant to their daily work. This confusion is enhanced by the fact that many AI technology's such as ASI are still in the theoretical phase of development and difficult for most people to grasp their mechanics.

Clearly defining what constitutes an AI system is likely to become more difficult over the coming years. Education is a theme that is central to many of this study's aims. Ordinary people need to be taught about AI, to understand its use and application. For people to have an informed opinion on AI governance, they need to be made aware of the implications for them. If we want our next generation to guide the direction of AI development, we need to improve AI literacy, particularly as the technology becomes more complex (Ng et al, 2021).

4.2.2 What challenges does the scale and complexity of AI bring to its effective governance?

The difficulty in defining the issue of AI governance is complicated in several ways by the scale and complexity of the technology. AI is utilised by companies across a wide variety of industries such as medicine, aeronautical, automotive, and manufacturing making it a challenge to create a blanket legislation that complies with the policies of the individual sectors; therefore, individual industries may require specific governance clauses. The rapid pace of the technologies development challenges regulatory bodies to keep pace with its advances. *'The technology industry operates at much faster pace than the legislative world'* (Participant 5). This issue is further enhanced by the need for global cooperation, not only from the legislative bodies but also from the technology companies designing the systems.

As AI system rely on vast amounts of data to operate, it causes ethical and privacy concerns about the data source and storage, and the prevision of consent for its use. If AI systems advance and make decisions autonomously, it will become more difficult to assign accountability for their actions (Dwivedi et al, 2021a).

To overcome this, three common themes emerged from the study's interviews. Collaboration between private companies across the various industries that use AI and the regulatory bodies that govern the technology is essential. If one territory has lax AI laws compared to the rest of the world, it is likely that less scrupulous companies will conduct their research and development in this territory to avail of the lack of regulation, negating the effectiveness of legislation in other territories.

Accountability was also regarded as essential to overcome the complexity of the task. By holding companies responsible for their actions and sanctioning them when acting inappropriately, it could foster responsible AI development. However, interviewees expressed scepticism as to the current penalties effectiveness in enhancing accountability due to the resources of the corporations driving its development.

The study also highlighted transparency as a key theme in overcoming the complexity of effective governance. This topic be discussed in the next segment of this chapter but as with accountability, there was scepticism about whether the large corporations leading AI's development will be transparent about their actions given the potential financial gains involved.

4.2.3 Is greater transparency a key parameter to achieving success in AI governance?

Transparency as with education is a theme that was common from all participants. The complexity of some AI systems results in those who have designed the systems are the only people who understand how outputs are reached. The financials benefit for companies leading the technology's development are immense and because of this, there has been a lack of transparency as to how systems work, the source of the data and whether any bias was inherent in the system that achieved the outcomes. Participant 2 described an irony in Open AI's name, the parent company of chatbot Chat GPT *' the irony of being called Open AI when it is pretty much closed'*, which appears to mirror the beliefs of all of the interviewees who expressed doubt as to whether AI companies will be transparent in their actions. This theme is echoed by the findings of Bommasani et al. (2023) that reported an average transparency score of 37 out of 100 for the ten major developers of AI across indicators such as data usage, risk, capabilities, and policies.

Felzmann et al (2020), also note how integrating transparency into the design and use of AI system is important but will not be easy due to the pace of the technology, the different dimensions to transparency as a concept and where it is required as well as how to communicate knowledge with different stakeholders and integrate it into accountability measures. Ulnicane et al. (2021) describe how for AI to deliver on its potential, AI needs a global collaborative where it is not important whether one country or company dominates. Whether this happens remains to be seen but transparency will play a vital role in achieving effective governance. It may also foster public trust in AI which is an essential aspect of accepting the technology into the workplace and into society in general, by showing people how a system works and makes decisions, and who is accountable for its actions.

4.2.4 What role can ordinary people play in designing an effective governance strategy, if any?

The qualitative data suggests transparency and education are also key in enhancing the role of ordinary people in designing effective governance strategies. The study found that ordinary people have a minor role to play at the moment because *'the majority of people don't know the difference between machine learning (ML) and AI or LLM, the current role of an ordinary person is as a guinea pig for AI developers'* (Participant 5)

The complexity of AI systems and the lack of transparency from the systems creators means that ordinary people have little understanding of how systems work and thus struggle to analyse the effectiveness of a governance strategy or add to the narrative in designing new ones. This naivety causes insufficient scrutiny of AI governance proposals and of the practices of the leading developers. Also, it means that the general public's opinions may be based on misconceptions, incomplete information or the opinion of a technologist they choose to follow on social media. If people have limited knowledge about AI, they struggle to understand the implications of any new AI law. A lack of understanding may cause confusion, leading to resistance or apathy to new measures making it harder for regulatory bodies to gain support for new strategies (de Almeida et al., 2021). This study found that to overcome these issues, greater AI literacy is required throughout society and not just the technology industry. This aligns with NG et al. (2021) who found it essential that AI becomes a fundamental skill for everyone to guide the technology in a positive way. They recommend that global school curriculum be adapted to include AI literacy, so that future generations are better equipped to influence AI development and governance.

4.3 Objective 2: Govern the present. Are the current regulatory approaches adequate?

The second objective of this study was to assess the current regulatory approaches of the leading territories in AI development. The data found a gap in knowledge among the interview candidates as to the workings of the EU, US, and China's approach to governing the technology, with only one candidate having a strong familiarity with the different strategies. It appears that technology sector professionals rely on their employer to interpret legislation and ensure they are compliant. Despite the lack of awareness among the population regarding regulatory approaches, several common themes emerged during the process which helped answer the research questions central to achieving the study's second goal. They included transparency, accountability, agile governance mechanisms, auditing, stricter enforcement mechanisms and industry specific legislation.

4.3.1 Are current regulatory approaches adequate to address the recent level of AI development?

A recurring theme in the qualitative data was that the current approaches would not be effective. One of the limitations of AI governance is that it requires global cooperation. While some international legislation is designed to promote safety and security, other policies are designed to promote innovation, similarly industries such as the financial sector may require legislation to prevent fraud while healthcare needs legislation to safeguard patients and protect their privacy. The complexity of the problem means that designing a blanket policy will be difficult. Also, the study found that companies are likely to move their operations to territories where the legislation best suits their ambitions for AI development. This aligns with Higgins at al. (2023) analysis of the EU AI act citing the complexities of the policy as well as the penalties for breaches of the regulation as a reason that many start-up companies will move their operations to the US where the laws are laxer.

