UNIVERSITY OF CAPE TOWN

FACULTY OF LAW

SCHOOL FOR ADVANCED LEGAL STUDIES

Examining Copyright Infringement and Liability in Generative Artificial Intelligence Training and Use: A Legal Perspective in South Africa and Beyond

Tendai Mikioni

(MKNTEN007)

Minor Dissertation as a prerequisite for completion of the LLM in Intellectual Property Law

Supervisor: Prof Caroline Ncube

Word Count: 25 819

Research dissertation presented for the approval of Senate in fulfilment of part of the requirements for the LLM in Intellectual Property Law in approved courses and a minor dissertation. The other part of the requirement for this qualification was the completion of a programme of courses.

DECLARATION

I hereby declare that I have read and understood the regulations governing the submission of LLM in Intellectual Property Law dissertations, including those relating to length and plagiarism, as contained in the rules of this University, and that this dissertation conforms to those regulations.

Signed by candidate

Signature:

Date: 27 May 2024

TABLE OF CONTENTS

DEDICATION

To my beloved family, whose unwavering love, support, and sacrifices have been the cornerstone of my life's journey.

To Merrill Carmichael, my mother, who opened her heart to me when I needed it most. Your unconditional love, guidance, and unwavering belief in my potential have profoundly shaped me.

To Gibson Mikioni, who has been my closest companion and confidant, sharing in both the joys and challenges of growing up together.

To my mentor, Alouis Sagota, who instilled in me the values of perseverance, integrity, and compassion.

To Seppo and Oili, for your dedication to empowering vulnerable children, offering hope where it is needed most.

To Zama, whose boundless love, support, encouragement and belief in my abilities have propelled me forward, even in the face of adversity.

To the children who face seemingly impossible odds, may this work contribute to a brighter future for you.

To the creators and artists whose resilience and passion inspire me daily, may your talents continue to shine amidst the challenges.

ACKNOWLEDGEMENTS

I would like to acknowledge God, from whom I draw all strength and who continuously opens new opportunities for me.

I express my deepest gratitude to Professor Caroline Ncube, whose guidance, expertise, and steadfast support have been invaluable throughout this journey. Your nerve-wracking, yet insightful, comments have shaped my art of writing.

I am indebted to the Konrad Adenauer Foundation for the generous scholarship that made this research possible. Your assistance in networking and fostering connections within the academic community has been invaluable. The opportunities to engage with like-minded individuals and share moments of laughter have greatly enriched my experience.

My sincere appreciation goes to Professor Tobias Schonwetter and the iNtaka Centre for Law and Technology team for their support throughout.

I am incredibly grateful to Dr Ntandokayise Ndlovu for being both a brother and a mentor. You continue to support me without expecting anything in return.

I make a special mention to my esteemed colleagues whose love, support and camaraderie have been instrumental throughout this journey. Your encouragement and collaborative spirit have been truly inspiring. I express my heartfelt appreciation to Luxolo Tomsana, Tendai Nyakumbi, Joseph Mugauri, Mihlali Hoofman, Hanani Hlomani, Ben Mgcini, Zilungile Mbali, Yibanathi Nyeka, Ludwe Mbeceni, Asekho Phiwani, Samuel Kanhodo, Ayabonga Gantuntu, Xabiso Mpiko, Kyle Janse, Christine Immenga, Vhugala Nthakheni and Nathalie Baumgart.

Thank you to the University of Cape Town for creating a conducive environment necessary for my academic pursuits. I would be remiss without mentioning some individuals who have had a particular influence on my work: Prof Lee-Ann Tong, Nandi Ntengenyana, Katriina Hetemaa, Ghati Nyehita, Omowamiwa Kolawole and Anel Odendaal.

I extend my heartfelt appreciation to all those I have not mentioned by name who, directly or indirectly, contributed to this research.

ABSTRACT

Once again, humanity has welcomed technological advancement, this time around artificial intelligence, with mixed reactions. The creative industry is no exception to this rapidly evolving technology, with generative AI (genAI) deepening its claws in the creative industry. What lies within the fabric of genAI is a primary concern. In order to be trained (taught), genAI ingests enormous amounts of data, which is harvested indiscriminately. This is a cause for concern for those whose work is harvested and utilised without compensation, credit and consent. In addition, when genAI is deployed, the user's input prompts it to create works of their desires, ranging from images to musical lyrics. To that end, it remains to be answered whether the use of works for the purpose of training genAI and the generation of works by genAI trained using copyrighted works amount to copyright infringement.

It is the duty of this dissertation to examine whether the South African copyright regime would deem it an infringement to make use of copyrighted works for training genAI. In addition, this dissertation goes further to examine whether there is a possibility of copyright infringement materialising when a user generates works through genAI. In this dissertation, the training of genAI will be referred to as the "input phase", while the use of genAI by a user will be referred to as the "output phase." In addition, the issue of who is liable when copyright infringement materialises will be analysed.

In the end, the author submits recommendations for South Africa to address copyright infringement liability. Leading jurisdiction in copyright and AI regulations will be infused to enrich the discussion.

INDEX OF ABBREVIATIONS

AI-	Artificial Intelligence
AGI-	Artificial General Intelligence
ANI-	Artificial Narrow Intelligence
ASI-	Artificial Super Intelligence
CC -	Constitutional Court of South Africa
DSRPAI -	Dartmouth Summer Research Project on Artificial Intelligence
EU -	European Union
GenAI -	Generative AI
LAION -	Large-Scale Artificial Intelligence Open Network
LLMs -	Large Language Models
ML -	Machine learning
NN-	Neural Networks
SCA -	Supreme Court of Appeal
US/USA -	United States of America
UK -	United Kingdom
WIPO -	World Intellectual Property Organisation
WTO -	World Trade Organisation

TERMINOLOGY

In this dissertation, unless the context indicates otherwise, the terms are to be understood as defined in this section

AI: refers to the use of computer science methods that enable machines to perform tasks traditionally viewed as requiring human intellect, in other words, the ability of computers to exhibit human-like cognitive abilities. Used in a broad sense.

AI company: refers to the entity or organisation that recruits, employs, and provides the necessary resources for AI programmers to carry out their work, including the development, testing, and deployment of AI models. At times, it is referred to as an AI producer.

AI programmer: refers to a software developer or coder, who writes, tests, debugs, and maintains the code that enables software applications to function. Programmers use and understand programming languages.

AI model: used interchangeably with "AI system" and means a program that functions as defined above under AI.

Author: Refers to Tendai Mikioni, who authored this dissertation.

GenAI: is an abbreviation for generative AI, which refers to a computational technique capable of generating seemingly new, meaningful content, such as text, images, or audio, based on training data.

GenAI-generated: refers to works or outputs produced by genAI models, for example art, music, videos, images.

***In casu**:* a Latin phrase that translates to "in this case" or "in the present case."

Reader: refers to any person reading this dissertation

Time mark: refers to a specific moment within a YouTube video, cited using minutes and seconds, to pinpoint a particular section for reference or discussion.

CHAPTER 1: INTRODUCTION

Introductory Remarks

Throughout history, humanity's progress has been intertwined with its ingenuity to harness various tools. In the present technological era, this is exemplified in the realm of Artificial Intelligence (AI). The rapid advancement of AI has sent shockwaves across multiple industries and transformed the landscape of intellectual property rights. As the author navigates the complexities of this evolving landscape, it becomes imperative to explore how AI's intersection with intellectual property poses both opportunities and challenges, paving the way for the bruising and battering of copyright. In *Payer Components South Africa Ltd v Bovic Gaskins,* the court distinguished between "computer-aided" and "computer-generated" works.[1] On one hand, computer-aided works are those works where a computer is a mere tool, while on the other hand, computer-generated works are those works whose creation is performed by the computer itself with relatively little human input.[2] In understanding this dissertation, the reader should view generative AI (genAI) as creativity by a computer model with relatively little human input.

The advent of genAI has revolutionised the creative landscape, empowering individuals to generate compelling works from the comfort of their homes.[3] A textbook illustration of this phenomenon is evident in the creation of a song titled "Heart on My Sleeve." Though seemingly a collaborative effort between renowned artists Drake and The Weeknd, this captivating track was, in fact, the brainchild of an anonymous creator known as Ghostwriter, who dexterously harnessed genAI to replicate the distinctive voices of these musicians and craft an entirely new musical composition.[4] It is reported that upon its release, this song went viral on TikTok and amassed more than 230,000 plays on YouTube, along with over 625,000 plays on Spotify.[5]

Digging into this case of "Heart on My Sleeve," gives valuable insights into the challenges posed by genAI creativity and the pressing need for a thoughtful and comprehensive approach to

---

[1] *Payer Components South Africa Ltd v Bovic Gaskins & Others* 1995 (2) All SA 600 (A) at 603.
[2] *Ibid.*
[3] Forbes "The Best Free AI Music Generators Of 2023" available at *https://www.forbes.com/sites/ariannajohnson/article/the-best-free-ai-music-generators/?sh=1fa5215d4ffb*, accessed on 3 July 2023.
[4] Sky News 'AI Drake and The Weeknd: Song called Heart On My Sleeve - made with cloned voices - removed from streaming services' available at *https://news.sky.com/story/ai-drake-and-the-weeknd-song-called-heart-on-my-sleeve-made-with-cloned-voices-removed-from-streaming-services-12859951*, accessed on 3 July 2023.
[5] *Ibid.*

safeguarding artistic expression and intellectual property rights in the digital age. The reality of genAI in the creativity industry raises several questions, for example: (i) whether the use of copyrighted material to train genAI amounts to copyright infringement; (ii) whether the works generated by genAI infringe existing copyright; (iii) who is liable for copyright infringement - the genAI developer, the genAI user, or the genAI model itself; and (iv) what the implications are for holding each party liable.

This dissertation examines the potential copyright infringements associated with genAI in South Africa. It specifically investigates two key areas: whether the training of genAI models prior to deployment constitutes a violation of copyright and whether the subsequent generation of works by genAI in response to user prompts also amounts to copyright infringement. Through analysing statutes, existing jurisprudence and scholarly commentaries, this study aims to shed light on critical questions in the wake of genAI. Consequently, recommendations are made to address copyright infringement associated with genAI.

Justification and Impact

The Copyright Act[6] provides for infringement of copyright and attaches consequences. The Copyright Act provides that copyright is infringed when a person who is not the owner of the copyright and without the licence of the owner does or causes any other person to do any act which the owner has the exclusive rights to do or to authorise.[7] The other form of infringement is when a person without a licence from the copyright owner deals in infringing articles.[8]

The mushrooming of works by genAI, which in some cases infringes copyright, puts the law to the test. The question of copyright infringement liability surfaces, specifically on whose shoulders the liability lies. Should the infringement liability be imputed on the genAI programmer? Should the users of the genAI model bear the liability for infringement? Possibly, is it appropriate for the genAI model itself to assume responsibility for the liability for infringement? As the lines blur as to whose shoulders should copyright infringement liability fall, it becomes paramount to explore this subject.

[6] s23(1) of the Copyright Act 98 of 1978.
[7] s23(1) of the Copyright Act.
[8] s23(2)(a)-(d) of the Copyright Act.

The arguments made in this dissertation are premised on the need to address the emerging legal challenges to ensure that copyrights are protected in the digital age. In addition, the current copyright laws were drafted without advanced technologies such as AI in mind. The Copyright Act was assented to on the 20th of June 1978 and came into force on the 1st of January 1979. Despite several amendments, the Copyright Act craves harmonisation with the current advancement in the digital age. This research embarks on an academic journey of finding answers as to whether, despite their archaic nature, the current laws can assist in addressing the question of copyright infringement and liability.

In addition, this research will provide recommendations to guide lawmakers and policymakers in developing comprehensive and future-proof copyright laws that account for genAI creations.

Research Question

This dissertation seeks to answer the main question: "Whether the use of copyrighted material for training genAI and the subsequent generation of works by genAI constitute a violation of copyright law in South Africa, and if so, who should be held liable— the human developer, the users of the genAI, or the AI model itself—and what are the implications of holding each party liable for such infringement?"

This research question is further broken down into the following sub-questions:

1. Whether the use of copyrighted material during the genAI-training phase amounts to copyright infringement?
2. Who is responsible for copyright infringement during the genAI training phase?
3. Whether the genAI-generated works amount to copyright infringement?
4. Who is liable for copyright infringement where genAI-generated works infringe copyright?
5. What are the implications of holding each party responsible?

Research Methodology

This research will employ doctrinal legal research methodology. Doctrinal research examines the law in a particular area whereby the researcher collects and analyses a body of case law and legislation and may include secondary sources such as journal articles or written commentaries on

case law and legislation.[9] The doctrinal (or classical) legal research method has been labelled "the gold standard of legal research."[10] Doctrinal research methodology involves establishing legal rules and principles and applying them to a specific factual context.[11] In addition, the doctrinal legal research methodology inherently includes describing the state of the law and conducting legal analysis, which may also involve comparative study, and when necessary, it provides recommendations for potential law reform.[12] On this basis, the author adopts doctrinal legal research methodology as a seamless-fit methodology for this research.

Thesis Structure

The structure of the remainder of this thesis is set out as follows:

Chapter 2: *Literature Review*

This chapter focuses on a literature review of the subject matter, providing an analytical discussion and overview of existing academic research on genAI and copyright infringement liability. Given the scarcity of literature on the topic, the author consults journals, commentaries, and podcasts from credible institutions.

Chapter 3: *An Overview of AI and Copyright Infringement*

This chapter broadly explains AI technology's fundamental functionalities and copyright in South Africa. Regarding AI technology, the focus is on its definition, historical development, various types, current trends, and future prospects. Concerning copyright, the chapter defines copyright, underscores its significance, and elucidates the copyright protection offered in South Africa. The subsequent part of this chapter examines copyright infringement within the South African context.

Chapter 4: *Untangling Copyright Infringement Liability in genAI Training Data – Input Phase*

Chapter four focuses on copyright infringement at the training stage of AI models– the input phase. In this chapter, the author examines whether the use of copyrighted material for the purpose of training genAI amounts to copyright infringement. In addition, the chapter will examine the

[9] Mike McCnville & Wing Hong Chui *Research Methods for Law* 2 ed (2017) 21.
[10] Mkhululi Nyathi 'Re-Asserting the Doctrinal Legal Research Methodology in the South African Academy: Navigating the Maze' (2023) 140 *SALJ* 367.
[11] *Ibid.*
[12] *Ibid* at 369.

assignment of copyright infringement liability for the use of copyrighted material to train genAI, specifically the suitability of the following candidates: genAI developers, genAI users, and the genAI model itself. The chapter further highlights the implications of holding each party liable for infringement. It will synthesise the South African approach and examine how other jurisdictions would handle the assignment of liability.

Chapter 5: *Untangling Copyright Infringement Liability in genAI-Generated Works – Output Stage*

In chapter five, the author's attention is directed toward copyright infringement involving works generated by genAI - the output stage. The chapter addresses the assignment of copyright infringement liability when genAI produces a work. The candidates that will be explored are genAI developers, genAI users, and even the AI model itself, as well as the implications of holding each party liable. Most importantly, the author examines other jurisdictions that would handle copyright infringement associated with genAI.

Chapter 6: *Conclusion*

The final chapter summarises the key findings from the research conducted throughout the study. It provides a concise overview of the main findings and makes recommendations. Additionally, it concludes with the author's final thoughts and insights on genAI and copyright infringement liability.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

Owing to the deficiency of literature regarding copyright infringement by genAI, the author relies heavily on available journals, conference papers, an amalgamation of blog posts, commentaries and a legal podcast. The novelty of the topic and other factors play a massive role in the unavailability of dense literature in this area. Hopefully, this research will alleviate such scarcity and advance knowledge in copyright infringement liability through genAI-generated works.

2.2. Literature Appraisal

Guadamuz voices his view in a beautifully crafted opinion-piece[1] where he explores the copyright liability and exceptions that exist in AI input and output. He begins by unpacking AI, acknowledging the multiplicity of definitions advanced, but a common feature stands out, which is the requirement for data as an input of whichever model is used. He further explains that the term "big data" is often used in conjunction with AI. Big data simply refers to "a confluence of volume, velocity, and variety of information, and when applied to artificial intelligence, it means that the faster, larger, and varied the amount of data, the better the AI model will be."[2] This comes in handy in understanding the thematic topics and data, which is critical to understanding this research.

Guadamuz's paper focuses on two particular copyright infringement questions in the form of inputs and outputs.[3] On one hand, the input part of the article focuses on whether the act of accessing, reading, preparing, analysing, and mining data is an act of copyright infringement, and should this be affirmative, and what remedies can be employed. He argues that a question of copyright infringement will depend on the dataset used; for example, some data is comprised of data collected lawfully in the public domain or unprotected data or with permission of the author. However, Guadamuz did not dismiss the possibility of infringement in some cases.

On the other hand, the outputs part of the article explores whether the copyright holder of the works used to train (teach) the AI can sue the maker for copyright infringement from the

[1] Andrés Guadamuz 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs' (2023) Available at SSRN: http://dx.doi.org/10.2139/ssrn.4371204.
[2] *Ibid* 3.
[3] *Ibid.*

resulting works. Despite being placed in a setting that is foreign to Africa, these two pertinent are critical to this research. The paper mainly focused on whether infringement occurs in the input and output phases and passed remedies for such, which leaves room for exploring the imputing liability and harmonisation of thoughts shared in the article with a South African setting.