Another common theme is that legislation by its nature is not agile enough to keep pace with the technology industry. The final draft of the EU AI act was published in June 2024, yet it does not become law for another two years. The development of AI has been staggering over the last two years with LLM now able to analyse significantly more data than their counterparts from two years ago meaning that the legislation that was designed to govern this aspect of AI may be outdated by the time the law comes into effect.

Also, different territories and companies are likely to interpret and govern the technology that best suits their own interests. Some large technology companies have historically interpreted legislation in a way that best suited their operations, weighing the potential benefits for non-compliance against the cost of the fines or penalties associated with breaches of rules (Clayton, 2022). The challenge for the regulatory bodies lies in finding a way of working collectively to design models that are agile enough to keep pace with the technology. A successful collaboration will involve the transparent participation of the companies leading the technologies development which may prove difficult. A global collaboration like this is possible though as witnessed during the Covid-19 pandemic (Aunger et al., 2022).

4.3.2 Would Industry specific policies be more effective in governing AI development and deployment?

All the study's participants agreed that industry specific legislation would be more effective in governing the technology. 'I'm not sure how someone from the finance industry could add to the narrative around the measures needed for medicine for example, they don't know what is needed or not need' (Participant 4). Ulnicane et al. (2021) agrees in part with this theme as governance at an industry specific level means that a diverse range of stakeholders are involved in the narrative as well as the large companies that are driving the technologies development. Regulations that are tailored to a specific sector can adapt to the unique challenges AI poses to that industry. However, they see the involvement of individual states as key though to ensure that governance strategies benefit all of society, not allowing one sector to adopt practices that give them an advantage over others. Although the data suggests that industry specific guidelines would provide clearer guidance for a company within a sector making it easier for them to comply with regulations, the recurring themes of transparency, collaboration, and accountability emerged as the central in achieving a successful system. Without these actions, industry specific governance may fail or become inconsistent as different sectors prioritise the specific need of their industry. This is why there exists a need for a body with overall control of the direction of AI governance within a territory (de Almeida et al., 2021).

4.3.3 How can the global governing bodies design legislation that fosters innovation with the requirement for regulations to prevent bias and protect privacy?

There was a common consensus on the need to prevent bias and protect people's privacy among the candidates. How this can be achieved created much debate. The data suggests that in current AI systems, it is impossible to eliminate bias as the systems are designed by humans and designers are likely to unintentionally and inadvertently incorporate a bias into a system. Transparency is key to reduce bias, as it will allow governing bodies to monitor how systems are developed and trained. Transparency is also essential in protecting people's privacy. It is important that regulators can see how data was mined to create individual systems, as well as how it was stored and deployed. Balancing these two ethical dilemmas with innovation will be difficult. Ulnicane et al. (2021) suggest that while it is important to have regulations to prevent risks associated with AI, it is also important not to have too many or it will slow down or stop innovation. One way of overcoming the innovation versus regulation dilemma would be to change the whole culture in the technology industry. Participant 3 suggests that '*fines won't work, money means nothing to these large companies, it has to be about their reputation*'. If AI's leading companies adopted collectively high ethical standards, the need for heavy regulation would not exist. This approach would require companies to take responsibility for their actions, creating a culture where poor behaviour is unacceptable and companies adopting unethical practices are shunned and boycotted by society and other businesses. A change towards greater self-regulation is needed as the next segment of this chapter will discuss why the current enforcement mechanisms are unsuccessful.

4.3.4 How effective are the current enforcement mechanisms and will they be successful in ensuring the ethical development of AI?

The current enforcement mechanisms are reactive and viewed as ineffective in ensuring the ethical development of AI by the interview candidates. Fining a company $\notin 100$ million when their market cap is $\notin 1$ trillion is not an adequate deterrent to bad behaviour. Financial penalties are often absorbed by large corporations as a cost of doing business. Companies utilise public relations (PR) firms after a breach of policy to create a narrative that removes blame from their actions and use prominent law firms to extend legal proceedings elongating the process of achieving compliance with laws.

Additionally, the current mechanisms also slow innovation by smaller firms as they fear receiving financial penalties that may bankrupt them. This enhances the larger industry players position as it creates less competition, but also hinders overall innovation.

de Almeida et al. (2021) describe regulation and its enforcement mechanism as a means of altering behaviours. However, the central theme in the data suggest that the current mechanisms are unlikely to alter behaviour and will not prevent unethical actions by the companies leading AI development. It does indicate that a new agile hybrid governance strategy is needed where leading players are regularly audited to ensure they are complying with regulations. The success of this strategy will involve international cooperation and a cohesive approach because companies will naturally conduct research and development in territories where the laws are laxer.

4.4 Objective 3: Anticipate the future

As described in chapter 2, the Collingridge dilemma (1980), technological determinism, and the pacing problem outline the challenges that will need to be overcome to successfully guide the direction of AI. The pacing problem describes how regulation consistently struggles to keep pace with innovation. The Collingridge dilemma explains that when a technology is in its infancy, there is not enough known about its pitfalls to enact concise governance strategies and by the time its pitfalls become known the technology is too entrenched in society to enact effective changes. Advocates of technological determinism believe that new technologies are naturally going to shape society outside of human control.

The final objective of this study is to ask whether society can guide the direction of AI development effectively, overcoming the pacing problem and technological determinism. It also attempts to discover how this might be achieved, how the Collingridge dilemma can be overcome and what challenges it will pose. Finally, it asks the industry experts whether development should be slowed to allow legislation keep pace with the technology's advances.