Vesala and Ballardini penned a think-piece[4] on copyright infringement triggered by the use of third-party data in the computational processes involved in training neural networks (NN) or in using it once it is trained. They aptly argue that in order to develop AI applications using neural networks trained with data, it is necessary to have access to and authorisation to use suitable training data.[5] Vesala and Ballardini posit that training neural networks with third-party data could infringe upon copyright and other intellectual property rights. This dissertation focuses on the former. The premise upon which they hinge their argument is that the act of using copyright-protected works to develop and train neural networks may involve actions that are covered by the exclusive rights of copyright holders, in most cases, the right of reproduction.

Vesala and Ballardini grounded their premise on the following: first, it is possibly necessary to make digital copies of the data when collecting and preparing such data for the purpose of training neural networks. Second, the computational processes that take place within neural networks might result in the reproduction of identical or substantially similar copyright-protected aspects of training data. Vesala and Ballardini sum up by advancing that unless an exception or limitation can be employed, copyright infringement may occur should the training of an NN be conducted without the consent of copyright holders. It is apparent from the aforementioned commentary that they invested the writing in the infringement phase, which is equally important to the ongoing discussion on copyright infringement in AI-generated creations. However, the lack of particular attention as to whose shoulders should the liability of copyright infringement be imputed is a testimonial to the need for completion of the ongoing research paper. This should not be taken as a flaw of the paper authored by Vesala and Ballardini, but rather, their work must be prized for its role as water from which withered thoughts absorb moisture.

[4] J Vesala & R M Ballardini 'AI and IPR Infringement: A Case Study on Training and Using Neural Networks' in R M Ballardini *et al* (eds) *Regulating Industrial Internet through IPR, Data Protection and Competition Law* (2019) 99-114.
[5] *Ibid.*

Bonadio and McDonagh crafted an insightful article tackling a wide ride range of topical intellectual property rights issues in works by genAI. [6] Amidst all the hot topical issues raised, they dealt with the risk of copyright infringement and how fair use[7] as an exemption can be incorporated to assist in escaping infringement liability. They submitted their noble viewpoint that "expressive uses of copyright works by AI systems should not be exempt from copyright infringement via the fair use defence, especially if the final output is going to be exploited commercially and not, for example, used for research purposes."[8] It is evident that their argument is hinged on the premise that where fair use remedy is employed, it deprives the authors and copyright holders of the opportunity to exploit the market.[9]

Furthermore, this paper adeptly and incisively raises a critical observation by advancing the assertion that if the programmers, owners or users of genAI models want to claim exclusive rights to the final outputs generated by the genAI model, they must bear responsibility for those works, not only in relation to copyright infringement but also for any other form of liability.[10] The line of reasoning adopted by the paper is based on the logic that where one claims to be an author/owner of a work produced by the generative process carried out by the genAI model, they should equally be held liable for when the computer infringes the rights of others. In the same vein, the paper questions the reasoning and logic behind exempting the owner or user of genAI, which consumes huge quantities of copyrighted works without asking for permission to access such data, while a human creator replicating the same act at a much lower scale cannot take advantage of the exemption. Thus, the selective application of the exemption and discrimination of human authors compared to genAI would be evident. In sum, Bonadio and McDonagh propose that those who claim copyright ownership in the genAI final outputs must be liable for copyright infringement. The perspective shared by Bonadio and McDonagh will serve as a well of knowledge the author can tap into in order to iron out the issue of copyright infringement by genAI in South Africa.

---

[6] Enrico Bonadio & Luke McDonagh 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity' (2020) 2 *Intellectual Property Quarterly* 112-137.
[7] *South Africa uses fair dealing per s12(1)(a)-(c) of the Copyright Act 98 of 1978.
[8] Supra note 6 at 9.
[9] *Ibid* 9.
[10] *Ibid* 9.

Naqvi wrote an in-depth analysis[11] of copyright infringement by genAI from the USA standpoint, precisely and effectively addressing the core issues and how the current law would grapple with these issues. He contends that since Al is ineligible to be a copyright holder, it cannot be sued for damages, and there must be an alternative party that may hold the copyright and be liable for the infringement at the same time.[12] In pursuit of who can be liable, he captivatingly applied the law governing agency, intertwining the agent-principal relationship in the vast part of the paper's substance. He infuses a principle that liability can be imputed to the principal where a third party is harmed by an agent who is acting within the scope of their employment, and the agent's conduct is tortious, or such a principal conducted himself negligently when selecting, supervising, or controlling the agent who harmed the third party.[13] Furthermore Al can be treated as an agent of a principal because the principal exerts control over and supplies all aspects of Al to produce work. However, a stumbling block remains: Al does not have the legal capacity to assent to be an agent for a principal, but Naqvi submits that "law must treat the Al as an agent to avoid any verbiage issue." It is worth noting that the programmers in the case of genAI may not be the owners of the genAI because they were hired by a company which, in the end, owns the genAI model, argues Naqvi.[14]

In addition, Naqvi posits that, under a principal-agent relationship, the principal should be held liable for copyright infringement as dictated by the fact that the AI model is an agent synthesised with reliance on the *respondeat superior*[15] principles, which provides that a principal is liable for an agent's actions if the agent's action occurs within the scope of their agency. For the court to hold the principal liable, the following factors will be employed to enforce the principle: (1) control and (2) direct financial benefit.[16] Naqvi observes that with AI, the principal controls how and what the Al produces, and s/he is the one who benefits from the work created by the Al. Therefore, holding the principal liable for copyright infringement from the agent-principal law lens is fair.

---

[11] Zack Naqvi 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 *Marquette Intellectual Property Law Review* 15-52.
[12] *Ibid* 25.
[13] *Ibid* 26.
[14] *Ibid* 31.
[15] A Latin phrase that literally means "let the master answer." This is akin to vicarious liability in South Africa.
[16] *Supra* note 11 at 31.

Furthermore, the other angle that can have the producer of AI (above analysed as principal) held liable for copyright infringement hinged on secondary liability. Naqvi argues that secondary liability applies when "one who, with knowledge of the infringing activity, induces, or materially contributes to the infringing conduct of another." He further points out that a special case for secondary liability revolves around contributory infringement that can arise when a defendant sells goods used by a direct infringer. He argues that secondary liability can be noted because of the enticement of the consumer by the producer through "marketing, advertising, subtle hints, and/or examples of how the consumer's ability to infringe on copyright work is easier."[17] However, Naqvi laid bare that in some cases, where a company produces and sells AI to a user who later uses that Al in a manner that infringes the copyrights, the company will not be liable.[18] The premise for this reasoning is that the Al in question has substantial lawful uses, and courts are willing to limit liability to cases of more acute fault, even though some of the Al will be misused.[19] In such cases, the AI user or consumer is liable for copyright infringement.

Therefore, Naqvi authored a paper, which is a treasure trove of information that has advanced the reasoning that liability cannot be imputed on the AI model but rather on the AI programmer through secondary liability or as a principal in an agent-principal relationship. In addition, where the user or consumer misuses an AI model, liability must be on their shoulders. The paper eloquently and rigorously addresses the central topical issues from the USA perspective. It will serve as an enriching ingredient to the finalisation of the issue of copyright infringement liability in South Africa, as well as understanding the implication of holding the parties involved liable. The knowledge harvested from the paper by Naqvi will be a guiding insight for an exploration of the South African position.

Bonadio, Dinev and McDonagh contributed a perceptive paper[20] exploring the possibility of allocating copyright infringement liability to "end user; the seller or developer/programmer (in a broad sense); and the AI system itself." Dealing with the possibility of imputing liability on the AI model itself, Bonadio, Dinev and McDonagh present the argument that most institutions and

---

[17] *Ibid* 38.
[18] *Ibid* 35.
[19] *Ibid.*
[20] Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022) 245 – 257.

scholars have explored granting AI legal personality. It is argued that the European Union has considered equating intelligent machines with corporations which already have "legal personhood," which could pave the way for liability to be imputed on the AI in the long run.[21] The scholars noted with utmost care reports that Saudi Arabia granted Sophia (a humanoid robot) citizenship and highlighted that these reports stray far from the shores of accuracy, making them "nothing more than a publicity stunt." Based on that, no major jurisdiction, the USA, EU or UK, has recognised the AI or robots as entities capable of acquiring legal status.[22] They tie their argument on AI and liability by asserting that their lack of legal personality means they cannot be held accountable at present.[23]

Bonadio, Dinev and McDonagh went on to explore the possibility of holding a user or AI developer liable. Regarding the AI developer, in some cases, the AI company merely provides the AI as a tool to the user who has complete control over what is fed into the AI model.[24] In this instance, the developers effectively provide the necessary code, which is then automatically customised in tandem with the needs of the user. In such a case, the developer does not encourage infringement, lacks control of any control over the process and does not monitor or deal with the AI-generated outputs. Bonadio, Dinev and McDonagh express that the AI developer may be absolved in such a case, and the user may be held liable for primary infringement.[25] However, as reflected in the *locus classicus*: *Metro-Goldwyn-Mayer Studios Inc v Grokster Ltd*,[26] it is possible to hold the developer (company) liable for copyright infringement through the application of the secondary liability where the seller encourages users to infringe by using their AI products. To put it differently, the developer is absolved from liability in instances where the AI has substantial lawful use, and the AI was appropriately marketed and advertised.[27]

They went a step further to table thoughts on instances where the AI model has the capability to alter its own programming, extinguishing the possibility of neither developers nor

[21] *Ibid* 11.
[22] *Ibid.*
[23] *Ibid.*
[24] *Ibid* 12.
[25] *Ibid.*
[26] *Metro-Goldwyn-Mayer Studios Inc v Grokster Ltd,* 545 U.S. 913 (2005).
[27] Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022) 12.

users being able to appreciate the risk of infringement.[28] Bonadio, Dinev and McDonagh postulate that, on the one hand, holding the users liable is unfair because users, as consumers, "are often legally unsophisticated individuals and adopting a punitive approach towards users may also discourage the use of (otherwise helpful) AI."[29] On the other hand, developers or programmers are better positioned to appreciate the copyright infringement liability and are most likely to acquire economic value from AI through selling and licensing it.

However, they noted that expecting developers to anticipate infringement risk will not always be realistic given the nature and aim of processes like machine learning; some degree of case-by-case evaluation is likely to be necessary and desirable.[30] In light of the failure to find a candidate to be held accountable for copyright infringement involving AI, they predict that this problem might indirectly encourage the use of the genAI model for infringement purposes by individuals.[31] Wrapping up their discussion, they lament: "Why should the use of AI be given special treatment while holding human actors liable in the same or similar set of circumstances?"[32] This paper is a goldmine of expertise on copyright infringement by AI models and will be a practical asset to the completion of this research in South Africa.

Gervais enriches the current discussion by formulating a paper[33] which reviews arguments for and against the protection by copyright of literary and artistic productions made by AI machines, which brings to the spotlight the issue of liability as a result of AI. He opines that the outputs of AI's deep learning processes must not be recognised (copyrighted) unless the genAI model as the purported "author" can accept full responsibility for "its" creation.[34] He adopts a mantra that "no rights without responsibilities" is quelling arguments that advocate for the protection of AI-generated works.[35] The fact that the paper was an exploration of the copyrightability of AI-generated works must not be a deterrence to the use of this paper[36] because

---

[28] *Ibid* 13.
[29] *Ibid.*
[30] *Ibid.*
[31] Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022) 11.
[32] Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022) 14.
[33] Daniel J Gervais 'The Machine as Author' (2019) 105 *Iowa Law Review* 2053-2106.
[34] *Ibid* 2087.
[35] *Ibid*.
[36] Daniel J Gervais 'The Machine as Author' (2019) 105 *Iowa Law Review* 2053-2106.

it is a reservoir of learning that in order for genAI works to be protected, they (AI or who acquires rights) must assume responsibility in cases where there is copyright infringement.

Watson supplements the ongoing conversation by producing an article[37] which deals with liability for patent infringement in cases involving multi-parties such as a producer, seller and user of AI. He suggests that parties must resort to contracts or "interim contractual solutions" to escape liability since it is unclear how courts will allocate infringement liability when an artificial intelligence model used by multiple parties infringes a patent.[38] Watson stresses that a contractual solution is needed by the parties as it will be a predictable solution to liability, should infringement occur.[39] A notable suggestion is that a party, developer and not seller, must satisfy three requirements in order to shield itself against liability issues arising from direct infringement by genAI model users. These requirements are: updating software, notifying the selling party of any new inventions, and getting feedback from a patent attorney, and refraining from encouraging infringement.[40] While the paper indeed cantered on patent infringement rather than copyright, this must not dent the significance of this paper as an authority because the same proposed contractual solution may be transplanted into the copyright regime; in any event, copyright does not operate in a vacuum but together with other intellectual property rights.

Saputr and Budianto enter the fray of the ongoing discussion with the authorship of a paper[41] exploring the infringement liability involving AI from an Indonesian perspective. Regarding imputing liability on the AI model, they argue that AI is neither a legal subject nor given by any court legal standing.[42] In light of this, AI lacks the capacity to defend its own rights. Therefore, in order to carry out legal actions such as making licencing agreements or filing lawsuits for copyright infringement, a legal entity or a non-AI legal entity is required.[43] In sum, humans are still required to handle the rights and obligations associated with copyright because AI is unable to act independently. This opinion suggests that AI models cannot be held liable as

---

[37] Bridget Watson 'A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence' (2017) 58 *IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property* 65-93.
[38] *Ibid* 83.
[39] *Ibid.*
[40] *Ibid.*
[41] Dimas Saputr & Azis Budianto 'Legal Analysis of Artificial Intelligence Responsibility in Copyright Infringement in the Digital Era' (2023) Proceedings of the 3rd *International Conference on Law, Social Science, Economics, and Education*.
[42] *Ibid* 2.
[43] *Ibid.*

human candidates are required to fulfil their obligations. They focused, in particular, on the allocation of copyright to either the programmer, the company, the user, or the AI model itself.

Kumar and Kumar expand upon the ongoing exchange by presenting an article[44] on the training of AI and copyright infringement. The paper densely focuses on the training, the exceptions and suggestions; however, it raises the possibility of using the fair use doctrine in relation to the use of data when training AI.[45] It has been argued that failure to allow the exemption of data in AI training can result in biased Al algorithms. This is based on the observation that the current and contemporary works are not generally available in the public domain.[46] In addition, a striking argument was advanced that "the use of archaic material in training often leads to the perpetuation of old-fashioned beliefs and attitudes."[47] Kumar and Kumar noted that subjecting Al's training to stringent liability for copyright infringement hampers the public's right to information, which captures the right to conduct and access research.[48] However, Kumar and Kumar did concede that the use of copyrighted works by an Al developer to create new expressive works cannot be deemed fair, especially if the output is exploited financially because the right holder's market will unavoidably be adversely harmed unfairly.[49] Regardless of focusing on the "fair use" doctrine, a doctrine not aligned with the "fair dealing" doctrine applied in South Africa, the paper makes outstanding observations that will be useful to the discussion herein.

Craig supplements the discussion in an iNtaka podcast,[50] where she notes that genAI is trained on vast quantities of copyright-protected material. Craig holds that the use of copyrighted material for AI training purposes should not be regarded as infringing.[51] Her argument is premised on the fact that the use of copyrighted material in training AI is not a copyright use or use of work as work because it is not used as original expression.[52] She further adds that these

---

[44] Ravindra Kumar & Pankaj Kumar 'Training ai and copyright infringement: where does the law stand?' (2022) 2 *Indian Journal of Integrated Research in Law* 1-14.
[45] *Ibid* 5-9.
[46] *Ibid* 8.
[47] *Ibid.*
[48] *Ibid.*
[49] *Ibid.*
[50] Carys Craig "The AI Copyright Trap" *iNtaka Centre for Law & Technology* note 20:00 (time mark) available at *https://www.youtube.com/watch?v=vX-c-YOKs_8,* accessed on 23 May 2024.
[51] *Ibid*, 21:59 (time mark).
[52] *Ibid,* note 22:22 (time mark).

works are not being used for their meaning or understood by the AI model in that the data is simply being pulled out and turned into statistics and tokens.[53] Therefore, the use is beyond the scope of the legitimate interest of any copyright owner. Craig cautions that if it were to be that training could not be done without permission of every copyright owner of every piece of work in a data set; copyright basically could be a thing that changes the trajectory of AI development or obstructs it and significantly limits the capacity of AI development and deployment.[54]

In relation to works generated by genAI, Craig argues that when a person who, by design, intends to make money from infringing activity at the output stage, there is potential for infringement liability, but it depends on the business model and the defendant can demonstrate that they have done what they can to avoid the model from producing infringing outputs.[55] On the question of finding genAI users liable, Craig argues that the inquiry should be whether the prompts used by the user were to produce an infringing copy.[56]

2.3. Conclusion

In conclusion, this chapter has provided insight on copyright infringement and liability associated with genAI, which informs the arguments made in this dissertation. Flowing from this insight, the dissertation proceeds to examine deeply the ongoing discussion in a South African context.

[53] *Ibid,* note 22:39 (time mark).
[54] *Ibid*, note 20:12 (time mark).
[55] *Ibid*, note 35:55 (time mark).
[56] *Ibid*, note 36:24 (time mark).

CHAPTER 3: AN OVERVIEW OF AI AND COPYRIGHT INFRINGEMENT

3.1. Introduction

This chapter provides an in-depth overview of AI technology as a broader concept and copyright. First, in addressing AI technology, the author will provide the definition, historical development, types of AI, current trends, and future prospects. Regarding copyright, the chapter defines copyright, underscores its significance, and explains the copyright protection offered in South Africa. The subsequent part of this chapter examines copyright infringement within the South African context in its crude sense.

3.2. An overview of AI technology

3.2.1. Definition of AI

To begin, there is no widely agreed-upon definition of AI.[1] Despite this lack of consensus on the definition of AI, researchers in the field have submitted substantially similar definitions. AI is a broad term that encompasses a wide range of technological solutions from state-of-the-art deep learning models to human-coded rules thus, it can be challenging to determine which tools qualify as AI and which do not.[2] AI involves the use of computer science methods that enable machines to perform tasks traditionally viewed as requiring human intellect, in other words, the ability of computers to exhibit human-like cognitive abilities.[3]

GenAI refers to computational techniques capable of generating seemingly new, meaningful content, such as text, images, or audio, based on training data.[4] GenAI produces novel and realistic content such as texts, images, or programming code for various domains based on basic user prompts.[5] Furthermore, while acknowledging the absence of a universally agreed-upon definition of AI, it is contended that a practical definition is that AI is any machine capable of executing tasks that would typically require human intelligence.[6] AI is not a single piece of hardware or software, but rather, a constellation of technologies that allows a computer system the

---

[1] European Parliamentary Research Service *Civil liability regime for artificial intelligence* (2020) 3.
[2] Dariusz Szostek & Mariusz Załucki (eds) *Legal Tech* (2021) 81.
[3] *Ibid* at 81.
[4] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch & Patrick Zschech 'Generative AI' (2023) *Business & Information Systems Engineering* 1.
[5] Leonardo Banh & Gero Strobel 'Generative artifcial intelligence' (2023) 63 *Electronic Markets* 1.
[6] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 84.

ability to solve problems and perform tasks that would normally need human intelligence.[7] Therefore, while there is no single-agreed definition of AI, the author observes that AI mimics human cognitive abilities and performs tasks that traditionally require human abilities.

3.2.2. Development of AI

Exploring AI requires tracking its inception to the current *status quo*. The technological progress of AI models is not yet fully understood or known technologically.[8] The genesis of AI is difficult to pinpoint, but the roots of AI can probably be traced back to the 1940s, specifically 1942, when a short story, Runaround, was authored and published by the American Science Fiction writer Isaac Asimov.[9] The storyline of Runaround centres on a robot created by engineers Gregory Powell and Mike Donovan.[10] The narrative revolves around the Three Laws of Robotics: (1) a robot is prohibited from causing harm to a human or, through inaction, allowing harm to befall a human; (2) a robot is obliged to follow human orders unless such commands contradict the First Law; and (3) a robot must protect its own existence as long as it does not conflict with the First or Second Laws. Asimov's work is credited for serving as a source of inspiration for numerous generations of scientists in the field of robotics, AI, and computer science, such as Marvin Minsky.[11]

Around the same period, the English mathematician Alan Turing engaged in considerably less fictional pursuits and dedicated his efforts to creating a code-breaking machine named The Bombe for the British government, aiming to decrypt the Enigma code[12] employed by the German army during the Second World War.[13] The remarkable capability of The Bombe to decipher the Enigma code, a task deemed impossible for even the most skilled human mathematicians, led Turing to wonder about the intelligence of such machines. In 1950, Alan Turing penned an article

---

[7] *Ibid.*
[8] European Parliamentary Research Service *Civil liability regime for artificial intelligence* (2020) 3.
[9] Michael Haenlein & Andreas Kaplan 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) *California Management Review* 61(4) at 6.
[10] *Ibid.*
[11] *Ibid.*
[12] During World War II, the Germans used the Enigma, a cipher machine, to develop nearly unbreakable codes for sending top secret messages. See Central Intelligence Agency "Enigma Machine" available at *https://www.cia.gov/legacy/museum/artifact/enigma-machine/#:~:text=During%20World%20War%20II%2C%20the,able%20to%20crack%20its%20code,* accessed 10 February 2024.
[13] Michael Haenlein & Andreas Kaplan 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) *California Management Review* 61(4) at 7.

titled "Computing Machinery and Intelligence."[14] Turing posed the question of whether machines could think and learn from experience in a manner similar to a child.[15] One of the critical steps was Turing's introduction of "The Turing Test," an experiment to assess a computer's ability to think and behave like a human.[16] The Turing Test continues to be regarded as a contemporary benchmark for assessing the intelligence of artificial systems: if a human, while interacting with both another human and a machine, is unable to distinguish the machine from the human, the machine is considered intelligent according to this test.[17]

Alan Turing's work could not take off immediately, given the obstacles that existed at that time. First, computers needed a fundamental change due to the fact that before 1949, computers lacked a key element for intelligence: they could not store commands; they could only execute commands without the capacity to retain them.[18] In other words, computers could be told what to do, but they were unable to recall or remember the actions they did. Second, computing was extremely expensive, with the monthly cost of leasing a computer soaring to $ 200,000.00 (equivalent to today's South African ZAR 3 765,000.00) in the early 1950s.[19] Only prestigious universities and large technology firms had the financial means to explore the computing realm. Therefore, like any development in humanity's history, the development of AI was neither a linear one nor an exponential progression.

Professor John McCarthy is widely credited with coining the term "Artificial Intelligence."[20] It is argued that the official coining of the term AI occurred in 1956 during the Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI), a roughly eight-week-long initiative hosted by Marvin Minsky and John McCarthy, a computer scientist at Stanford, at

---

[14] A M Turing 'Computing Machinery and Intelligence,' (1950) 49 *Mind* 433-460, available at *https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf,* accessed *10* February 2024.
[15] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 7.
[16] *Ibid.*
[17] Michael Haenlein & Andreas Kaplan 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) *California Management Review* 61(4) at 7.
[18] Rockwell Anyoha 'The History of Artificial Intelligence' *Harvard Special Edition on Artificial Intelligence* 28 August 2017 available at *https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/* accessed 10 February 2024.
[19] *Ibid.*
[20] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 7.

Dartmouth College in New Hampshire.[21] This workshop marked the birth of AI, was funded by the Rockefeller Foundation, and brought together individuals later recognised as the founding fathers of AI. Notable participants included computer scientist Nathaniel Rochester, the designer of the IBM 701, the first commercial scientific computer, and mathematician Claude Shannon, the founder of information theory.[22] DSRPAI aimed to unite researchers from diverse fields to establish a new research area focused on constructing machines capable of simulating human intelligence. The importance of this event cannot be understated, as it served as a catalyst for the subsequent two decades of AI research.[23] Thus, DSRPAI planted the seed for AI research to germinate.

Following the Dartmouth Conference, there was a period of almost two decades marked by substantial achievements in the field of AI. An early example is the renowned ELIZA computer program, developed between 1964 and 1966 by Joseph Weizenbaum at MIT. ELIZA, a natural language processing tool, could simulate conversations with humans and was one of the pioneering programs to attempt to pass the Turing Test.[24] The other cited success story of the early days of AI was the General Problem Solver program that was developed by Nobel Prize winner Herbert Simon and RAND Corporation scientists Cliff Shaw and Allen Newell, which was able to automatically solve certain kinds of simple problems, such as the Towers of Hanoi.[25] This milestone, coupled with inspiring stories, resulted in substantial funding being directed towards AI research, leading to more and more projects.

Despite these fascinating and inspiring stories, AI research caught another cold; for example, in 1973, the US Congress expressed strong criticism regarding the substantial expenditure on AI research.[26] In the same year, AI sustained criticism from the British mathematician James Lighthill, who released a report commissioned by the British Science Research Council. In the report, he questioned the optimistic outlook presented by AI researchers. He asserted that machines would, at best, achieve the proficiency of an "experienced amateur" in

[21] Michael Haenlein & Andreas Kaplan 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) *California Management Review* 61(4) at 7.
[22] *Ibid.*
[23] *Supra* note 18.
[24] *Supra* note 21 at 3.
[25] *Supra* note 21 at 3.
[26] *Supra* note 21 at 3.

games like chess and that common-sense reasoning would perpetually remain beyond their capabilities.[27] In response to the scathing paper authored by James Lighthill, the British government withdrew support for AI research in all universities except three, which are Edinburgh, Sussex, and Essex. Subsequently, the United States government followed suit, marking the commencement of what has been dubbed "AI Winter."[28] AI Research would experience lows and dwindling funding in the ensuing years.

Regardless of the setbacks experienced during the AI winter, over the last two decades, AI has emerged as a transformative technology in the twenty-first century.[29] Experts attribute this transformation to several factors working synergistically, including the development of complex algorithms, improvements in computational speed, the creation of new sensors, an unprecedented surge in data availability, and the rise of cloud computing and machine learning.[30] It is crucial to bring artificial neural networks to the fore in the development of AI. Artificial neural networks and deep learning serve as the foundation for many applications falling under the AI label. They underpin image recognition algorithms employed by platforms like Facebook and speech recognition algorithms that power smart speakers and contribute to the functionality of self-driving cars. Artificial neural networks experienced a resurgence through Deep Learning in 2015 when AlphaGo, a program developed by Google, was able to beat the world champion in the board game Go.[31] However, as early as the 1940s, the Canadian psychologist Donald Hebb developed a theory of learning known as Hebbian Learning that replicates the process of neurons in the human brain.[32]

In the captivating narrative of advancement, the author finds himself a privileged witness to the relentless march of AI. This extraordinary phenomenon, burgeoning and evolving with each passing moment, compels the author to acknowledge that the history of AI is not a static tale penned in the past but an ever-unfolding saga being written anew every single day. The awe-inspiring journey of AI is not confined to the pages of a bygone era; rather, it is an ongoing narrative, captivating as it advances into uncharted territories of technological prowess. In this

---

[27] *Supra* note 21 at 3.
[28] AI Winters refer to times when funding, public interest and research in the artificial intelligence space decline. For a discussion of this concept and related matters, see: Ellen Glover 'What Is AI Winter?' available at *https://builtin.com/artificial-intelligence/ai-winter* accessed accessed 10 February 2024.
[29] *Supra* note 20 at 7.
[30] *Supra* note 20 at 7.
[31] *Supra* note 21 at 4.
[32] *Supra* note 21 at 4.

grand unfolding, the undeniable truth remains – much more is yet to be seen, and the unwritten future of AI holds promises and possibilities beyond the current comprehension.

3.2.3. Synergistic Factors Shaping the Advancement of AI

In this section, the author will engage the key synergistic factors that have shaped and continue to play a crucial role in advancing AI and making it a transformative technology. An understanding of these factors is crucial for comprehending the landscape of AI. These are briefly discussed below.

To begin, the first key factor is complex algorithms. Complex algorithms are mathematical equations embedded in software code that find, sort, and look for meaning in data for example an algorithm can be as simple as the command "Insert."[33] To paint a complete picture of the factors steering AI advancements, it is imperative to underscore the profound impact of computational speed. Computational speed is the speed at which a computer can process data.[34] The scientific arrangement behind computational speed is rooted in silicon transistors in computer chips that are responsible for powering computers and communicating through electric pulses represented as zeros and ones.[35] A positive energy pulse corresponds to a one, while a zero denotes the absence of a pulse. The arrangement of these zeros and ones forms computer code. The ongoing miniaturisation of transistor circuitry enables processing larger volumes of data within progressively smaller spaces, leading to enhanced processing speed.[36] Thirdly, sensors are worth a discussion. Sensor technology is found in driverless cars, cell phones, and home devices. In addition, personal assistants like Siri,[37] Watson,[38] and Alexa[39] all use sensors to collect data.[40]

---

[33] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 7.

[34] IT Law Wiki 'Computational speed' available at *https://itlaw.fandom.com/wiki/Computational_speed#:~:text=Computational%20speed%20is%20the%20speed,Millions%20of%20Instructions%20Per%20Second* accessed 12 February 2024.

[35] *Ibid.*

[36] *Ibid.*

[37] Apple 'Siri' available at *https://www.apple.com/siri/* accessed on 12 February 2024.

[38] IBM "IBM Watson Assistant" available at https://www.ibm.com/docs/en/cloud-private/3.1.0?topic=services-watson-assistant accessed 12 February 2024.

[39] Amazon 'Alexa features' available at *https://www.amazon.com/b?ie=UTF8&node=21576558011* accessed 12 February 2024

[40] *Supra* note 33 at 8.

Furthermore, delving into the dynamic factors of technological advancement demands a recognition of the indispensable influence wielded by data. It is argued that data drives the AI revolution, and the more data one has, the easier it is to train a computer system to perform a task or solve a problem, and likely the more accurate the result will be.[41] The data needed for AI training include images, videos, audio/speech, and text.[42] It is argued that Al models undergo training using massive amounts of data to recognise patterns and make intelligent decisions or predictions.[43] This dissertation revolves around the use of this same massive data when such data is copyrighted. A thorough examination of the factors propelling the progress of AI necessitates an acknowledgement of the significance of cloud computing. Cloud computing refers to the provision of computing services, including servers, storage, databases, networking, software, analytics, and intelligence, over the internet (commonly known as "the cloud") to offer faster innovation, flexible resource allocation, and economies of scale.[44] Cloud computing allows more data to be stored on a permanent, retrievable basis.[45]

Most importantly, a thorough examination of the factors influencing AI development would be incomplete without mentioning machine learning (ML). Machine learning essentially refers to different methodologies that program software-driven machines to learn on their own and consequently enhance and optimise their functions.[46] Machine learning is a subfield of AI that refers to a machine's ability to imitate intelligent human behaviour.[47] The discussion in this section has pointed out the undeniable fact that many factors are at play in the AI field. This view is aptly put by the National Security Commission on AI, quoted by James E Baker *et al,*[48] stating that "AI

---

[41] *Supra* note 33 at 8.
[42] Dusan Simic 'Mastering the Essentials of AI Training Data: A Comprehensive Guide' available at *https://www.linkedin.com/pulse/mastering-essentials-ai-training-data-comprehensive-guide-dusan-simic-fclxf/* accessed 12 February 2024.
[43] Ravindra Kumar & Pankaj Kumar 'Training ai and copyright infringement: where does the law stand?' (2022) 2 *Indian Journal of Integrated Research in Law* 1.
[44] Microsoft Azure 'What is cloud computing?' available at *https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing#:~:text=Simply%20put%2C%20cloud%20computing%20is,resources%2C%20and%20economies%20of%20scale* accessed 12 February 2024.
[45] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 8.
[46] *Ibid.*
[47] Sara Brown 'Machine learning, explained' available at *https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained* accessed 12 February 2024.
[48] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 1.

is not a single piece of hardware or software, but rather, a constellation of technologies that gives a computer system the ability to solve problems and to perform tasks that would otherwise require human intelligence."

3.2.4. Types of AI

Having brought to light the inception of AI, the next step is to engage the various types of AI. AI can be broadly classified into Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGN) and hypothetically Artificial Super Intelligence (ASI). These types of AI are explored in detail below.

(i) Artificial Narrow Intelligence

First, Artificial Narrow Intelligence, also known as weak AI, refers to the "ability of computational machines to perform singular tasks at optimal levels, or near optimal levels, and usually better than, although sometimes just in different ways, than humans."[49] It is argued that weak AI lacks the ability to think or self-learn and the capacity to perform a set of pre-defined functions.[50] Weak AI refers to programs that merely try to replicate or duplicate whatever function it is meant to do.[51] In addition, weak AI encompasses many single-purpose technologies, *inter alia*, facial recognition, driverless vehicles, and drones.[52]

Furthermore, it is fascinating to observe that the narrow AI is "intelligent" in only one domain, a limitation on their ability to be used for multiple purposes or deal with certain complex situations outside their intended purpose.[53] Swamy concurs with the specialist view that the present AI is frail as its distinct qualities are twofold: (a) it involves direct human intercession in its development, and (b) it is confined to a solitary errand.[54] Arguments have been put forward that all AI currently in use falls under the weak AI category, and such AI is good at correlating, connecting, and classifying data, recognising patterns, and weighting probabilities.[55] Therefore,

---

[49] *Ibid* at 9.
[50] Binder Dijker Otte South Africa "An Introduction to Artificial Intelligence and Machine Learning" available at *https://www.bdo.co.za/getattachment/Industries/Financial-Services/Emerging-Tech-Toolbox/Artificial-Intelligence-and-Machine-Learning/FS-AI-Whitepaper-design-July-2020.pdf* accessed 13 February 2024.
[51] Rex Martinez 'Artificial Intelligence: Distinguishing Between Types & Definitions' (2019) 19 *Nevada Law Journal* 1015 at 1027.
[52] *Supra* note 48 at 9.
[53] *Supra* note 48 at 9..
[54] Raju Narayana Swamy 'Strong AI to Super-intelligence: How is AI placed vis-à-vis Intellectual Property Rights' (2021) 3 *International Journal of Computer Communication and Informatics* at 2.
[55] *Supra* note 48 at 9.

the author argues that genAI such as ChatGPT,[56] Google Gemini (Bard),[57] Midjourney,[58] GitHub Copilot,[59] form part of the weak AI given that they are confined to a solitary errand which is content generation and rarely for multipurpose use.