4.4.1 What are the future challenges for governing bodies. Is it possible to anticipate and prepare for them?

The study found that the challenges currently faced by governing bodies such as preventing bias, and protecting privacy will continue to be an issue for future regulators. The issues are likely to persists due to the way AI systems are trained using data that often contains existing bias, as well as the large amount of data required to operate different systems. Without change, we risk making similar mistakes to those encountered during the advent of the internet and social media, although the consequences of getting AI's governance wrong may be more profound because AI may be used as a means of increasing an entity's power rather than enhancing society as a whole (Chatham house, 2024).

However, these challenges are likely to become more pronounced due to the pace of the technology's development. As system become more advanced, they could potentially create new AI systems which will significantly accelerate the development of the technology. As AI approaches 'singularity', where they surpass human intelligence, it is impossible to comprehend their potential outputs (Kurzweil, 2024). This opinion was echoed by the interviewees who agreed that due to the complexity of the technology, it is very difficult to anticipate its future challenges which supports the Collingridge dilemma. Transparency, collaboration, education and accountability are seen as essential in overcoming these

challenges. Agility is a key feature in the design of future legislation if it is to overcome the pacing problem, but as this will involve the collaboration of multiple stakeholders, it is considered a further challenge and while the data did not show a belief in technological determinism, it did indicate that one of the challenges that lay ahead would be designing governance strategies that allowed society dictate the direction of the technology rather than a small number of companies.

4.4.2 What strategies can be employed to ensure that legislation keeps pace with the technology's rapid development.

To ensure that regulations keep pace with the technology's development, transparency and agility were the common theme in the qualitative data offered as a solution. The current systems were deemed cumbersome, and inadequate to keep pace with developments which aligns with Taeihagh (2021) call for a reassessment of the efficacy of the current governance approaches. There are also information gaps between governments and the companies driving AI development that need to be addressed. Without clarity on how AI systems work, it is difficult to design effective governance strategies.

To overcome the first hurdle will involve designing governance frameworks that are agile and easily adapted to keep pace with the technology. (Participant 6) describes the current models as 'overly bureaucratic to keep pace with a technology that is moving so fast'. The creation of agile laws would allow quick action to deal with evolutions of the technology without the need for an overhaul of the entire framework. As part of this, the data suggests the need for dedicated bodies to monitor developments with the authority to conduct regular audits on the companies involved in the industry. These bodies need to be equipped with the potential to impose sanctions on companies that are misbehaving that go beyond the current controls.

The qualitative data also suggests the need for a change of culture in the industry as a whole, where bad behaviour is not tolerated, and transparency is promoted as a norm to help the legislator direct the development of AI in a positive way that maximises its potential to society. This will involve global cooperation from all stakeholders, or as discussed previously, companies may gravitate to territories with laxer AI laws.

4.4.3 Is it possible to design governance models that guide the direction of AI development in a way that enhances society or is the technology destined to follow its own course?

Collaboration and learning from past mistakes in governing emerging technologies were the themes central to this topic which aligns with Taeihagh (2021) call for cross-cultural cooperation, adopting regulations previously used to govern emerging technologies, using the lessons learned from these strategies to improve the effectiveness of AI governance.

The lax approach adopted in the early stages of the internet led to issues with data and privacy that saw reactive laws such as Europe's GDPR enacted to protect people's data and right to privacy, highlighting the risk of delaying intervention. Future governance models must be proactive rather than reactive due the potential pitfalls of AI that may pose an existential threat (Kurzweil, 2024). This will be challenging due to the reach and complexity of the technology. Collaboration is central to designing effective governance models. AI is a technology with global reach spanning various industries and because of this, governance models will only be as effective as the weakest link. International cooperation is needed to create consistent policies that prevent unethical behaviour by companies in a region because the laws are less stringent.

However, the data suggests that this may be difficult to achieve due the influence of the companies driving the technology's development as well the competing national interests of the territories leading AI development. Without collaboration, the technology may not follow its own course, but it is likely to follow the direction of a small group of companies or countries. Despite this, recent history has shown that international multi stakeholder cooperations is possible through the design of legislation to combat climate change as well the cooperative effort that was used to overcome the Covid-19 pandemic.

4.4.4 Should AI development be slowed to allow legislation keep pace with its advances?

This topic generated the most polarised responses from the interviewees and forms part of a greater philosophical debate about regulation over innovation. The only common theme in the data is that it is unlikely to happen. (Participant 7) described how '*you can't stop evolution, and I think technology and evolution are the same thing*', which aligns with believers in technological determinism that proposes the technology is destined to follow its own path.

There is considerable academic debate about whether the technology's development should be slowed to allow legislation keep pace with its development. Judge et al. (2024) suggest that until more is known about how AI systems work, and strategies are designed that create transparency in how to minimise risks to society that the technologies development should be slowed. Taeihagh (2021) argues that rather than slowing development there is a need to design effective governance models that protect society while at the same time encourage innovation because of the many positive possibilities that AI offers.

Current data would suggest that it is advisable to slow the development of the technology as the current regulatory approaches have struggled to keep pace with the technology's development. However, this is unlikely to happen due to the practical challenges of implementing a slowing of the technology's development. This would require the cooperation of all the stakeholders involved in the technology's development, including researchers, national governments, and the companies driving its development which is difficult given the competitive nature of the industry. Instead, the focus should be on creating new agile governance strategies that can keep pace with the technology ensuring its safe and ethical development. This involves transparency and collaboration which may be difficult to achieve.

Chapter 5 - Conclusion and Recommendations

The potential benefits from the safe and ethical development of AI and the potential consequences from its unsafe development are numerous (Dafoe, 2018). This study explored the complexity of the problem faced by legislators, where even agreeing on a consensus of what constitutes an AI system is problematic. The thesis examined the regulatory frameworks in the US, EU and China, reviewing the current literature and collecting primary data through semi-structured interviews to assess these policies' ability to govern the technology effectively. The study also probed the future challenges that the technology is likely to create, seeking advice from technology experts on how to mitigate against future issues.

The findings from the study revealed several key insights:

The term AI is too broad and needs to be more clearly defined, the complexity of the technology and its reach mean that society is facing a governance challenge unlike others.

Transparency from all stakeholders and global cooperation are essential to maximise the benefits of the technology and education to improve AI literacy is essential to give ordinary people a role in the narrative around its governance.