(ii) Artificial General Intelligence

Secondly, Artificial General Intelligence (AGI), also known as strong AI, form part of the AI family. Predictions have been made in the computer engineering field that AGI will emerge this century. This type of AI is an AI multitasking capacity that can serve multiple purposes.[60] AGI has the ability to think and make decisions like humans.[61] Therefore, this means that AGI, as predicted, will possess autonomy and the capability to think and behave as humans would. The most fascinating yet unsettling fact is that AGI is a model that can write and rewrite its own code and shift from task to task.[62] AGI is arguably the actual instantiation of the thing called "intelligence."[63] This means that AGI is a concrete or genuine manifestation of intelligence.

(iii) Artificial Super Intelligence

Thirdly, the two widely accepted types of AI discussed above, namely strong AI and weak AI. However, some voice the potential emergence of Artificial Super Intelligence (ASI).[64] ASI refers to an AI that is beyond general intelligence, where computers are generally smarter than humans with unlimited sources of information and energy.[65] Put differently, ASI is defined as a type of AI that can surpass human intelligence by demonstrating cognitive abilities and developing its own thinking skills.[66] However, the credibility of this concept has sustained some criticism from many

---

[56] OpenAI 'Introducing ChatGPT' available at *https://openai.com/blog/chatgpt* accessed 13 February 2024.
[57] Gemini 'Gemini' available at *https://gemini.google.com/?hl=en* accessed 13 February 2024.
[58] Midjourney 'About' available at *https://www.midjourney.com/home* accessed 13 February 2024.
[59] GitHub "The world's most widely adopted AI developer tool" available at *https://github.com/features/copilot* accessed 13 February 2024.
[60] James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023) 12.
[61] Binder Dijker Otte South Africa 'An Introduction to Artificial Intelligence and Machine Learning' available at *https://www.bdo.co.za/getattachment/Industries/Financial-Services/Emerging-Tech-Toolbox/Artificial-Intelligence-and-Machine-Learning/FS-AI-Whitepaper-design-July-2020.pdf* accessed 13 February 2024.
[62] *Supra* note 60 at 13..
[63] Raju Narayana Swamy 'Strong AI to Super-intelligence: How is AI placed vis-à-vis Intellectual Property Rights' (2021) 3 *International Journal of Computer Communication and Informatics* at 2.
[64] *Supra* note 61.
[65] *Supra* 60 at 14.
[66] Vijay Kanade 'What Is Super Artificial Intelligence (AI)? Definition, Threats, and Trends' available at *https://www.spiceworks.com/tech/artificial-intelligence/articles/super-artificial-intelligence/* accessed 13 February 2024.

computer engineers and government officials who dismiss ASI as science fiction and a distraction from the real and immediate challenges of today's narrow AI.[67] Therefore, while AGI is hypothetical, ASI is even more hypothetical. However, despite the futuristic possibility of the Super AI being dented by criticism, curious and enthusiastic minds are set to gather under one roof in Marina Bay Sands, Singapore[68], between 5 and 6 June 2024 to explore ASI. Until then, the concept remains a hypnotical one. The above-discussed AI is adequate to examine the issue of AI-enabled copyright infringement liability.

3.3. An overview of copyright and copyright infringement

3.3.1. Copyright and meaning in South Africa

In order to appreciate the next chapters, the reader must understand copyright, its national and international legal framework and the protection it offers. Copyright is a statutory creature regulated by the Copyright Act 98 of 1978[69] in South Africa. In *Moneyweb (Pty) Limited v Media 24 Limited and Another,*[70] the Court stated that copyright is an intellectual property right protected under section 25(1) of the Constitution. Section 25(1) of the Constitution provides that "no one may be deprived of property except in terms of law of general application, and no law may permit arbitrary deprivation of property."[71]

The Copyright Act is destitute of a definition of copyright. *Claasen and Another v TEC Novation Solution (Pty) Limited and Another,*[72] states that copyright law in South Africa is embodied in the Copyright Act 98 of 1978. The case further states that copyright governs the right to control the use of and distribution of artistic and creative works. The case quoted Harms JA's observation that "copyright, it has been said, is a technical subject."[73] Article 9(2) of the TRIPS Agreement provides that "[c]opyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such."[74] Therefore, protection is granted to works that have been reduced to a material form. This resonates with section 2(2) of the

[67] *Supra* note 60 at 15.
[68] Super AI 'Where AI meets the world' available at *https://www.superai.com/* accessed 13 February 2024.
[69] *King v SA Weather Services* 2009 (3) SA 13 (SCA) para 6.
[70] *Moneyweb (Pty) Limited v Media 24 Limited and Another* 2016 (3) All SA 193 (GJ), para 108.
[71] Constitution of the Republic of South Africa, 1996.
[72] *Claasen and Another v TEC Novation Solution (Pty) Limited and Another* (2017/40521) 2018 (ZAGPPHC), para 5.
[73] *Memory Institute SA CC t/a SA Memory Institute v Hansen and Others* (253/2002) 2003 (ZASCA), para 5.
[74] Agreement on Trade-Related Aspects of Intellectual Property Rights, 1994.

Copyright Act. Dean and Dyer also weigh in on the definition of copyright, stating that it grants the copyright holder exclusive statutory rights, limited in duration, in rendering or performing certain dealings or acts related to specific works, and conversely, copyright offers the right to prevent others from enjoying the exclusive rights.[75]

The rationale behind copyright is to grant the copyright owner the opportunity, throughout the duration of the right, to exploit and profit from the protected work, which serves as a motivation for further innovation and creativity.[76] In addition, copyright is the only form of intellectual property protected without registration, and it is subject to meeting all protection conditions in the Act.[77] However, copyright protects only the form of expression of ideas and not the ideas themselves.[78] Copyright protection is afforded to the following works: literary works; musical works; artistic works; cinematograph films; sound recordings; broadcasts; programme-carrying signals; published editions; and computer programs.[79] Most importantly, in order for works to be eligible for protection, they must be original,[80] reduced to material form,[81] by a qualified person[82] or the first publication took place in South Africa.[83]

Copyright has its footprints in the international legal frameworks as well. Since copyright is a territorial asset,[84] an effort to ensure reciprocal protection of copyright in multiple jurisdictions is reflected in international law. For example, the Berne Convention for the Protection of Literary and Artistic Works (Berne Convention)[85] is one of the relevant international legal instruments and it contains the national treatment principle.[86] The national treatment principle means that the rights of foreigners and local subjects are protected equally under domestic law in signatory states.[87] Dean and Dyer distil the concept in detail, stating that "in the copyright context, national treatment promises foreign copyright owners that their works will enjoy the same treatment as the protecting

---

[75] Owen Dean & Alison Dyer (eds) *Dean & Dyer Introduction to Intellectual Property Law* (2014) 3.
[76] *Ibid* at 3.
[77] ss2-4 of the Copyright Act.
[78] World Intellectual Property Organisation *WIPO Intellectual Property Handbook* (2004) 40.
[79] s2(1)(a)-(i) of the Copyright Act.
[80] s2(1) of the Copyright Act.
[81] s2(2) of the Copyright Act.
[82] s3(1) of the Copyright Act.
[83] s4(1) of the Copyright Act.
[84] Owen Dean & Alison Dyer (eds) *Dean & Dyer Introduction to Intellectual Property Law* (2014) 6.
[85] Berne Convention for the Protection of Literary and Artistic Works, 1979.
[86] Article 2(6) and 5(1) of the Berner Convention.
[87] Article 3(1) of the Agreement on Trade-Related Aspects of Intellectual Property Rights, 1994.

country gives to the works of its own nationals."[88]

The other important international instrument is the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement)[89] which deals with, *inter alia*, intellectual property rights, enforcement and dispute resolution. Article 9(2) of the TRIPS Agreement requires member states (such as South Africa) to comply with Articles 1 through 21 of the Berne Convention except in relation to moral rights provided in article 6*bis* of the Berne Convention. Therefore, the TRIPS Agreement builds on the Berne Convention. There are other instruments that deal with copyright at the global stage; *viz*, WIPO Copyright Treaty (WCT)[90] and Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled.[91] Therefore, while copyright regimes are territorial, legal mechanisms are in place to ensure seamless reciprocal enforcement of copyright in different jurisdictions which are signatories to international frameworks.

3.3.2. Exclusive Rights

The copyright holder is entitled to a range of economic rights that grant them exclusive control over their works and their moral rights. In the realm of economic rights, the copyright holder possesses the authority to reproduce, publish, make adaptations, distribute, and perform their work. This is a display of the power the copyright holder possesses to dictate how their work is disseminated and consumed by the public, ensuring that they can derive profit from its use and exploitation. On the moral front, the copyright holder or the author is protected by paternity and integrity rights. The author will explore some of these economic rights and moral rights, shedding light on their being and the legal framework that upholds such rights.

(i) Economic Rights

The copyright holder is endowed with the right to reproduce their literary or musical work in terms of section 6(a) of the Copyright Act.[92] In *Galago Publishers (Pty) Ltd v Erasmus,* the court stated

[88] Owen Dean & Alison Dyer (eds) *Dean & Dyer Introduction to Intellectual Property Law* (2014) 6.
[89] Agreement on Trade-Related Aspects of Intellectual Property Rights, 1994.
[90] WIPO Copyright Treaty, 1996.
[91] Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled, 2013.
[92] Copyright Act 98 of 1978. For similar provisions in relation to other types of works, see: ss7(a); 8(a); 9(a); 11A; 11B(a).

that to "reproduce" means to copy.[93] Tong argues that the reproduction of work is pivotal in infringement claims and remains inevitably a precursor to infringement rights reserved for the copyright holder by the Act.[94] Furthermore, the right to publish literary works is vested in the copyright holder in terms of section 6(b) of the Copyright Act, which allows them to publish unpublished works or creations to the public through publications means including online platforms. In addition, the copyright holder possesses the exclusive right to perform work in public in terms of section 6(c) of the Copyright Act, which allows them to control how and when their work is performed. In *SAMRO vs Svenmill Fabrics*[95] the court stated that the meaning of the word "public" must be considered in the light of the relationship between the audience and the copyright holder and not in the light of the relationship between the audience and the performer.

Expanding on economic rights, the copyright holder possesses the exclusive right to make adaptations of their work in terms of section 6(f) of the Copyright Act.[96] While section 1(1) of the Copyright Act defines the term "adaptation," a more digestible definition is provided by WIPO[97] wherein it states that an "adaptation is commonly defined as the modification of a work from one form to another, such as transforming a novel into a motion picture, or modifying a work to make it suitable for various conditions of exploitation or utilisation in different conditions." Another crucial economic right pertains to the authority to engage in, or authorise, the direct or indirect distribution of program-carrying signals by any distributor to the public or a section of the public in South Africa. While economic rights vary depending on the work in question, these economic rights collectively form the bedrock of the copyright holder's control.

(ii) Moral Rights

Furthermore, shifting attention to the moral rights domain, the right to paternity and integrity engraved in section 20(1) of the Copyright Act demands to be elucidated. Section 20(1) of the Copyright Act provides that "the author shall have the right to claim authorship of the work, … and to object to any distortion, mutilation, or other modification of the work where such action is or would be prejudicial to the honour or reputation of the author." Infused in the aforesaid

---

[93] *Galago Publishers (Pty) Ltd & Another v Erasmus* 1989 (1) All SA 431 (A) at 432.
[94] Lee-Ann Tong "Copyright Protection for Computer Programs in South Africa: Aspects of Generis Categorisation" (2009) *JWIP* 284.
[95] *Southern African Music Rights Organisation Ltd v Svenmill Fabrics (Pty) Ltd* 1983 (1) SA 608 (C), para 33.
[96] Act 98 of 1978. For similar provisions, see: ss7(e); 8(e) and 11B(f).
[97] *Supra* note 78 at 45.

provision is the right to claim authorship of the work, which is the right to paternity. Equally significant within the moral right provision is the author's right to object to any distortion, mutilation, or other modification of the work where such action is or would be prejudicial to the honour or reputation of the author, the author's right to the integrity of the work which is referred to as the right to integrity. In addition, a breach of moral rights will be treated as an infringement of copyrights.[98]

3.3.3. Infringement of Copyright

In order to gain a comprehensive understanding of this dissertation, it is imperative to have a comprehensive understanding of infringement within the copyright ecosystem of South Africa. The Copyright Act delineates two forms of civil infringement: direct (primary) infringement and indirect (secondary) infringement.[99] A detailed exploration of these forms of infringement follows below.

(i) Direct Infringement

On the one hand, in terms of section 23(1) of the Copyright Act, direct infringement transpires when "any person, not being the owner of the copyright, who, without the licence of such owner, does or causes any other person to do, in the Republic, any act which the owner has the exclusive rights to do or to authorise." The exclusive rights in the aforesaid provision were explained above.[100] This provision outlaws the exercising of exclusive rights to a person who is the copyright holder or without authorisation from the copyright holder. In addition, it is essential to read section 23(1) in tandem with section 1(2A) of the Copyright Act, where reference to the doing of any act in relation to any work should be interpreted as a reference to the doing of any such act in relation to any substantial part of such work. Thus, engaging in exclusive acts concerning an insubstantial part without authorisation does not amount to infringement.[101]

Furthermore, section 21(1) extends liability to a person who "causes any other person to do" exclusive acts. This direct infringement is known as causal infringement and may occur in two forms: first, where a person causes another person to do such exclusive acts, and second, where a

---

[98] s20(1) of the Copyright Act.
[99] Lee-Ann Tong "Copyright Protection for Computer Programs in South Africa: Aspects of Generis Categorisation" (2009) *JWIP* 283.
[100] *Supra* section 3.3.2: Exclusive Rights.
[101] *Supra* note 99.

person causes the authorisation of a restricted act to be carried out.[102] Direct infringement may occur when one unconsciously copies a and the absence of intention to create an infringing work.[103] An observation is made that direct copyright infringement does not only extend to a person who does any of the exclusive acts without the authorisation of the copyright holder, but it extends to a person who caused such an act to be done.[104]

The courts have been instructive in examining copyright infringement. Dealing with reproduction as an alleged infringing act, the court in *Galago Publishers (Pty) Ltd & Another v Erasmus*[105] can assist shed more light. *In casu*,[106] the court stated that to prove copyright infringement of the original work; one must establish two things: first, "that there is sufficient objective similarity between the alleged infringing work and the original work, or a substantial part thereof, for the former to be properly described, not necessarily as identical with, but as a reproduction or copy of the latter," and second, "that the original work was the source from which the alleged infringing work was derived, i.e. that there is a causal connection between the original work and the alleged infringing work." This test of examining infringement was confirmed in *Dexion Europe Ltd v Universal Storage Systems (Pty) Ltd*[107] where a two-stage inquiry was employed to establish copying. In *Haupt t/a Softcopy v Brewers Marketing Intelligence,*[108] the court stated that whether a substantial part has been produced lies more on quality than quantity of the reproduced. Therefore, the author can argue that examining substantial similarity is not a numerical exercise but an examination of the quality of the work reproduced.

Second, the second leg of the two-stage inquiry establishes a causal connection between the original work and the alleged infringing work. When examining the existence of a causal connection between the original work and the alleged infringing work, it simply indicates that the alleged infringing work has been derived from the original work as a source.[109] Therefore, direct infringement would have been established when the alleged work is objectively similar to the

---

[102] Sylvia Papadopoulos & Sizwe Snail ka Mtuze *Cyberlaw @ SA IV: the law of the Internet in South Africa* 4 ed (2022) 218.
[103] *Ibid* at 214.
[104] *Supra* note 99.
[105] *Galago Publishers (Pty) Ltd & Another v Erasmus* 1989 (1) All SA 431 (A) at 432.
[106] *Ibid.*
[107] *Dexion Europe Ltd v Universal Storage Systems (Pty) Ltd* 2002 JOL 10172 (SCA) para 4.
[108] *Haupt t/a Softcopy v Brewers Marketing Intelligence* 2006 (4) SA 458 (SCA) para 44.
[109] *Galago Publishers (Pty) Ltd & Another v Erasmus* 1989 (1) All SA 431 (A) at 432.

original work and a causal link exists.

(ii) Indirect Infringement

On the other hand, indirect infringement, also known as secondary infringement, occurs when an action falls within the ambit of subsections 23(2) and 23(3) of the Copyright Act. It is argued that indirect infringement aims to safeguard the copyright holder's interests in controlling the market for their work during the subsistence of copyright and the public performance of the work.[110]

First, section 23(2) of the Copyright Act encapsulates instances under which, when done without authorisation from the copyright holder, would constitute infringement. Section 23(2)(a)-(d) reads as follows:

> "Copyright shall be infringed by any person who, without the licence of the owner of the copyright and at a time when copyright subsists in a work - "(a) imports an article into the Republic for a purpose other than for his private and domestic use; (b) sells, lets, or by way of trade offers or exposes for sale or hire in the Republic any article; (c) distributes in the Republic any article for the purposes of trade, or for any other purpose, to such an extent that the owner of the copyright in question is prejudicially affected; or (d) acquires an article relating to a computer program in the Republic, if to his knowledge the making of that article constituted an infringement of that copyright or would have constituted such an infringement if the article had been made in the Republic."