AI has the potential to significantly change our lives and change society, but its future direction is uncertain. It may create many benefits but there is also the concern that it could leave many people disadvantaged. The decisions made over the coming years are likely to have an impact on the current society and on the lives of future generations (Dwivedi et al, 2021a).

The insights gathered during the research determined the following recommendations to enhance AI governance and ensure that its development remains on a positive trajectory that benefits society. Further research is warranted on how to make these recommendations a reality.

- 1) Clear globally accepted definitions and classifications of AI are essential to its effective governance.
- 2) Enhanced transparency and accountability are essential for its ethical development.
- 3) AI literacy need to be improved through education to allow ordinary people play a proactive role in its governance

- 4) New agile governance models are needed to keep pace with the speed of the technology development
- 5) New enforcement mechanisms are essential for breaches of regulation that are a proper deterrent to bad behaviour by AI companies.
- 6) Global cooperation is essential or bad actors will operate in territories where the least regulation exists.
- 7) The culture needs to change within the technology industry, where bad behaviour is not tolerated.

Bibliography

Adeoye-Olatunde, O.A. and Olenik, N.L. (2021) 'Research and scholarly methods: Semistructured interviews', *Journal of American College of Clinical Pharmacy*, 4(10), pp.1358-1367. doi: 10.1002/jac5.1441

Adhabi, E. and Anozie, C. B. (2017) 'Literature Review for the Type of Interview in Qualitative Research. *International Journal of Education*. 9(3), pp. 86-97. doi: 10.5296/ije.v9i3.11483 Alshenqeeti, H. (2014) 'Interviewing as a Data Collection Method: A Critical Review', *English Linguistics Research*, 3(1), pp.39-45. doi:10.5430/elr.v3n1p39

Aunger, J.A., Millar, R., Rafferty, A.M. and Mannion, R. (2022) 'Collaboration over competition? Regulatory reform and inter-organisational relations in the NHS amidst the COVID-19 pandemic: a qualitative study', *BMC Health Services Research* 22, 640. doi.org/10.1186/s12913-022-08059-2

Azungah, T. (2018). Qualitative Research: Deductive and Inductive Approaches to Data Analysis. *Qualitative Research Journal*, 18(3), pp. 383-400. doi: 10.1108/QRJ-D-18-00035 Babbie, E.R. (2020). *The practice of social research*. 15th edn. Available at: <a href="https://books.google.ie/books?hl=en&lr=&id=lFvjDwAAQBAJ&oi=fnd&pg=PP1&dq=Babbie,+E.R.+(2020).+The+practice+of+social+research.+Cengage+Au.&ots=I4yQ3B5ON4&sig=fLS6-zcZDX-Q9Gg51qtUuxhMvv0&redir_esc=y#v=onepage&q&f=false [Accessed 23 July 2024].

Baldassarre, M.T., Caivano, D., Fernandez Nieto, B., Gigante, D. and Ragone (2023) 'The social impact of generative AI: An analysis on ChatGPT', in GoodIT '23: *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. Lisbon, Portugal, 6-8 September 2023, pp. 363-373. doi: 10.1145/3582515.3609555

Barroga, E., Matanguihan, G.J., Furuta, A., Arima, M., Tsuchiya, S., Kawahara, C., Takamiya,
Y., and Izumi, M. (2023) 'Conducting and writing quantitative and qualitative research', *Journal of Korean Medical Science*, 38(37), e291. doi: 10.3346/jkms.2023.38.e291
Beauchamp, D. (2018) *How public opinion shapes the future of AI*. Available at:
<u>https://www.smartdatacollective.com/public-opinion-shapes-future-ai/</u> [Accessed 17 May 2024].

Benbya, H., Davenport, T.H. and Pachidi, S. (2020) 'Artificial intelligence in organizations: Current state and future opportunities', *MIS Quarterly Executive*, 19(4), pp. 9-21. doi: 10.2139/ssrn.3741983

Birkstedt, T., Minkkinen, M., Tandon, A. and Mäntymäki, M. (2023) 'AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, 33(7), pp.133-167. doi: 10.1108/INTR-01-2022-0042

Bleiker, J., Morgan-Trimmer, S., Knapp, K., and Hopkins, S. (2019) 'Navigating the maze: Qualitative research methodologies and their philosophical foundations', *Radiography*, 25(1), pp. S4-S8. doi: 10.1016/j.radi.2019.06.008

Bommasani, R., Klyman, K., Kapoor, S., Longpre, S., Xiong, B., Maslej, N. and Liang, P. (2024) The Foundation Model Transparency Index after 6 months. *Stanford CRFM*. Available at: <u>https://crfm.stanford.edu/2024/05/21/fmti-may-2024.html</u> [Accessed 2 Aug. 2024].

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), pp.77-101. doi: 10.1191/1478088706qp063oa

Bughin, J., Seong, J., Manyika, J., Chui, M. and Joshi, R. (2018) 'Notes from the AI frontier: Modelling the impact of AI on the world economy'. Available at: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligen ce/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on %20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx [Accessed 25 June 2024]. Buhmann, A. and Fieseler, C. (2023) 'Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence', *Business Ethics Quarterly*, 33(1), pp.146-179. doi: 10.1017/beq.2021.42

Burt, A. (2021) New AI regulations are Coming. Is Your Organisation Ready? Available at: https://hbr.org/2021/04/new-ai-regulations-are-coming-is-your-organization-ready [Accessed 31 May 2024].

Cammaerts, B. and Mansell, R. (2020) 'Digital platform and regulation. Toward a radical democratic turn'. *International Journal of Communication*, 14, pp. 135-154. Available at: <u>https://ijoc.org/index.php/ijoc/article/viewFile/11182/2901</u> [Accessed 19 May 2024].

Carter, W., Hyman, K. and Silver, H. J. (2024) *The EU Artificial Intelligence (AI) Act: A pioneering framework for AI governance*. Available at: https://www.womblebonddickinson.com/us/insights/alerts/eu-artificial-intelligence-al-actpioneering-framework-aigovernance#:~:text=The%20AI%20Act%20categorizes%20AI,fundamental%20EU%20right s%20are%20banned [Accessed 24 June 2024].