The notable essential element is the requirement of knowledge on the part of the person alleged to be infringing. In addition, this provision means dealing in infringing copies amounts to infringement.

Second, in terms of section 23(3) of the Copyright Act, indirect infringement of copyright in copyright in a literary or musical work may occur when "any person who permits a place of public entertainment to be used for a performance in public of the work, where the performance constitutes an infringement of the copyright in the work: Provided that this subsection shall not apply in a case where the person permitting the place of public entertainment to be so used was not aware and had no reasonable grounds for suspecting that the performance would be an

[110] *Supra* note 102.

infringement of the copyright." Therefore, letting a place of public entertainment to be used for a performance of a copyrighted work without proper authority constitutes infringement. However, the person must have guilty knowledge, meaning that the person who permitted the performance must have been aware or must have grounds to suspect that the performance would infringe copyright.

3.4. Conclusion

This chapter has explored the landscapes of both AI and copyright law and provided a deep understanding of the intersection. In the realm of AI, the chapter examined the various types of AI, ranging from ANI to the hypothetical concept of ASI, and further explained the synergistic factors driving the advancement of AI, including complex algorithms, computational speed, sensor technology, data availability, cloud computing, and machine learning, all of which contribute to AI's progress. This dissertation focuses on genAI. In the realm of copyright law, the author examined the protection offered in South Africa, including economic rights such as reproduction, publication, and distribution, as well as moral rights such as the right to paternity and integrity of the work. A further step was taken to explore copyright infringement, specifically direct and indirect infringement.

CHAPTER 4: IMPUTING LIABILITY FOR COPYRIGHT INFRINGEMENT - GenAI INPUT PHASE

4.1. Introduction

This chapter examines the assignment of liability for copyright infringement that occurs at the training phase of genAI, herein referred to as the "input phase." In doing so, the author will provide an insightful background of how infringement occurs first. Second, it imputes liability on potential candidates and analyses the implication of holding each party liable; and third, it examines how other jurisdictions deal with copyright infringement during the training phase. This dissertation focuses on genAI. Therefore, the reader must bear in mind that genAI refers specifically to models that are capable of generating "new" content in response to user inputs called "prompts."[1]

4.2. How does infringement occur during the genAI training phase?

4.2.1. Simplifying the Puzzle: Understanding the GenAI Training Process

In order to function, genAI must learn from existing materials, some of which may be protected by copyright owned by another party. For example, in an interview with Forbes, David Holz, the CEO of Midjouney, was questioned about obtaining consent from the living artists or work still under copyright, and his response was "No. There isn't really a way to get a hundred million images and know where they're coming from."[2] It is against this backdrop that copyright infringement at the genAI training phase is rooted. However, the author must understand the genAI training process to comprehend how copyright infringement possibilities arise.

According to Ravindra Kumar & Pankaj Kumar,[3] in a general sense, the training of genAI can be categorised into six stages, outlined below. The initial stage involves data ingestion, wherein a vast amount of data is collected. After data collection, the next step involves the removal of inconsistent, incorrect, and skewed information. The source data must undergo filtration, and any prejudices or biases will be removed. Further, the data must be formatted to ensure uniformity and remove anomalies. The data formatting process should be tailored to best suit the machine

[1] *Authors Guild v. OpenAI, Inc.*, No. 1-23-cv-08292 Document 39 (S.D.N.Y. 2023), at 9.
[2] Forbes "Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy" available at *https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=65ab1ad82d2b* accessed 20-02-2024.
[3] Ravindra Kumar & Pankaj Kumar 'Training ai and copyright infringement: where does the law stand?' (2022) 2 *Indian Journal of Integrated Research in Law* 3.

learning model. Further, data is converted into patterns that provide relevant information that can be fed into the learning algorithms. Finally, the data is divided into a "training set" and an "evaluation set." The "training set" comprises the data utilised by the model for making predictions.

GenAI uses substantial amounts of input data, including images, videos, text and other artistic content, as part of its learning process.[4] One illustration involves a text-generating genAI. GenAI models designed to recognise input text and generate output text are built on large language models (LLMs).[5] LLMs undergo training by ingesting large amounts of text referred to as the "training dataset."[6] LMs copy text from the training dataset, and extract expressive information.[7] Once an LLM has copied and ingested the text in its training dataset, it can emit convincingly naturalistic text outputs in response to user prompts.[8] ChatGPT[9] is a language model designed for paying users to enter text prompts, and it responds by simulating human reasoning.[10] Thus, genAI ingestion of pre-existing material during its training phase enables it to simulate or mimic human capability when responding to user prompts.

An additional illustration is Anthropic's Claude AI,[11] which undergoes training with material collected by scraping the text directly from websites and other digital sources onto Anthropic's servers.[12] This process is facilitated through the use of automated tools, such as bots and web crawlers, and/or third parties provide material which may have been harvested through web scraping.[13] In sum, it is claimed that Stability AI copies billions of text-and-image pairings, loading them into computer memory for the purpose of training a model.[14] It is further claimed that Stable Diffusion underwent training using 5 billion image-text pairs sourced from datasets curated by Large-Scale Artificial Intelligence Open Network (LAION), a German entity that works

[4] Enrico Bonadio & Luke McDonagh 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity' (2020) 2 *Intellectual Property Quarterly* 3.
[5] *Supra* note 1.
[6] *Ibid.*
[7] *Ibid.*
[8] *Tremblay, et al., v OpenAI, Inc., et al*., No. 3:23-cv-03223 Document 1 (N. D. Cal. 2023), at 1.
[9] OpenAI 'Introducing ChatGPT' available at *https://openai.com/blog/chatgpt* accessed 13 February 2024.
[10] *Supra* note 8.
[11] Anthropic 'Talk to Claude' available at *https://claude.ai/login?returnTo=%2F* accessed 27-02-2024.
[12] *Concord Music Group, Inc v Anthropic PBC*, No. 3-23-cv-01092 Document 1 (M.D. Tenn. 2023), at 17.
[13] *Ibid.*
[14] *Getty Images (US) Inc v Stability AI Inc* No. 1:23-cv-00135 Document 1 (D. Del. 2024), at 12.

in conjunction with and is sponsored by Stability AI.[15] An interesting observation is that Stability AI encodes the images, which involves creating smaller versions of the images that take up less memory. The above section provided insight into the training of genAI and how a massive amount of pre-existing material is copied.

4.2.2. Copyright Infringement Risks in GenAI Training

The training of genAI has revealed that substantial volumes of material are consumed, raising concerns about the heightened risk of copyright infringement during the genAI training process. This section will explain possible copyright infringement when one, without authorisation, engages in the reproduction and adaptation of works while training genAI. Pursuant to section 23(1) of the Copyright Act, direct infringement occurs when "any person, not being the owner of the copyright, who, without the licence of such owner, does or causes any other person to do, in the Republic, any act which the owner has the exclusive rights to do or to authorise."[16] The author can observe that the aforesaid provision means that in South Africa, engaging in or causing another person to exercise exclusive rights held by copyright owners without authorisation for the purpose of training genAI would constitute copyright infringement.

(i) Infringement of the Right to Reproduce the Work in Any Manner

First, section 6(a) of the Copyright Act provides that the exclusive right to reproduce or authorise reproduction of works vests in the copyright holder.[17] The Copyright Act further provides that reproduction, in relation to:

> "a literary or musical work or a broadcast, includes a reproduction in the form of a record or a cinematograph film; an artistic work, includes a version produced by converting the work into a three-dimensional form or, if it is in three dimensions, by converting it into a two-dimensional form; any work, includes a reproduction made from a reproduction of that work; and references to 'reproduce' and 'reproducing' shall be construed accordingly."[18]

This definition is open-ended because of the legislature's use of the word "includes," denoting that it can be interpreted to capture acts of reproduction that are novel owing to

[15] *Andersen et al v. Stability AI Ltd. et al* No. 3:23-cv-00201-WHO Document 117 (N. D. Cal. 2023), at 2.
[16] Copyright Act 98 of 1978.
[17] s6(a) of the Copyright Act 98 of 1978. For similar provisions, see: ss7(a); 8(a); 9(a); 11A; 11B(a).
[18] s1(1)(a)-(c) of the Copyright Act.

technological advancement. In addition, section 1(2A) of the Copyright Act provides that the reproduction of a work shall, unless the context otherwise indicates, be construed as a reference to the reproduction of a substantial part of the work.

In examining the definition of reproduction, insights from judicial precedents can shed more light. In *Galago Publishers (Pty) Ltd v Erasmus,*[19] the court stated that to "reproduce" means to copy. A similar perspective is expressed in the case of *Payen Components SA Ltd v Bovic CC & Others.*[20] In *Haupt t/a Softcopy v Brewers Marketing Intelligence,* the Supreme Court of South Africa stated that whether a substantial part of the work has been reproduced depends much more on quality than quantity of the reproduced.[21] In this particular case, the Court determined that the copied components of the software were deemed valuable ingredients of the program, leading to the conclusion that the copied material was a substantial part of the original work.[22]

As illustrated earlier, [23] genAI is provided with extensive amounts of material during the training process, a procedure accomplished through the act of copying. For instance, Stability AI copies billions of text-and-image pairings, loading them into computer memory for the purpose of training a model.[24] The reader must recall that copying amounts to reproduction as per *Galago Publishers (Pty) Ltd v Erasmus*.[25] This is extracted from section 1(1) of the Copyright Act, which states that "copy" means a reproduction of a work. Therefore, where a person or company copies copyrighted material without authorisation from the copyright holder would *prima facie* constitute infringement in South Africa. In *Haupt t/a Softcopy v Brewers Marketing Intelligence,* the Court deemed copied components of the software valuable ingredients of the program, leading to the conclusion that the copied material was a substantial part of the original work.[26] In the same vein, when analysing whether genAI training violated copyright, the fact that text, images, audio or videos are valuable ingredients renders the act of reproducing the material an infringement of copyright to reproduce.

---

[19] *Galago Publishers (Pty) Ltd & Another v Erasmus* 1989 (1) All SA 431 (A) at 432.
[20] *Payen Components SA Ltd v Bovic CC & Others* 1995 (2) All SA 600 (A), at 606.
[21] *Haupt t/a Softcopy v Brewers Marketing Intelligence* 2006 (4) SA 458 (SCA) para 45.
[22] *Ibid,* para 45.
[23] See subheadings: 4.2.1.
[24] *Getty Images (US) Inc v Stability AI Inc* No. 1:23-cv-00135 Document 1 (D. Del. 2024), at 12.
[25] *Galago Publishers (Pty) Ltd & Another v Erasmus* 1989 (1) All SA 431 (A) at 432.
[26] *Ibid,* para 45.

In addition, this view is buttressed by legal scholars. It is argued that during the training process, literal reproduction of the copyrighted work may occur, as the same input datasets are copied multiple times throughout the learning process.[27] Bonadio and McDonagh submit that copyright infringement occurs where a violation of the reproduction materialises because machines need to reproduce the data which they access during the learning process.[28] Therefore, reproduction as an exclusive right reserved for the copyright holders may be breached during the training phase of a genAI model. However, this infringement could materialise specifically when the genAI training is conducted without proper authorisation.

(ii) Infringement of the right to making an adaptation of the work

Second, section 6(f) of the Copyright Act provides that the exclusive right to make an adaptation of the work or to authorise such vests in the copyright holder. The Copyright Act offers the definition of "adaptation," which varies depending on the type of work under consideration. In relation to a literary work, adaptation includes:

> "in the case of a non-dramatic work, a version of the work in which it is converted into a dramatic work; in the case of a dramatic work, a version of the work in which it is converted into a non-dramatic work; a translation of the work; or a version of the work in which the story or action is conveyed wholly or mainly by means of pictures in a form suitable for reproduction in a book or in a newspaper, magazine or similar periodical."[29]

In relation to a musical work, adaptation "includes any arrangement or transcription of the work, if such arrangement or transcription has an original creative character."[30] The Copyright Act further provides that in relation to an artistic work, adaptation "includes a transformation of the work in such a manner that the original or substantial features thereof remain recognizable."[31] Lastly, in relation to a computer program, adaptation includes –"a version of the program in a programming language, code or notation different from that of the program; or a fixation of the

[27] Ravindra Kumar & Pankaj Kumar 'Training ai and copyright infringement: where does the law stand?' *Indian Journal of Integrated Research in Law* (2022) 4.
[28] Enrico Bonadio & Luke McDonagh 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity' (2020) 2 *Intellectual Property Quarterly* 8.
[29] s1(1)(a)(i)-(iv) of the Copyright Act.
[30] s1(1)(b) of the Copyright Act.
[31] s1(1)(c) of the Copyright Act.

program in or on a medium different from the medium of fixation of the program."[32]

As indicated earlier, copyright infringement occurs when the exclusive act to make an adaptation is performed by an unauthorised person. Instances where copyright infringement would occur at the training stage can emerge. For instance, it is argued that "Stability AI encodes the images, which involves creating smaller versions of the images that take up less memory."[33] The author observes that adaptation occurs at the training stage because the definition of adaptation, under artistic work, captures creating small versions of images. The basis for this reasoning is that adaptation can be raised where one transforms work in such a manner that the original or substantial features thereof remain recognisable. Therefore, since the original and/or substantial feature of the image remains recognisable when transformed into smaller versions, the argument stands that the genAI trainers adapted the images. Hence, it logically follows that adaptation of images occurs, in this case without authorisation, leading to copyright infringement.

In addition, should a party evade adaption due to circumstances, it can still emerge during the altering and adding of audio. It is submitted[34] that Stability AI adds visual "noise" to encoded images, intentionally modifying them to make it progressively more challenging to discern the visual quality. This intentional degradation in visual quality is aimed at training the model to eliminate the introduced "noise."[35] The addition of noise and intentionally modifying quality implies that the genAI trainer is making an adaptation. This is an exclusive act or right, reserved for the copyright holder and those they authorise to make an adaptation.

However, it is not always crystal clear that copyright infringement has occurred during the training stage, considering the complexities of the ever-evolving field of computer science. An example can be extracted from *Huckabee et al v. Bloomberg L.P. et al*.[36] In this case, it is raised that LLMs work by leveraging neural networks, specifically recurrent neural networks. A noteworthy detail is that after collecting a training dataset, AI companies "tokenize," or break down the text contained in a training dataset into tiny pieces called tokens. These tokens can be as short as a single letter or as long as a word or even a phrase, which helps the LLM understand the

[32] s1(1)(c) of the Copyright Act.
[33] *Getty Images (US) Inc v Stability AI Inc* No. 1:23-cv-00135 Document 1 (D. Del. 2024), at 12.
[34] *Ibid.*
[35] *Ibid.*
[36] *Huckabee et al v. Bloomberg L.P. et al,* No. 1:23-cv-09152 Document 1 (S.D.N.Y. 2023), at 10.

structure of language. This means that it will prove a *cal-de-sac* for one to raise adaptation in such a case given the complexities attached to training. The type of work and the conduct of the genAI trainer will influence the establishment of the violation of the right to make an adaptation.

(iii) Seeking refuge of fair dealing doctrine

Attempts may be made to rely on fair dealing as a defence against copyright infringement, but its application varies in degree depending on the works.[37] The Copyright Act is instructive. Dealing with literary and musical work, section 12(1)(a)-(c) of the Copyright Act provides, *inter alia*, that:

> "Copyright shall not be infringed by any fair dealing with a literary or musical work- (a) for the purposes of research or private study by, or the personal or private use of, the person using the work; (b) for the purposes of criticism or review of that work or of another work; or (c) for the purpose of reporting current events - (i) in a newspaper, magazine or similar periodical; or (ii) by means of broadcasting or in a cinematograph film: Provided that, in the case of paragraphs (b) and (c) (i), the source shall be mentioned, as well as the name of the author if it appears on the work."

It is important to note that while section 12 deals with literary and musical works, it applies *mutatis mutandis* to other works in varying degrees. In relation to artistic works, section 12(1) applies without limitation.[38] With regard cinematograph films, section 12(1)(b) and (c) apply excluding 12(1)(a).[39] This means that that research, or private study, or the personal or private use of cinematograph films is not covered by fair dealing. In addition, with respect to sound recordings, the Legislature applies only section 12(1)(b) and (c) to the exclusion of 12(1)(a) of the Copyright Act. This means that private use is not justified and as such format shifting it outlawed.[40] The Legislature also limits application of section 12(1) only to 12(1)(b) and (b) regarding computer programmes.[41] However, section 12(1) applies broadcasts and published editions as a whole.[42] It is worth noting that the legislature found it unnecessary to replicate the provisions of section 12 of the Copyright Act for other types of works, opting instead to use the phrase: "[t]he provisions of section 12… shall *mutatis mutan*dis apply with reference to [adds the specific work]."

---

[37] Owen Dean & Alison Dyer (eds) *Dean & Dyer Introduction to Intellectual Property Law* (2014) 83.
[38] s15(4) of the Copyright Act.
[39] s16(1) of the Copyright Act.
[40] s17 of the Copyright Act.
[41] s19B(1) of the Copyright Act.
[42] ss 18 and 19A of Copyright Act.