Castleberry, A. and Nolen, A. (2018) 'Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6), pp. 807-815. doi: 10.1016/j.cptl.2018.03.019

Chatham House, (2024) *Artificial intelligence and the challenge for global governance: Nine essays on achieving responsible AI.* Research paper. Updated 10 June, ISBN: 978 1 78413 608 6. doi: 10.55317/9781784136086.

Clayton, J. (2022) 'Europe agrees new law to curb Big Tech dominance', *BBC News*, 25 March. Available at: <u>https://www.bbc.com/news/technology-60870287</u> [Accessed 03 August 2024)].

Collingridge, D. (1980) *The social control of technology*. Milton Keynes: Open University Press.

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pontes Pinto, E., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E. and de Oliveira, N. (2023) 'Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance', *Patterns*, 4(10), 100857. doi: 10.1016/j.patter.2023.100857

Creswell, J.W. and Poth, C.N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. 4th edn, Available at: https://books.google.ie/books?hl=en&lr=&id=DLbBDQAAQBAJ&oi=fnd&pg=PP1&dq=Cr eswell,+J.+W.,+%26+Poth,+C.+N.+(2017).+Qualitative+Inquiry+and+Research+Design:+C hoosing+Among+Five+Approaches.+Sage+Publications.&ots=ip19aGVPt&sig=SUQjfekm2-v9TE9mRJOJinITEYk&redir_esc=y#v=onepage&q&f=false [Accessed 23 July 2024].

Dafoe, A. (2018). 'Global politics and the governance of artificial intelligence'. Journal ofInternationalAffairs,72(1),pp.121-126.Availableat:https://research.ebsco.com/c/x47ol5/viewer/pdf/rnb7rprkfz[Accessed 09 June 2024].

Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., Wang, W.W. and Witteborn, S. (2019). AI, governance and ethics: Global perspectives. *University of Hong Kong Faculty of Law*. Research Paper No.2019-15. doi: 0.2139/ssrn.3414805

de Almeida, P.G.R., dos Santos, C.D. and Farias, J.S. (2021) 'Artificial intelligence regulation: A framework for governance', *Ethics and Information Technology*, 23, pp.505-525. doi: 10.1007/s10676-021-09593-z

Di Pietro, R. and Cresci, S. (2021)' Metaverse: Security and Privacy Issues'. in 2021Third IEEE International Conference on Trust, Privacy and Security in Intelligence Systems and Applications (TPS-ISA). Atlanta, GA, USA, 13-15 December 2021, pp. 281-288. doi: 10.1109/TPSISA52974.2021.00032

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi. Et al. (2021a) 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges. opportunities, and agenda for research. practice and policy', *International* Journal of Information Management, 57, 101994. doi: 10.1016/j.ijinfomgt.2019.08.002

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Baabdullah, A.M., Koohang, A. et al. (2021b) 'So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', *International Journal of Information Management*, 58, 102282. Available at: <u>https://www.sciencedirect.com/science/article/pii/S0268401223000233#bib190</u> [Accessed 15 May 2024].

Easton, K.L., McComish, J.F. and Greenberg, R. (2000) 'Avoiding common pitfalls in qualitative data collection and transcription', *Qualitative Health Research*, 10(5), pp.703-707. doi: 10.1177/104973200129118651

Edwards, R. and Holland, J. (2013) What is qualitative interviewing? Available at: https://library.oapen.org/viewer/web/viewer.html?file=/bitstream/handle/20.500.12657/58752 /9781849668026.pdf?sequence=1&isAllowed=y [Accessed 23 July 2024].

European Parliament (2024) *EU AI Act: First regulation on artificial intelligence*. Available at: <u>https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-</u>

intelligence#:~:text=The%20Parliament%20adopted%20the%20Artificial%20Intelligence%2 0Act%20in,Council%20followed%20with%20its%20approval%20in%20May%202024 [Accessed 22 June 2024].

Fanni, R., Steinkogler, V.E., Zampedri, G. and Pierson, J. (2023) 'Enhancing human agency through redress in artificial intelligence systems. *AI & Society*, 38, pp.537-547. doi: 10.1007/s00146-022-01454-7

Fazlioglu, M. (2023) US federal AI governance: Laws, policies and strategies. Available at: <u>https://iapp.org/resources/article/us-federal-ai-governance/</u> [Accessed 09 June 2024].

Felzmann, H., Fosch-Villaronga, E., Lutz, C. *and* Tamò-Larrieux, A. (2020) 'Towards Transparency by Design for Artificial Intelligence', *Sci Eng Ethics* 26, pp. 3333–3361. doi.org/10.1007/s11948-020-00276-4

Fossey, E., Harvey, C., McDermott, F. and Davidson, L. (2002) 'Understanding and evaluating qualitative research', *Australian & New Zealand Journal of Psychiatry*. 36(6) pp.717-732. doi: 10.1046/j.1440-1614.2002.01100.x

Gbadegeshin, S.A., Al Natsheh, A., Ghafel, K., Tikkanen, J., Gray, A., Rimpilainen, A., Kuoppala, A., Kalermo-Poranen, J. and Hirvonen, N. (2021) 'What is artificial intelligence (AI): A simple buzzword or a worthwhile inevitability?', in *ICERI2021 Proceedings*, 14th annual International Conference of Education, Research and Innovation. 8-9 November 2021, pp. 468-479. doi: 10.21125/iceri.2021.0171

Genus, A. and Stirling, A. (2018) 'Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, 47(1), pp.61-69. doi: 10.1016/j.respol.2017.09.012

Ghauri, P. and Gronhaug, K. (2010) Research methods in Business Studies. 4th edn. Harlow: FT Prentice Hall.