In *Moneyweb (Pty) Limited v Media,* the Court stated that fair dealing is an elastic concept; it is impossible to lay any hard-and-fast definition because it is a matter of fact.[43] The interpretation of fair dealing, however, sheds light on the fact that it must be research, private study, private use of the work, and reporting the current events. Arguments supporting genAI training as a non-infringing act will not hold water because the reproduction is neither for personal use nor private study, as shown in the US, where companies such as OpenAI profit from the deployment of these genAI models. For example, OpenAI's ChatGPT charges $20 USD (R363.54) per month for premium subscription.[44] In addition, it is difficult to rely on reporting current events because the names of the authors and sources are impractical to mention, given that most of the works do not have metadata which can be used to track its source. Therefore, it will be a bold move to attempt and use the current fair dealing regime to allow the training of genAI. Until amendments are made, South Africa will remain unattractive to AI start-ups and likely to fail to catch up on the AI race. Amendment of the current regime to permit the training of AI without infringing copyright could change the aforesaid perception. However, such relaxation of copyright to accommodate training of genAI should be cautiously implemented to avoid unfairly treating creators.

4.3. How would other jurisdictions address copyright infringement?

After examining the assignment of copyright infringement and liability during the training phase in South Africa, it would be advantageous to investigate how other jurisdictions currently handle or would potentially address this matter. The author endeavours to ascertain whether a copyright infringement case could indeed emerge during the training phase, and if so, to determine who is liable.

4.3.1. Japan

It is unsurprising that Japan has garnered the distinction of being labelled the "world's most AI-friendly" country and a "machine learning paradise," given its tolerant stance on the ingestion of copyrighted material for both non-commercial and commercial purposes. [45] In Japan, article 30-

[43] *Moneyweb (Pty) Limited v Media 24 Limited & Another* 2016 JOL 35803 (GJ), para 114.
[44] OpenAI 'Pricing' available at *https://openai.com/chatgpt/pricing/* accessed 15 November 2024.
[45] Scott Warren & Joseph Grasser 'Japan's New Draft Guidelines on AI and Copyright: Is It Really OK to Train AI Using Pirated Materials?' available at *https://www.privacyworld.blog/2024/03/japans-new-draft-guidelines-on-ai-and-copyright-is-it-really-ok-to-train-ai-using-pirated-materials/* accessed 14 March 2024.

4(ii) of the Copyright Act,[46] provides that it:

> "[I]s permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation: …(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data."

Article 30-4(ii) of the Copyright Act means that it is permissible to use data for data analysis, which is "the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data." This provision applies to genAI. However, such use is limited where it unreasonably prejudices the interests of the copyright holder. The other caveat is that the use of copyrighted works is not permitted where the purpose is to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work. It has been argued that the use of data for training genAI does not aim to create enjoyment of the ideas or sentiments expressed in the work.[47] Legal scholars posit that article 30-4 of the Copyright Act applies to the training of genAI models used for commercial and non-commercial purposes.[48] Dermawan holds that:

> "[T]he Japanese TDM exception is regarded as the 'broadest TDM exception in the world' for the following reasons: (1) TDM applies to both commercial and non-commercial purposes; (2) the Japanese TDM exception applies to any exploitation regardless of the rightsholders reservations; (3) exploitation by any means is permitted; and (4) no lawful access is required."[49]

---

[46] Copyright Act 48 of 1970, (as amended up to January 1, 2022)
[47] Shinnosuke Fukuoka, Tomonobu Murata and Atsuki Mizuguchi 'Legal Issues in Generative AI under Japanese Law-Copyright' available at *https://www.nishimura.com/sites/default/files/newsletters/file/robotics_ai_230711_en.pdf*, accessed 14 March 2024.
[48] Michihiro Nishi 'Japanese Law Issues Surrounding Generative AI: ChatGPT, Bard and Beyond' available at *https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2023/10/Japanese-Law-Issues-Surrounding-Generative-AI.html*, accessed 14 March 2024.
[49] Artha Dermawan 'Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?' (2023) *Journal of World Intellectual Property Law* 54.

Furthermore, analysis has indicated that companies are permitted to incorporate copyrighted material into their genAI training data without limitations, however, AI output that is excessively similar to, and reliant upon, such copyrighted works will not be permitted.[50] Michihiro suggests that it is not prudent to train the genAI model using a narrow set of works, such as specific types of works or works authored by particular individuals.[51] This caution arises from the heightened risk that a small training dataset could lead to the genAI-generated output resembling pre-existing copyrighted works too closely, potentially resulting in copyright infringement.[52] This means that copyright infringement would arise in the output phase where works are generated by genAI, but at the training phase, the genAI companies have flexible legal insulation at their disposal. South Africa can undoubtedly borrow this exception at the genAI training phase in order to catch up with AI development at a global stage.

4.3.2. United Kingdom

In the United Kingdom (UK), the text and data mining (TDM) exception might tempt one to argue that genAI training is accommodated in the copyright regime. However, the reality is not as straightforward as one might assume. In terms of section 29A(1)(a)-(b) of UK Copyright, Designs and Patents Act:

> "the making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that— (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise)."[53]

TDM is defined as the utilisation of computational techniques to analyse extensive amounts

---

[50] John Donegan 'The US should look at Japan's unique approach to generative AI copyright law' available at *https://insights.manageengine.com/artificial-intelligence/the-us-should-look-at-japans-unique-approach-to-generative-ai-copyright-law/#:~:text=Unlike%20the%20U.S.%20and%20E.U.,the%20gen%20AI%20arms%20race,* accessed 14 March 2024.

[51] Michihiro Nishi 'Japanese Law Issues Surrounding Generative AI: ChatGPT, Bard and Beyond' available at *https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2023/10/Japanese-Law-Issues-Surrounding-Generative-AI.html*, accessed 14 March 2024.

[52] *Ibid.*

[53] s29A of Copyright, Designs and Patents Act 1988.

of information with the aim of identifying patterns and trends.[54]

This means that in order for one to be insulated by the TDM exception, one must; first, have lawful access to the work; and second, use it for the sole purpose of research for non-commercial purposes. One can observe that unlawful scrapping of works which one does not have lawful access to do does not fall under this category. Most importantly, unlike in Japan, the use must be for non-commercial purposes. Berry argues that under the current UK law, should a commercial AI company intend to use a third-party material publicly accessible on the internet for training its models, it might be necessary to seek permission from each relevant copyright holder.[55] She goes on to voice the concern that identifying the relevant copyright holders can often be challenging, particularly considering the vast amount of data involved and the lack of a streamlined licensing framework for this purpose. Consequently, obtaining permission to train genAI models can become a cumbersome process. Therefore, the UK is not flexible in accommodating the genAI training phase as compared to its peers such as Japan. The lack of this flexible accommodation of genAI at the training phase would hold the genAI company liable for copyright infringement.

4.3.3. European Union

The European Union (EU) permits text and data mining, which is an exception in whose arms genAI companies might seek refuge. Article 2(2) of the EU Directives defines TDM as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations."[56] Article 4(1) of the EU Directives provides that Member states shall provide an exception which allows "reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining." However, article 4(3) of the Directives provides that:

> "[T]he exception or limitation provided … shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightsholders

[54] Rachel Montagnon & Sungmin Cho 'UK withdraws plans for broader Text and Data Mining (TDM) copyright and database right exception' available at *https://hsfnotes.com/ip/2023/03/01/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception/,* accessed 15 March 2024.
[55] Kathy Berry 'UK fails to agree AI/copyright code of practice,' available at *https://www.linklaters.com/en/insights/blogs/digilinks/2024/february/uk-fails-to-agree-ai---copyright-code-of-practice#:~:text=Under%20the%20current%20legal%20framework,their%20works%20by%20AI%20developers,* accessed 15 March 2024.
[56] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

> in an appropriate manner, such as machine-readable means in the case of content made publicly available online."

Article 4(3) of the Directives provides an opt-out mechanism whereby rightsholders can prevent an entity or another from conducting TDM.[57] Margoni and Kretschmer voice that Article 4 of the EU Directives is available to any type of beneficiaries for any type of use; however, it can be explicitly reserved by rightsholders, meaning it may be overridden through mechanisms such as "opt-out" or "contract-out" clauses. Therefore, while the EU is headed in a promising future for text and data mining, which can be key for genAI, the fact that some rightsholders might opt out somewhat hinders the advancement of genAI. Thus, the use of work where a rightsholder has opted out might result in infringement liability imputed on the genAI programmers.

In addition, the EU Council approved an Artificial Intelligence Act aiming to harmonise rules on artificial intelligence.[58] In terms of section 107 of the EU AI Act:

> "In order to increase transparency on the data that is used in the pre-training and training of general-purpose AI models, including text and data protected by copyright law, it is adequate that providers of such models draw up and make publicly available a sufficiently detailed summary of the content used for training the general-purpose model."[59]

Therefore, it will be a requirement to publish summaries of copyrighted material used in the training of genAI. This allows copyright holders to make use of the required summaries to enforce their rights. In addition, the need for a summary will lead to companies using authorised material. This is a giant step in ensuring compliance with copyright protection.

4.3.4. United States of America

Copyright infringement in the US occurs when anyone violates the exclusive rights of the

---

[57] Artha Dermawan 'Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?' (2023) *Journal of World Intellectual Property Law* 51.

[58] Council of the European Union "Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI" available at *https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/,* accessed on 22 May 2024.

[59] European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), available at *https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf,* accessed on 22 May 2024.

copyright owner or the author or imports copies or phonorecords into the US.[60] The exclusive rights in §106 of the grant the copyright holder the sole permission to exercise exclusive rights or authorise another to exercise them, for example, reproduction. This means that in the US, the exercise of exclusive rights without permission would amount to copyright infringement. This means that the reproduction of works at the training stage of genAI would, in a similar manner, amount to infringement.

However, the US Copyright Act incorporates the doctrine of fair use. In terms of §107 of the US Copyright Act, "the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright." The use of the word "including" suggests that the use of the works is but not limited to the ones listed in this provision. This is a different approach to the South African doctrine of fair dealing which has a closed list which does not permit gravitating towards outside the listed uses.

§107 of the US Copyright Act provides tightens screws when it provides a test of whether certain uses fall within fair use. It further provides that:

> "In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include— (1) the purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work."

An interesting factor is commercial use. One relies on *Google LLC v. Oracle Am., Inc*,[61] where the Supreme Court stated that "there is no doubt that a finding that copying was not commercial in nature tips the scales in favour of fair use. But the inverse is not necessarily true, as many common fair uses are indisputably commercial." The Supreme Court gave an example of new reporting, which is done for commercial purposes but falls under fair use. Thus, genAI

[60] §501 of the Copyright Act, 17 U.S.C.
[61] *Google LLC v. Oracle America, Inc*. 593 U. S. (2021), at 27.

companies, most of whom are in court in the US,[62] will have to be determined on a case-by-case basis by the court. Until a solid precedent is set, the factors determining fair use are instrumental in informing whether cases of infringement will surface.

4.3.5. Canada

In section 27(1) of the Copyright Act,[63] "it is an infringement of copyright for any person to do, without the consent of the owner of the copyright, anything that by this Act only the owner of the copyright has the right to do." Therefore, where one engages in acts that are reserved for the copyright owner without being the owner or without consent thereof. However, the Canadian copyright regime incorporates an exception of fair dealing that absolves one from copyright infringement when making use of the work for the purpose of research, private study, education, parody or satire.[64] The doctrine of fair dealing offers exceptions to limited circumstances which are listed in the Act. In *CCH Canadian Ltd. v. Law Society of Upper Canada*,[65] the Supreme Court of Canada held that "research must be given a large and liberal interpretation in order to ensure that users" rights are not unduly constrained and is not limited to non-commercial or private contexts. The training of genAI involves copying of existing works, and to qualify such as research appears to be a daunting task.

Interesting to note is section 30.71(a)-(c) of the Copyright Act which provides that "it is not an infringement of copyright to make a reproduction of a work or other subject-matter if (a) the reproduction forms an essential part of a technological process; (b) the reproduction's only purpose is to facilitate a use that is not an infringement of copyright; and (c) the reproduction exists only for the duration of the technological process." The internal limitation, such as the reproduction existing during the technological process, makes it difficult to seek refuge under this exception when training genAI. This is because the datasets used to train are the fibre and being of genAI. The so-called AI autonomy is hinged on works it is trained upon. Thus, it is not a walk

---

[62] See*: Authors Guild v. OpenAI, Inc.,* No. 1-23-cv-08292 Document 39 (S.D.N.Y. 2023); *Tremblay, et al., v OpenAI, Inc., et al.,* No. 3:23-cv-03223 Document 1 (N. D. Cal. 2023); *Concord Music Group, Inc v Anthropic PBC*, No. 3-23-cv-01092 Document 1 (M.D. Tenn. 2023); *Getty Images (US) Inc v Stability AI Inc* No. 1:23-cv-00135 Document 1 (D. Del. 2024); *Andersen et al v. Stability AI Ltd. et al* No. 3:23-cv-00201-WHO Document 117 (N. D. Cal. 2023); *Huckabee et al v. Bloomberg L.P. et a*l, No. 1:23-cv-09152 Document 1 (S.D.N.Y. 2023);*Google LLC v. Oracle America, Inc*. 593 U. S. (2021).

[63] Copyright Act (R.S.C., 1985, c. C-42).

[64] s29 of the Copyright Act (R.S.C., 1985, c. C-42).

[65] *CCH Canadian Ltd. v. Law Society of Upper Canada*, 2004 1 S.C.R, at 342.

in the park, but relying solely on this provision to justify genAI training on data as non-infirming proves to be grim. In any event, the exception applies to the use of the work for purpose of research, private study, education, parody or satire.

4.4. Assigning Responsibility: Imputing Liability and Implications

In this section, the author examines the assigning of liability for copyright infringement in training genAI and determines the implications that accompany such assignments. The author applies the legal framework of the South African copyright ecosystem in efforts to find suitable candidate(s) for copyright infringement liability. The candidate(s) under consideration include the genAI company/programmer and the genAI model. The genAI user is of no significance at the training phase, given that they are the user and not involved in the training process.

4.4.1. AI company/programmer

First, the possible candidate to impute copyright infringement liability on is the genAI programmer. In terms of section 23(1) of the Copyright Act, direct infringement occurs when "any person, not being the owner of the copyright, who, without the licence of such owner, does or causes any other person to do, in the Republic, any act which the owner has the exclusive rights to do or to authorize." It has been shown above that the training phase risks resulting in "reproduction" and "making an adaption" of works without permission. Therefore, at the training phase, the person responsible for infringement is a genAI programmer or company. This is premised on the fact that the programmer engaged in reproduction and made adaptions of the works without consent from the copyright holder. Hence, in terms of the South African law, the responsibility for copyright infringement will be imputed on the genAI programmer.

However, it must be noted that the genAI programmer is responsible when they are self-employed or working alone. Where the programmer is a mere computer engineering genius working for a technology company, responsibility must be imputed on the company. This is premised on two grounds, first (i) the direct infringement per section 23(1) of the Copyright Act; and (ii) the common law doctrine of vicarious liability. In relation to section 23(1) of the Copyright Act, a copyright can be infringed by any person who, without permission, "causes any other person" to do any act which the owner has the exclusive rights to do or to authorise. This means that where a person causes another to exercise exclusive rights, such as making an adaption or

reproducing the works, such permission has infringed the copyright. Thus, since a programmer, who is an employee of the AI technology company, exercised the rights of the copyright holder, infringement must be imputed to the company.

In addition, the genAI programmer can be absolved from infringement and have it imputed on the company (employer) through the application of vicarious liability. In *K v Minister of Safety and Security,*[66] the Supreme Court of Appeal reflected that an employer will be vicariously liable for the delict of an employee if the delict is committed by the employee in the course and scope of his or her employment. The following requirements must be present in order for vicarious liability to be imposed on an employer:

> "(i) the person committing the wrongful act must be an employee of the employer; (ii) a third person must have suffered damages because of the unlawful act; and (iii) the act must have been committed in the course and scope of the wrongdoer's employment.[67] One can argue that, in order to impute liability on the technology company, an employment relationship must exist between the genAI programmer and the company."

This is generally the case since technology companies such as Midjourney Inc[68] have full-time employees programming a genAI model. In addition, the copyright holder (third person), must demonstrate damage suffered. Most importantly, the genAI programmer's act must be committed during the course and scope of employment. In the genAI training phase, this cannot be argued to be a hurdle since the reproducing and adapting of works is done within the scope of training genAI. Hence, the employer can have copyright infringed imputed to them genAI made was for a company.

The implications that come with holding the programmer and/or company liable for copyright infringement is that each party will be accountable and legally responsible when training genAI. Such accountability necessitates proactive measures to obtain consent from copyright holders, mitigating the risk of legal action. This will minimise the exploitation of the copyrighted works without any economic return for the copyright holders. However, this potentially stifles the

---

[66] *K v Minister of Safety and Security* 2005 (3) All SA 519 (SCA), para 4.
[67] Karin Calitz and Cecile de Villiers 'Sexual abuse of pupils by teachers in South African schools : the vicarious liability of education authorities' (2020) 137 *SALJ* 90.
[68] Midjourney 'About' available at *https://www.midjourney.com/home* accessed 1 March 2024.

development of highly effective genAI. This is because limiting genAI's access to data may compromise its accuracy and reliability. It has been argued that access to and use of archaic material in training often leads to the perpetuation of old-fashioned beliefs and attitudes.[69] This is because the data falling under open access often consists of old and archaic information, laden with values that contemporary society is actively trying to unlearn, such as racism and sexism, to name a few.