Gherghina, S., Mokre, M. and Miscoiu, S. (2020) 'Introduction: Democratic deliberation and underrepresented groups', *Political Studies review*, 19(2), pp. 153-163. doi: 10.1177/1478929920950931

Gikay, A.A. (2024) 'Risks, innovation and adaptability in the UK's incrementalism versus the European Union's comprehensive artificial intelligence regulation', *International Journal of Law and Information Technology*, 32, eaae013. doi: 10.1093/ijlit/eaae013

Gleckman, H. (2018) *Multistakeholder governance and democracy: A global challenge*. London: Routledge. doi: 10.4324/9781315144740

Goertzel, B. (2014) 'Artificial general intelligence: Concept, state of the art, and future prospects', *Journal of Artificial Intelligence*, 5(1), pp. 1-46. doi: 10.2478/jagi-2014-0001

Gong, J., Qu, H. and Dorwart, H. (2024) AI governance in China: Strategies, initiatives, and key considerations. Available at: <u>https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations</u> [Accessed 22 June 2024].

Grass, K. (2024) 'The three logics of qualitative research: Epistemology, ontology, and methodology in political science', *American Journal of Qualitative Research*. 8(1), pp.42-56. doi: 10.29333/ajqr/14083

Greener, S. (2018) 'Research limitations: the need for honesty and common sense', *Interactive Learning Environments*. 26(5) pp. 567-568. doi: 10.1080/10494820.2018.1486785

Groumpos, P.P. (2021) 'A critical historical and scientific overview of all industrial revolutions', *IFAC-PapersOnLine*, *54*(13), pp.464-471. doi: 10.1016/j.ifacol.2021.10.492

Hall, P.A. (2003) 'Aligning ontology and methodology in comparative Politics'. in Mahoney,
J. and Rueschemeyer, D. (eds.) *Comparative historical analysis in the social sciences*, pp. 373-404. doi: 10.1017/CBO9780511803963.012

Hallström, J. (2022) 'Embodying the past, designing the future. Technological determinism reconsidered in technology education. *International Journal of Technology and Design Education*, 32, pp.17-31. doi: 10.1007/s10798-020-09600-2

Hayes, M. (2022) 'Incrementalism and Public policy-making', *Oxford Research Encyclopaedia of Politics*. doi: 10.1093/acrefore/9780190228637.013.133

Higgins, T., Jackson, R. and Korten, P. (2023) 'Understanding the EU AI Act: Concerns and
criticisms', Clifford Chance. 6 April. Available at:
https://www.cliffordchance.com/insights/resources/blogs/talking-
tech/en/articles/2023/04/the-eu-ai-act--concerns-and-criticism.html [Accessed 24 June 2024].

Hilb, M. (2020) 'Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance', *Journal of Management & Governance*, 24, pp.851-870. doi: 10.1007/s10997-020-09519-9

Heikkila, M. (2023) 'Five things you need to know about the EU's new AI Act', MITTechnologyReview.Availableat:https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/[Accessed 24 June 2024].

Hopgood, A. (2005) 'The state of artificial intelligence', *Advances in Computers*, 65. pp. 1-75. doi: 10.1016/S0065-2458(05)65001-2

Huttenlocher, D., Ozdalgar, A. and Goldston, D. (2023) *A Framework for US AI Governance: Creating a safe and thriving AI sector*. Available at: <u>https://computing.mit.edu/wp-content/uploads/2023/11/AIPolicyBrief.pdf</u> [Accessed: 15 July 2024].

Ignatius, A. and Bernstein, A. (2023) 'How generative AI changes strategy', HBR Idea Cast, [Podcast]. 5 May. Available at: <u>https://hbr.org/podcast/2023/05/how-generative-ai-changes-strategy</u> [Accessed 15 May 2024].

Jones, K. (2023) *AI governance and human right: Resetting the relationship*. Available at: <u>https://www.chathamhouse.org/sites/default/files/2023-01/2023-01-10-AI-governance-</u> <u>human-rights-jones.pdf</u> [Accessed 15 July 2024].

Jones, J., Casovan, A., Chaudhry, U., Bryant, N., Resmerita, L. and Spadea, M. (2024) *AI Governance in practice Report* 2024. Available at: <u>https://iapp.org/media/pdf/resource_center/ai_governance_in_practice_report_2024.pdf</u> [Accessed 15 July 2024].

Judge, B., Nitzberg, M and Russell, S (2024) 'When code isn't law: rethinking regulation for artificial intelligence', *Policy and Society*, puae020, <u>doi.org/10.1093/polsoc/puae020</u>

Kar, A.K. (2016) 'Bio inspired computing – A review of algorithms and scope of applications', *Expert Systems with Applications*, 59, pp. 20-32. doi: 10.1016/j.eswa.2016.04.018

Kelly, S., Kaye, S.-A. and Oviedo-Trespalacios, O. (2023) 'What factors contribute to the acceptance of artificial intelligence? A systematic review', *Telematics and Informatics*, 77, 101925. doi: 10.1016/j.tele.2022.101925

Knott, E., Rao, A.H., Summers, K. and Teeger, C. (2022) 'Interviews in the social sciences', Nature *Reviews Methods Primers*, 2(1), p.73. doi: 10.1038/s43586-022-00150-6

Kuhlmann, S., Stegmaier, P. and Konrad, K. (2019) 'The tentative governance of emerging science and technology- A conceptual introduction', Research *Policy*, 48(5), pp.1091-1097. doi: 10.1016/j.respol.2019.01.006

Kurzweil, R. (2024). The singularity is nearer: When we merge with AI. United States: Vintage Publishing

Lane, M., Williams, M. and Broecke, S. (2023) 'The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers', *OECD Social, Employment and Migration Working Papers, No.287,* Available at: https://one.oecd.org/document/delsa/elsa/wd/sem(2023)6/en/pdf [Accessed 25 June 2024].

Leslie, D., Rincón, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, S.J., Burr, C., Aitken, M., Katell, M. and Fischer, C. (2023) *AI Ethics and Governance in Practice: An Introduction*. Available at: <u>https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-ai-ethics-an-intro_1.pdf</u> [Accessed 22 July 2024].

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R., and Clark, J. (2024) Artificial intelligence index report 2024. Available at: <u>https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf</u> [Accessed 22 July 2024].