4.4.2. AI model

The next point of inquiry is whether copyright infringement during the training phase must be extended to the genAI model itself. Infringement occurs when a "person" does or causes exercise exclusive rights.[70] South African law recognises two categories of legal subjects namely human beings (natural persons) and juristic persons (artificial persons).[71] An example of a recognised juristic person in South Africa is a company which must be incorporated in terms of the Companies Act 71 of 2008. The use of the word "person" in section 23(1) of the Copyright Act implies that human beings and juristic persons (companies) are the ones who can be liable for infringement of a copyright. The word "artificial" used as a synonym of "juristic" must not be confused with genAI. Thus, in terms of the current law, liability cannot be imputed to genAI as it will be viewed as a legal object which people or companies are using. Hence, they are the ones who will be liable for infringement, as genAI is merely a tool which ingests data. There are theories suggesting that genAI must be recognised as a legal entity with rights and duties, but that is not the case in South Africa, and until such is done, genAI will remain beyond the unfeasible bearer of infringement liability. In the unlikely event that South Africa grants legal capacity to genAI, the implication is that there can be no source of practical remedy for the third party unless a clear legal framework is developed.

4.5. Conclusion

In conclusion, this chapter has explored copyright infringement at the input phase of genAI. It has emerged that under the current South African copyright regime, training genAI on copyrighted works will result in infringement. The discussion has brought to the fore that the suitable candidate

---

[69] Ravindra Kumar & Pankaj Kumar 'Training ai and copyright infringement: where does the law stand?' (2022) 2 *Indian Journal of Integrated Research in Law* 1-14.
[70] S23(1) of the Copyright Act.
[71] Jacqueline Heaton *The South African Law of Persons* 5 ed (2017) 3.

for copyright infringement liability at the training stage of genAI in South Africa is the company-that develops and deploys a genAI. The author further took an academy journey to jurisdictions beyond South Africa and found that Japan and the EU are more likely to be a breeding space for genAI training because they have flexible regulations. Copyright regimes in the US, UK, and Canada appear to be a rocky environment for genAI training firms, which are likely to be found in violation of copyright infringement liability. While there are limitations in decided cases in these jurisdictions on genAI and its training datasets, this research has demonstrated South Africa is not the only country legging behind in the AI race.

CHAPTER 5: IMPUTING LIABILITY FOR COPYRIGHT INFRINGEMENT – GenAI OUTPUT PHASE

5.1. Introduction

This chapter will examine the assignment of copyright infringement liability at the output phase through the lenses of the current South African copyright regime. It will scrutinise potential candidates, namely the AI programmer/company, the AI model, and the AI user, upon whose shoulders liability may be imputed. Furthermore, the implications of holding each party liable will be highlighted. In the spirit of clarity, the chapter will first explain how genAI-generated outputs pose a risk of infringing copyright before exploring possible candidates. Most importantly, it will draw insights from jurisdictions presently confronting copyright and infringement issues in AI creativity's output phase.

5.2. How GenAI-generated outputs pose a risk of infringing existing copyright

Copyright infringement is not novel to jurisprudence in South Africa. It is essential to bear in mind that one of the ways in which copyright infringement materialises is when a person does or causes another to exercise the exclusive rights reserved for the copyright holder without permission.[1] In order to establish copyright infringements, jurisprudence is instructive. In *Moneyweb (Pty) Limited v Media 24 Limited and Another*[2] Berger AJ applied the copyright infringement test from *Galago Publishers (Pty) Ltd v Erasmus*[3] while dealing with reproduction. Berger AJ reiterated that in infringement proceedings, it is not imperative for the plaintiff to prove the reproduction of the entire work, but rather, it suffices if a substantial part of the work has been reproduced.[4] The learned judge further stated that to "reproduce" within the meaning of the Copyright Act means to copy. Therefore, for genAI-generated outputs to infringe existing copyright, one does not have to prove the reproduction of the entire work as a substantial portion suffices.

In relation to the test to be applied when determining whether there has been copyright infringement *Moneyweb (Pty) Limited v Media 24 Limited and Another* remains instructive. *In casu,* Berger AJ stated:

---

[1] s23(1) of the Copyright Act 98 of 1978.
[2] *Moneyweb (Pty) Limited v Media 24 Limited and Another* 2016 (4) SA 591 (GJ), para 79.
[3] *Galago Publishers (Pty) Ltd & Another v Erasmus* 1989 (1) All SA 431 (A).
[4] *Moneyweb (Pty) Limited v Media 24 Limited and Another* 2016 (4) SA 591 (GJ), para 79.

> "[I]n order for there to have been an infringement of the copyright in an original work it must be shown (i) that there is sufficient objective similarity between the alleged infringing work and the original work, or a substantial part thereof, for the former to be properly described, not necessarily as identical with, but as a reproduction or copy of the latter; and (ii) that the original work was the source from which the alleged infringing work was derived, i.e. that there is a causal connection between the original work and the alleged infringing work."[5]

Against this backdrop, for genAI outputs to be found infringing existing copyright, one must demonstrate, first, that the genAI-generated work is objectively similar to the original work and, second, that the original work was the source of the alleged infringing work.

As stated above, in proving "objective similarity," the plaintiff must demonstrate that there is sufficient objective similarity between the alleged infringing work and the original work or a substantial part. It is essential to note that the work does not need to be identical.[6] What this entails is that a party that wishes to claim copyright infringement must demonstrate that the original work, which they have copyright over, is objectively similar to the alleged infringing work. In *Haupt t/a Softcopy v Brewers Marketing Intelligence*,[7] the court stated that whether a substantial part has been produced depends more on the quality than quantity of the reproduced. There is no blanket- approach test to objective similarity for genAI output, but determination must be made on a case-by-case basis. An interesting case is the *New York Times v MS and OpenAI*[8] where exhibit J demonstrated ChatGPT 4 reproducing verbatim, a publication series done by the New York Times in 2019. While this case is still under deliberation in the USA, it demonstrates clearly that genAI outputs can sometimes resemble the original works. Thus, genAI output, which has been subjected to substantial change, may escape infringement when one fails to establish objective similarity.

As mentioned above, the plaintiff must further establish a causal connection between the original work and the alleged infringing work. This means that the original work was the source from which the alleged infringing work was derived. Chapter Four dealt with the ingestion of large volumes of data by genAI during the training phase. This means that one must show that

---

[5] *Moneyweb (Pty) Limited v Media 24 Limited and Another* 2016 (4) SA 591 (GJ), para 79.
[6] *Ibid.*
[7] *Haupt t/a Softcopy v Brewers Marketing Intelligence* 2006 (4) SA 458 (SCA) para 44.
[8] *New York Times v Microsoft et al,* No 1:23-cv-11195 Document 1 (S.D.N.Y. 2023), at 30.

the genAI producing works similar to protected work had access to or derived its creativity from datasets containing the tokens or data of the protected works. This stage must appear easy, but it can prove to be difficult and impractical in the genAI sphere because genAI users or any other person besides the company can have access to the training datasets. In the event that datasets are concealed from the general public, most copyright holders will stagger, proving that the alleged infirming work was derived from the original protected work. Until parties can access training datasets, it remains a mountainous task to prove the causal connection.

5.3. How would other jurisdiction(s) address copyright infringement at the output phase?

5.3.1. China

One of the notable jurisdictions in the genAI field is China, and its deliberation will be valuable. Article 48(1)-(8) of the Copyright Law of the People's Republic of China[9] provides that one incurs civil liability or can be criminally prosecuted when engaged in acts such as reproduction without the permission of the copyright owner and such does not fall within the permitted parameters. This means that China, like most states, outlaws engaging in acts that infringe copyrights unless where the law permits. In a ruling directly involving AI, the Guangzhou Internet Court found an unnamed company to have committed copyright infringement for providing AI-generated text-to-image services.[10] At the core of the case lies the issue of reproduction. According to reports, a user requested Ultraman-related,[11] a well-known character owned by Tsuburaya Productions, from the AI platform, and the generated outputs were extremely similar to the plaintiff's original work.[12] This demonstrates that copyright infringement materialises when the generated work is substantially similar to the original work. The court found the AI company to be liable and ordered it to pay and awarded 10,000 yuan ($1,400) in damages.[13] Thus, this case is significant for; first, highlighting that copyright infringement occurs when the AI-

[9] Copyright Law of the People's Republic of China, 2010.
[10] Forbes 'China Rules AI Firm Committed Copyright Infringement' available at *https://www.forbes.com/sites/johannacostigan/2024/02/29/china-rules-ai-firm-committed-copyright-infringement/?sh=1b93132a454d,* accessed on 26 March 2024.
[11] Ultraman is a fictional superhero who would protect the Earth from monsters and alien threats. *See:* Ultraman Connection "Who is Ultraman" available at *https://www.ultramanconnection.com/pages/who-is-ultraman/,* accessed on 24 May 2024.
[12] *Supra* note 10.
[13] *Supra* note 10.

output is substantially similar to the original work; and second, the liability was imputed on the AI company and not the user. Thus, genAI companies are more likely to be held responsible than users.

5.3.2. United States of America

The US has a wealth of jurisprudence and laws dealing with copyright protection and infringement. In determining copyright infringement, *Tremblay et al v OpenAI Inc*[14] is instructive. *In casu,* the US District Court Northern District of California provided guidance on copyright infringement stating that:

> "Copyright infringement requires that a plaintiff show (1) 'he owns as valid copyright' and (2) the defendant 'copied aspects of his work' … The second prong 'contains two separate components: 'copying' and 'unlawful appropriation.' ... Copying can be demonstrated either through direct evidence or by showing that the defendant had access to the plaintiff's work and that the two works share similarities probative of copying, while the hallmark of 'unlawful appropriation' is that the works share substantial similarities."

It logically follows that ownership of copyright is the starting point. Upon satisfying that, one must show that the defendant copied the protected work. Copying and unlawful appropriation demand further explanation. The case held that one can demonstrate copying through direct access or showing that the alleged infringer had access to the copyright holder's work. However, it appears to be impractical because rarely can a plaintiff access data set used to train genAI. Most importantly, unlawful appropriation ought to be demonstrated by showing that the alleged infringing work and the original work share substantial similarities. Therefore, claims against genAI for copyright infringement at the output phase will possibly be successful where substantial similarity exists between the original and the genAI-generated work.

In addition, where the AI company is absolved of direct infringement, the vicarious liability trap might find them liable. Reliance can be placed on *Tremblay et al v OpenAI Inc,*[15] which states that:

> "Secondary liability for copyright infringement does not exist in the absence of direct

[14] *Tremblay et al v. OpenAI, Inc. et al* No 3:23-cv-3223-AMO Document 104 (N. D. Cal. 2024), at 4.
[15] *Ibid.*

> infringement by a third party". A plaintiff must then show that "the defendant has (1) the right and ability to supervise the infringing conduct and (2) a direct financial interest in the infringing activity."

This extract from the Trembley case means that in order for secondary infringement to materialise, the AI company must have the right and ability to supervise infringement. Supervision of genAI could get complicated as "certain Al is designed to rewrite their code to improve performance."[16] In relation to direct financial interest, the question can be blurry as most companies have a financial interest in the use of genAI. However, *Metro-Goldwyn-Mayer Studios v Grokster*[17] highlights that liability will be imputed where the objective of the distribution of a device in question was to promote its use to infringe copyright. Therefore, the issue of vicarious liability will be analysed on a case-by-case basis.

5.3.3. Canada

Section 27(1) of the Copyright Act,[18] provides that "it is an infringement of copyright for any person to do, without the consent of the owner of the copyright, anything that by this Act only the owner of the copyright has the right to do." Section 3(1) of the Copyright Act grants, among other rights, the right to reproduce the work or any substantial part copyright holder. In cases where one alleges that this right has been violated, reliance must be placed on case law. Hutchison[19] argues that in order to succeed in a breach of copyright "the plaintiff must prove two elements for substantial similarity infringement: (1) copying of, or access to, the work and (2) appropriation of the work—that is, all or a substantial part of the work has been taken." In *Cinar Corporation v Robinson*[20] the Supreme Court highlighted "the question of whether there has been substantial copying focuses on whether the copied features constitute a substantial part of the plaintiff's work - not whether they amount to a substantial part of the defendant's work."

The above means that in Canada, copyright infringement materialises where a substantial part has been taken. In assessing this, the court suggested that the focus is on whether the copied

---

[16] Zack Naqvi 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 *Marquette Intellectual Property Law Review* 20.
[17] *Metro-Goldwyn-Mayer Studios v Grokster,* 545 U.S. 913, 919 (2005).
[18] Copyright Act (R.S.C., 1985, c. C-42).
[19] Cameron J. Hutchison 'Substantial Similarity after Cinar Corp v Robinson' (2015) 31 *Canadian Intellectual Property Review* at 6.
[20] *Cinar Corporation v Robinson* 2012 (3) S.C.R 55 (SCC), para 39.

feature constitutes a substantial part of the plaintiff's work and not the inverse. In relation to genAI, based on the *Cinear Corporation v Robinson* analysis, the courts are more likely to assess whether the reproduced material constitutes a substantial part of the original work and not whether copied features amount to a substantial part of the genAI outputs. Therefore, the focus will be on the features reproduced.

5.3.4. United Kingdom

In the UK copyright law grants copyright holders exclusive rights such as copying and performing works.[21] Section 16(2) of the Copyright, Designs and Patents Act provides that "copyright in a work is infringed by a person who without the licence of the copyright owner does, or authorises another to do, any of the acts restricted by the copyright." In addition, section 16(3) of the Copyright, Designs and Patents Act provides that:

> "References … to the doing of an act restricted by the copyright in a work are to the doing of it - (a) in relation to the work as a whole or any substantial part of it, and (b) either directly or indirectly; and it is immaterial whether any intervening acts themselves infringe copyright."

In *Sheeran v Chokri*[22] the High Court of Justice highlighted that "to amount to an infringement, however, the copying must be of either the original work or a 'substantial part' of it… This is a qualitative, not quantitative question." In relation to genAI outputs, the plaintiff must demonstrate that the output was a copying of the original work which he/she owns, or a substantial part was copied. The case further highlights that "the test is whether the part in question contains elements which are the expression of the intellectual creation of the author of the work."[23] The author can extract from this case that the question of copying, which infringes upon the copyright, is a matter of fact. There are no hard and fast rules that dictate where the outcome will gravitate. Courts are more likely to determine that copyright infringement occurred when genAI output would have reproduced elements of the original work.

In addition, the case highlights that "an essential consideration is to ask whether a defendant has taken that which conferred originality on the claimant's copyright work (or a

[21] s16(1) of the Copyright, Designs and Patents Act 1988.
[22] *Sheeran v Chokri* 2022 EWHC 827 (CH), para 21.
[23] *Ibid*.

substantial part of it."[24] This underscores that when dealing with genAI outputs, the UK courts will assess whether elements that bestowed originality on the original work were taken and utilised by the alleged infringer in making the infringing work. It suffices to argue that this test depends on the comparisons of the original and alleged infringing work.

5.3.5. European Union

In terms of article 2 of the EU Information Society Directive[25] "Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part for, among others, authors, of their works." This means that the right to reproduce in whole or in part lies in the authors of the works. In relation to genAI, the author argues that a case of copyright infringement would materialise where a genAI output is the reproduction of the protected work without permission from the copyright holder.

5.4. Assigning Responsibility: Imputing Liability for Copyright Infringement and Implications

Once the potential for copyright infringement materialises at the genAI-output stage, the author shifts focus to determine liability among various parties. Candidates under scrutiny include the AI model itself, the AI programmer, and the AI user. It is noteworthy that liability for copyright infringement cannot be confined to one candidate, but depending on the merits of each case, multiple parties may potentially bear responsibility for infringement.

5.4.1. AI model

One might be tempted to argue that the AI model itself must be held responsible for copyright infringement. As explained above,[26] South Africa does not currently extend legal personhood or rights to entities such as AI and robots. This is because the rights recognised are that of a natural person and a juristic person, which is commonly the company. In addition, it has been argued that as of 2021, no major jurisdictions, such as the US, EU and UK, have recognised AI or robots as entities possessing legal personality.[27] Therefore, it is not only in South Africa where AI is

[24] *Ibid.*
[25] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001.
[26] *See above*: 4.4.2. "AI model"
[27] Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022) 12.

deprived of legal personhood. This logically follows that where AI as an entity cannot be a legal person, it cannot bear liability. In addition, AI is viewed as a tool and consequences that arise from its use are imputed on persons with legal capacity to be sued, such as a user or the company that developed that AI. Therefore, the assertion that AI models must bear responsibility for copyright infringement is flawed and will fail.

5.4.2. AI company/programmer

Owing to the failure to hold AI models liable since they are mere legal objects (tools), the author examines the possibility of holding the genAI company/programmer liable for copyright infringement. GenAI programmer and company carry two distinct meanings in this context: on the one hand, the former means the AI programmer is an unemployed solo genAI coder, and on the other hand, the latter means the genAI is developed by a company such as OpenAI with several programmers employed to create a genAI model. These distinctions have been highlighted earlier.[28] In terms of section 23(1) of the Copyright Act, direct infringement occurs when "any person, not being the owner of the copyright, who, without the licence of such owner, does or causes any other person to do, in the Republic, any act which the owner has the exclusive rights to do or to authorise." In the interest of brevity, where a genAI programmer is the sole AI programmer, an unemployed person, liability will be imputed to that one person. However, where an AI company employs a person, that company becomes liable should infringement materialise under the doctrine of vicarious liability.