Matthias, A. (2004) 'The responsibility gap: Ascribing responsibility for the actions of learning automata', *Ethics and Information Technology*, 6(3), pp. 175-183. doi: 10.1007/s10676-004-3422-1

Melnikovas, A. (2018) 'Towards an explicit research methodology: Adapting research onion model for futures studies', *Journal of Futures Studie*, 23(2), pp. 29-44. doi: 10.6531/JFS.201812_23(2).0003

62

Mukherjee, S.P. (2019) A guide to research methodology: An overview of research problems, task and methods. CRC Press. doi: 10.1201/9780429289095

Myers, M.D. (2008) Qualitative research in business & management 3rd ed. SAGE Publications. Available at: <u>https://uk.sagepub.com/en-gb/eur/qualitative-research-in-business-and-management/book244733#preview</u> [Accessed 13 July 2024].

Nassar, A. and Kamal, M. (2021) 'Ethical dilemmas in AI-powered decision-making: A deep dive into big data-driven ethical considerations', *International Journal of Responsible Artificial Intelligence*, 11(8), pp. 1-11 Available at: <u>https://neuralslate.com/index.php/Journal-of-Responsible-AI/article/view/43/28</u> [Accessed 17 May 2024].

Ng, D.T.K., Leung, J.K.L., Chu, S.K.W. and Qiao, M.S. (2021) 'Conceptualizing AI literacy: An exploratory review'. *Computers and Education: Artificial Intelligence*, *2*, p.100041. doi: 10.1016/j.caeai.2021.100041

Novelli, C., Casolari, F., Hacker, P., Spedicato, G. and Floridi, L. (2024) *Generative AI in EU Law: Liberty, privacy, intellectual property and cybersecurity.* doi: org/10.48550/arXiv.2401.07348

Ortiz-Ospina, E. (2019) The rise of social media. Available at: <u>https://ourworldindata.org/rise-of-social-media</u> [Accessed 19 May 2024].

Papagiannidis, E., Enholm, I.M., Dremel, C. Mikalef, P., Krogstie, J. (2023) 'Toward AI governance: Identifying best practices and potential barriers and outcomes', *Inf Syst Front*, 25(1), pp. 123–141. doi: 10.1007/s10796-022-10251-y

Payton, B. (2023) 'Power mad: AI's massive energy demand risks causing major environmental headaches' Reuters 4 December. Available at: <u>https://www.reuters.com/sustainability/climate-energy/power-mad-ais-massive-energy-</u> <u>demand-risks-causing-major-environmental-headaches-2023-12-04/</u> [Accessed 19 May 2024]. Rathi, R.A. (2018) 'Artificial Intelligence and the future of HR practices', *International Journal of Applied Research*, 4(6), pp.113-116. Available at: <u>https://www.allresearchjournal.com/archives/2018/vol4issue6/PartB/4-6-18-226.pdf</u> [Accessed 25 June 2024].

Raymond, M. and DeNardis, L. (2015) 'Multistakeholderism: Anatomy of an inchoate global institution. *International Theory*, 7(3), pp. 572-616. doi: 10.1017/S1752971915000081

Roberts, H., Hine, E., Taddeo, M. and Floridi, L. (2023) 'Global AI governance barriers and pathways forward', *International Affairs*, 100(3), pp.1275-1286. doi: 10.1093/ia/iiae073

Roser, M. (2022) Artificial intelligence is transforming our world – it is on all of us to make sure it goes well. Available at: <u>https://ourworldindata.org/ai-impact</u> [Accessed 30 May 2024].

Salsone, B., Stein, P. S., Parsons, K. G., Kent, T., Nielsen, K. and Nitz, D. T. (2022) 'Technological determinism', *in Science technology and society a student led exploration*. Available <u>https://opentextbooks.clemson.edu/sciencetechnologyandsociety/chapter/technological-</u> <u>determinism/</u> [Accessed 19 May 2024].

Santoni de Sio, F. and Mecacci, G. (2021) 'Four responsibility gaps with artificial intelligence: Why they matter and how to address them', *Philosophy & Technology*, 34(4), pp. 1057-1084. doi: 10.1007/s13347-021-00450-x

Saunders, M., Lewis, P. and Thornhill, A. (2007) Research methods for business students. 4th edn. Harlow: Pearson.

Saunders, M., Lewis, P. and Thornhill, A. (2009). *Research methods for business students*. 5th edn. Harlow: Pearson.

Saunders, M., Lewis, P. and Thornhill, A. (2019). *Research methods for business students*. 8th edn. Harlow: Pearson. Savai, S.F. (2016) 'Reflection on ontological, epistemological and methodological perspectives: What is the difference between qualitative and quantitative research? *International Journal of Innovative Research and Advanced Studies*, 3(9), pp. 70-72. Available at: https://www.ijiras.com/2016/Vol_3-Issue_9/paper_14.pdf [Accessed 22 July 2024].

Shams, R.A., Zowghi, D. and Bano, M. (2023) 'AI and the quest for diversity and inclusion: A systematic literature review', *AI and Ethics*. 1-28. doi: 10.1007/s43681-023-00362-w

Shaw, E. (1999), "A guide to the qualitative research process: evidence from a small firm study", *Qualitative Market Research*, Vol. 2 No. 2, pp. 59-70. doi.org/10.1108/13522759910269973

Sheehan, M. (2024) Tracing the roots of China's AI regulations: How face swap apps, investigative journalism and corporate thought leadership shaped governance. Available at: https://carnegieendowment.org/research/2024/02/tracing-the-roots-of-chinas-ai-regulations?lang=en [Accessed 22 June 2024].

Sheikh, H., Prins, C. and Schrijvers, E. (2023) 'Artificial Intelligence: Definition and Background', in *Mission AI. Research for Policy. Springer, Cham.* doi.org/10.1007/978-3-031-21448-6_2

Sioli, L. (2021) 'A European strategy for artificial intelligence', *CEPS webinar – European approach to the regulation of artificial intelligence*. Available at: <u>https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf</u> [Accessed 24 June 2024].

State Council of the People's Republic of China (2017) China issues guidelines on artificialintelligencedevelopment.Availableat:https://english.www.gov.cn/policies/latest_releases/2017/07/20/content_281475742458322.htmm[Accessed 22 June 2024].

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., & Leyton-Brown, K. (2022) Artificial Intelligence and life in 2030: The one-hundred-year study on Artificial Intelligence. doi: 10.48550/arXiv.2211.06318

Sun, T.Q. and Medaglia, R. (2019) 'Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare', *Government Information Quarterly*, 36(2), pp. 368-383. doi: 10.1016/j.giq.2018.09.008

Sundar, S.S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAII). *A Journal of Computer-Mediated Communication*, 25(1), pp. 74-88. doi: 10.1093/jcmc/zmz026

Sutrisna, M. (2009) 'Research methodology in doctoral research: Understanding the meaning of conducting qualitative research', *Proceedings of the Association of Researchers in Construction Management (ARCOM) Doctoral Workshop*. Liverpool, UK, 12 May 2009, pp. 48-57. Available at:

https://www.researchgate.net/publication/275956584 Research Methodology in Doctoral Research Understanding the Meaning of Conducting Qualitative Research [Accessed 22 July 2024].

Taeihagh, A. (2021) 'Governance of artificial intelligence'. *Policy and Society*, 40(2), pp. 137157. Available at: <u>https://academic.oup.com/policyandsociety/article/40/2/137/6509315</u>
[Accessed 03 June 2024].

Tallberg, J., Erman, E., Furendal, M., Geith, J., Klamberg, M. and Lundgren, M. (2023) 'Theglobal governance of artificial intelligence: Next steps for empirical and normative research,InternationalStudiesReview,25(3),viad040Availableat:https://academic.oup.com/isr/article/25/3/viad040/7259354[Accessed 13 July 2024].

Tirole, J. (2021) 'Competition and the industrial challenge for the digital age', *Annual Review* of *Economics*, 15, pp. 673-605. doi: 10.1146/annurev-economics-090622-024222

Tizhoosh, H.R. and Pantanowitz, L. (2018) 'Artificial intelligence and digital pathology: Challenges and opportunities', *Journal of Pathology Informatics*, 9 (1), 38. doi: 10.4103/jpi.jpi_53_18

Tshitadi, N. (2024) Understanding the research onion (Saunders et al., 2015). Available at: https://nephsonic.com/understanding-the-research-onion-by-saunders-et-al-2016-acomprehensive-guide-for-writing-an-effective-research-methodology-chapter/ [Accessed 07 July 2024].

Ulnicane, I., Knight, W., Leach, T., Stahl, B.C. and Wanjiku, W.-G. (2020)' Framing governance for contested emerging technologies: Insights from AI policy. *Policy and Society*, 40(2), pp. 158-177. doi: 10.1080/14494035.2020.1855800

Ulnicane I, Eke D.O, Knight W, Ogoh G, and Stahl B.C. (2021) 'Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies', *Interdisciplinary Science Reviews*. doi.org/10.1080/03080188.2020.1840

Vallance, C. (2024). UK competition watchdog has 'real concerns' over big tech AI dominance. *BBC News*, 11 April. Available at: <u>https://www.bbc.com/news/technology-68789880</u> [Accessed 17 May 2024].

van Klyton, A., Arrieta-Paredes, M.P., Palladino, N. and Soomaree, A. (2023) 'Hegemonic practices in multistakeholder internet governance: Participatory evangelism, quiet politics, and glorification of status quo at ICANN meetings', *The Information Society*, 39(3), pp. 141-157. doi: 10.1080/01972243.2023.2194295

van Teijlingen, E.R. and Hundley, V. (2002) 'The importance of pilot studies', *Nursing Standard*, *16*(40), pp.33-36. doi: 10.7748/ns2002.06.16.40.33.c3214

Veale, M., Matus, K., and Gorwa, R. (2023) 'AI and global governance: Modalities, rationales, tensions', *Annual Review of Law and Social Science*, 19, p.p. 255-275. doi: 10.1146/annurev-lawsocsci-020223-040749

Wallace, W. and Marchant, G. (2019) 'Toward the agile and comprehensive international governance of AI and robotics. *Proceedings of the IEEE*, 107(3), pp.505-508. doi: 10.1109/JPROC.2019.2899422

Wamba, S.F., Bawack, R.E., Guthrie, C., Queiroz, M.M., and Carillo, K.D.A. (2021) 'Are we preparing for a good AI society? A bibliometric review and research agenda', *Technological Forecasting and Social Change*, 164, 120482. doi: 10.1016/j.techfore.2020.120482

Wang, W. and Siau, K. (2018) 'Ethical and moral issues with AI – A case study on healthcare robots', in *Americas Conference on Information systems (AMCIS 2018)*. New Orleans, LA, pp. 1-5 Available at:

https://www.researchgate.net/publication/325934375_Ethical_and_Moral_Issues_with_AI [Accessed 17 May 2024].

Wheeler, T. (2023). *The three challenges of AI regulation*. Available at: <u>https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/</u> [Accessed 18 May 2024].

Whyman, B. (2023) 'AI regulation is coming- what is the likely outcome?', *CSIS*. 10 October. Available at: <u>https://www.csis.org/blogs/strategic-technologies-blog/ai-regulation-coming-what-likely-outcome</u> [Accessed 22 June 24].

Wuest, J. (2012) 'Grounded theory: The method', in Munhall, P.L (ed) *Nursing research. A qualitative perspective*, 5th edn. pp. 225-256. Available at: https://books.google.ie/books?hl=en&lr=&id=3T259dKHHq8C&oi=fnd&pg=PA225&dq=gr ounded+theory+as+a+qualitative+method&ots=VYPGU6Zo3B&sig=hbPS_0vhp2kqLxqX9n _cMCikGgY&redir_esc=y#v=onepage&q=grounded%20theory%20as%20a%20qualitative% 20method&f=false [Accessed 23 July 2024].

Yang, Z. (2024) 'Four things to know about China's new AI rules in 2024'. *MIT Technology Review*. Available at: <u>https://www.technologyreview.com/2024/01/17/1086704/china-ai-regulation-changes-2024/</u> [Accessed 22 June 2024]. Zhang, C. and Lu, Y. (2021) 'Study on artificial intelligence: The state of the art and future prospects', *Journal of Industrial Integration*, 23, 100224. doi: 10.1016/j.jii.2021.100224