Copyright infringement occurs when "any person, not being the owner of the copyright, who, without the licence of such owner, does or causes any other person to do… any act which the owner has the exclusive rights to do or to authorise."[29] It is fair to argue that at the AI output phase, the AI company/programmer does not have direct liability attached to doing any of the exclusive acts reserved for the copyright holder - this is more likely in the training phase of AI. However, a noteworthy conduit that may lead to holding the AI company/programmer liable at the output phase rooted on "causes any other person to do" any act which the owner has the exclusive rights to do or to authorise."[30] This is because while the user might exercise exclusive rights, where an

[28] *See above* section 4.3.1.
[29] s23(1) of the Copyright Act.
[30] s23(1) of the Copyright Act.

AI programmer causes that user to infringe copyright, liability will be extended to the AI developer. In this case, the plaintiff must demonstrate that the defendant caused another person to infringe copyright.

In the US, a party can be held liable for copyright infringement where a device was engineered to encourage copyright infringement. In *Metro-Goldwyn-Mayer Studios v Grokster*, the Supreme Court of the US stated that:

> "One who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties."[31]

This means that where an AI company would be shown to have developed a genAI model for the sake of promoting infringement, such a company will be liable for infringement by users of that genAI model. In South Africa, an AI that causes another person to engage in exclusive rights will be held liable because they have made it possible for another to infringe on protected works. Therefore, it goes down to what the AI program intended to serve and whether the AI company/ programmer has been directly or indirectly promoting infringement in efforts to attract users. While holding AI company/programmer liable seems a better option, it might impact the advancement of AI in South Africa or force AI firms to relocate to other jurisdictions that are conducive for AI firms.

5.4.3. AI user

A genAI user is another candidate to bear liability for copyright infringement. Given that infringement materialises when unauthorised person does or causes to do any act reserved for copyright holder,[32] an AI as a user will be liable where he/she does acts reserved for the copyright holder, such as reproduction.

In the event that a genAI user satisfies the test for infringement explored above, the presumption is that they will be directly liable for copyright infringement. This is premised on the fact that a genAI user reproduced a copyrighted work without permission from the copyright

[31] *MGM Studios Inc. v. Grokster, Ltd.* - 545 U.S. 913, 125 S. Ct. 2764 (2005), para 48.
[32] s23(1) of the Copyright Act 98 of 1978.

holder. AI users input prompts to trigger responses from genAI which emits outputs as prompted.[33] The author argues that users are responsible for what the genAI will have produced because they triggered it through prompts or were explicit in what they commanded the genAI to produce. Therefore, genAI users must bear responsibility for outputs that infringe on protected works. However, each case must be examined on its merits.

It has been argued that holding genAI users liable could be unfair when they are not necessarily aware that infringement may occur, and they are unsophisticated people.[34] An interesting argument was made by Enrico Bonadio, Plamen Dinev, and Luke McDonagh, who submit that should "programmers, owners or users of genAI systems want to claim exclusive rights over the final outputs generated by their machines, they should also accept responsibility for those works, in relation not only to copyright infringement."[35] Therefore, while in South Africa, a user is likely to bear responsibility for infringement because they prompted the genAI, imputing liability on those who claim ownership over the works will be another way to solve this problem and discourage the generation of infringing works since there will be no credit or cash obtained from such generated works. Caution must be exercised when holding the users liable because it might result in a decrease in the use of genAI and harm its development.

5.1. Conclusion

In summation, this chapter has examined copyright infringement and liability at the output phase of genAI. It has been established that genAI poses a risk of copyright infringement where objective similarity exists between a protected work and genAI-generated output. Jurisdictions explored differ in terms of the substance of the law, but generally, genAI at the output phase can be found to be in breach of copyright where it is similar to a protected work. Regarding the assignment of copyright infringement liability, the author has noted that genAI models cannot bear responsibility due to their lack of legal personhood. However, an AI company or programmer are the most likely candidates to be imputed with copyright infringement liability. It has been noted that holding genAI companies and programmers liable for the generated output might stifle the advancement

[33] *Tremblay, et al., v OpenAI, Inc., et al*., No. 3:23-cv-03223 Document 1 (N. D. Cal. 2023), at 1.
[34] Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022) 13.
[35] *Ibid.*

of AI as firms fear lawsuits. In the same vein, holding users liable will discourage AI use, harming the AI market.

CHAPTER 6: RECOMMENDATIONS AND CONCLUSION

6.1. Introduction

This concluding chapter summarises the findings and offers recommendations in relation to the cardinal question, "How can South Africa address copyright infringement involving genAI? Should responsibility for copyright infringement be attributed to the AI company/programmer, the users of the AI, or the AI model itself, and what are the implications of holding either party liable for such infringement?" This chapter will first provide a summary of findings about how the current copyright law addresses copyright infringement and the assignment of responsibility. Subsequently, the author makes recommendations for South Africa.

6.2. Summary and Findings

This dissertation focused mainly on whether copyright infringement would materialise during the training phase of genAI and the genAI output phase. The training phase relates to the stage at which genAI is developed and trained prior to deployment. The output phase relates to the stage when genAI is deployed and in circulation in the market. Noteworthy research sub-questions were probed and summarised below in accordance with South African copyright law.

6.2.1. Whether the use of copyrighted material during the genAI-training phase amounts to copyright infringement?

The dissertation examined whether the use of copyrighted material during the training of genAI amounts to copyright infringement. In addressing this question, it found that using copyrighted material during the training phase without permission from copyright holders amounts to copyright infringement by violating reproduction and adaptation rights. It arrived at this conclusion based on two premises in relation to reproduction. One of the premises is that reproduction may occur when the same input datasets are copied multiple times during the training of genAI. The additional premise is that the work that is ingested by genAI during the training phase amounts to "valuable ingredients." This is because genAI responds to prompts by emitting training datasets.

Copyright infringement may materialise under violation of adaptation rights based on the premise that AI companies/programmers may transform, modify or alter original works for several reasons, for example, encoding images and adding visual noise by Stability AI. This means the works are adapted to suit particular needs. Therefore, using copyrighted works to train genAI without

consent results in copyright infringement in South Africa, where the copyright holder's consent is absent. Most importantly, the fair dealing doctrine cannot insulate the genAI training process from breaching copyright infringement because genAI training strays beyond the confines of fair dealing, such as "research or private study and personal or private use" in terms of the Copyright Act.

6.2.2. Who is responsible for copyright infringement during the genAI training phase?

The dissertation further examined who is responsible for copyright infringement that materialises during the genAI training phase. In response to this question, the findings indicate that the AI programmer/company is responsible for copyright infringement liability. An AI programmer is liable when working individually without offering services to a technology company. However, where a programmer is an employee, their employer (company) becomes liable based on: (i) the direct infringement per section 23(1) of the Copyright Act and (ii) the common law doctrine of vicarious liability. Direct infringement occurs when an unauthorised does or causes another exercise rights reserved for the copyright holders. In addition, vicarious liability imputes liability on the employer because an employee commits a delict in the course and scope of his or her employment. Further, attempts to impute liability on the AI model for copyright infringement will fail in because South Africa does not extend legal personhood to AI models.

6.2.3. Whether the genAI-generated works amount to copyright infringement?

Furthermore, the dissertation examined whether genAI-generated works may infringe copyright. The answer to this key question is that it is possible for copyright infringement to occur at the output phase of generative genAI. However, in order for copyright infringement to occur, the inherent factors in the test must be met. This means that copyright infringement at the output phase will occur where (i) there is sufficient objective similarity between the alleged infringing work and the original work and (ii) a causal connection between the original work and the alleged infringing work must exist. Therefore, this question is destitute of a definite answer, but where one satisfies the test, work generated by genAI amounts to copyright infringement. It is worth stating that copyright infringement will occur when the work generated is objectively similar, not identical, work.

6.2.4. Who is liable for copyright infringement where genAI-generated works infringe copyright?

The dissertation examined who is liable for copyright infringement liability in relation to work

generated by genAI. In response to this question, it found that the genAI user and genAI company/programmer are the most suitable candidates to bear liability. The argument in support of finding the user liable is premised on the fact that he/she is the one who prompts genAI to produce an infringing work, this amounts to direct infringement. In addition, a genAI company/programmer will be liable where because they "caused another" to exercise the exclusive right to reproduction. In addition, the genAI companies that develop genAI software that encourages or commercialises the generation of infringing works will be liable. In addition, efforts to impute liability on genAI models are destined to fail because AI models lack legal personality and are mere legal objects under South African law.

6.2.5. What are the implications of holding each party responsible?

The dissertation went a step further to examine the implications of holding AI companies or programmers and genAI users liable for copyright infringement. It noted that holding an AI company/programmer liable for copyright infringement may negatively impact the advancement of genAI in South Africa or deter AI firms from expanding in South Africa. Furthermore, finding AI companies liable for copyright infringement at the training phase will discourage the use of more updated works, which may result in unreliable genAI or genAI that is outdated given that it has access to limited archaic works in the public domain, some of which may contain values that the society has evolved from such as racism and sexism. In addition, finding users liable may discourage the use of AI which harms the market and stifle AI's advancement

6.3. Charting the Path Forward: Recommendations for South Africa

While the author undertook to examine copyright infringement stemming from genAI, their aim extends to arming South Africa with recommendations that assuage conflicts associated with AI-related copyright infringement. In the paragraphs below, the author provides recommendations that foster an environment conducive to AI advancement while safeguarding the interests of copyright holders against unfair treatment. In order to create a thriving society, the implementation of these recommendations must be a collective effort by various stakeholders, including the government, businesses, copyright holders, and AI users.

6.3.1. Enacting AI Legislation

One of the ways to address genAI-related copyright infringement is by introducing AI legislation,

alternatively by amending existing copyright legislation. AI legislation will legislate issues that are related to AI, including copyright. The legislation will address issues ranging from the use, training, deployment, and liability attached to AI and the responsibilities of the parties involved. The carpet of AI regulation is unfolding across the globe, with several jurisdictions implementing AI legislation and bills, such as the EU[1] and Brazil.[2] However, South Africa must not simply transplant the EU or Brazilian approach to AI. The legislature must develop laws that address challenges unique to South Africa and its people. In addition, South Africa can amend existing copyright laws, for example, by introducing *a sui generis* exception for the training of genAI. A *sui generis* exception would, for example, provide that using data for the purpose of training genAI is not a breach of copyright. However, fairness factors can be incorporated, for example, the type of work, the impact of the training of copyright owners' market, and the annual turnover of the AI company. Alternatively, the legislature can amend the fair dealing exception to allow the use of data for the purpose of training genAI.

6.3.2. Licensing and Royalties

In efforts to mitigate conflict related to AI copyright infringement, licensing and royalties can be a crucial step forward. AI companies/programmers must obtain permission from copyright holders through licenses and royalties. Copyright holders have the authority to license the use of their work and, in return, be compensated. This can be a progressive step for AI companies to avoid infringing copyright. It will ensure that the data included in datasets is licensed and curb infringements during the training of genAI. After genAI has been deployed, licenses can be extended to ensure that works that the genAI users are protected and negotiated royalties be paid to the copyright owners. Fundamental to the sustenance of this model is that genAI companies/programmers train their genAI with data that is licensed or in the public domain. However, since data is indiscriminately harvested, companies/programmers might struggle to obtain permission from every copyright holder through licenses because contacting owners is a challenge. Regardless of this challenge, AI companies/programmers must abstain from using

[1] European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), available at *https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf,* accessed on 22 May 2024.
[2] Bill PL 1465/2024, [Portuguese version] available at *https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2414736&filename=PL%201465/2024*, accessed on 22 May 2024.

works whose origin is unknown to avoid risking infringement.

6.3.3. Collecting Societies

Combatting genAI-related copyright infringement could require reliance on collecting societies. The collecting society model has been a feature of copyright law in South Africa.[3] Collecting societies negotiate, collect and distribute royalties to copyright owners.[4] Given that several copyright owners exist, and AI companies could struggle to obtain permission from each individual owner, collecting societies can be a solution. Companies and collecting societies will agree on royalties and ensure that the training data has been licensed. However, while being interviewed by the author, Craig holds that collecting societies are not a workable model because "payments are going to be minuscule and transaction costs will be huge."[5] To this end, this cannot achieve the copyright holder's goals, such as a living wage or adequate funding.[6] Despite its flaws, a collective society can ensure remuneration.

6.3.4. Introduction of an AI taxation model

A proactive approach to mitigating genAI-related copyright infringement involves the introduction of a tax on genAI that uses works whose origin cannot be determined, otherwise known as orphan works. Data harvesting tools indiscriminately copy billions of works online, and some works lack the details that identify the owner. Copyright Amendment Bill defines orphan work as "a work in which copyright subsists and the owner of a right in that work cannot be identified; or is identified, but cannot be located."[7] The new tax will be imposed on AI companies that use orphan works, and the proceeds will be given to the collecting society or designated office. An AI company must declare when it has utilised orphan works. Proceeds from copyrighted work with identifiable copyright owners and proceeds from orphan work tax must be separate. This is because the former has identifiable copyright owners to whom such proceeds can be channelled, but proceeds from orphan work lack a direct recipient. The proceeds can assist in managing societies and combating copyright infringement by genAI through raising awareness among creators, users, and AI company/programmer. France is one of the countries headed

[3] s1(1) of the Copyright Act.
[4] s1(1) of the Copyright Act.
[5] Carys Craig "The AI Copyright Trap" *iNtaka Centre for Law & Technology* note 27:12 (time mark), available at *https://www.youtube.com/watch?v=vX-c-YOKs_8,* accessed on 27 April 2024.
[6] *Ibid.*
[7] s1(1)(i) of the Copyright Amendment Bill B 13D—2017.

towards taxing AI companies that use orphan works and proceeds utilised by management systems.[8] In addition, Craig argues that taxing companies that deploy or sell AI models can protect creators because money can be redistributed or put into cultural funds.[9] This is a shift from taxing the use of orphan works only to all works.

6.3.5. Implementation of technological measures that prevent copyright infringement

One avenue for effectively addressing genAI-related copyright infringement is implementing technological measures that prevent copyright infringement. First, companies can minimise the generation of infringing outputs by checking infringement before releasing outputs prompted by a user. Tools can be used to complement genAI models that generate music with a plagiarism-checker tool. An example is NVIDIA's plagiarism checker, which detects whether a generated musical piece is partly or fully plagiarised from the database from which the genAI learnt.[10] In addition, AI companies/programmers can internally block prompts that refer to third-party works. The importance of blocking prompts stating third-party works is highlighted clearly by a Guangzhou Internet Court ruling fining a company that infringed copyright when a user requested Ultraman-related[11] from the AI platform, and the generated outputs were extremely similar to the plaintiff's original work.

Additionally, copyright holders can proactively protect their work by detecting the use of their work in training datasets or sharing of works online. An example of these tools includes a free app, Glaze, which is an academic research project launched by the University of Chicago for artists to combat the theft of works - scraped into datasets to train genAI tools designed to mimic visual style- via the application of a high-tech cloaking technique.[12] Existing notice and takedown procedures will complement technological measures recommended herein to capture works that

---

[8] Article 4 of the Proposal No. 1630 aimed at regulating artificial intelligence through copyright (*Proposition de loi n°1630 visant à encadrer l'intelligence artificielle par le droit d'auteur*) 2023.

[9] Dr. Carys Craig "The AI Copyright Trap" *iNtaka Centre for Law & Technology* note 28:08 available at *https://www.youtube.com/watch?v=vX-c-YOKs_8,* accessed on 27 April 2024.

[10] NVIDIA "AI with the Heart of a Composer - Aiva | Season 1 Episode 1 | I AM AI Docuseries" note 2:48 available at *https://www.youtube.com/watch?v=CPh0bKcXgLo,* accessed on 27 April 2024.

[11] Ultraman is a fictional superhero who would protect the Earth from monsters and alien threats. *See:* Ultraman Connection "Who is Ultraman" available at *https://www.ultramanconnection.com/pages/who-is-ultraman/,* accessed on 24 May 2024.

[12] Glaze "About The Glaze Project" available at *https://glaze.cs.uchicago.edu/aboutus.html*, accessed on 27 April 2024.

escape measures put in place by AI companies.

6.3.6. AI Industry Guidelines and Code of Conduct

Introducing industry guidelines and codes of conduct could curb risks of copyright infringement. AI companies can have industry guidelines and a set code of conduct which governs how members train, deploy, update and profit from genAI. A similar setup can be extracted from the Internet Service Providers' Association (ISPA), which represents the South African Internet Industry with a code of conduct.[13] GenAI model providers can have members who train genAI tools on licensed data or are subject to open source and have plagiarism checks and blocks in place. This will encourage genAI customers to purchase genAI products or models that companies deploy in a certain association. In addition, to avoid liability, customers must inquire from genAI companies whether models were trained with copyrighted material, assess the terms of service, and avoid genAI tools that cannot confirm that their training data is properly licensed by copyright owners or collecting societies.

6.4. Concluding Remarks

In summation, it is possible for copyright infringement to materialise during the training of genAI when copyrighted material is ingested without authorisation from copyright holders. In addition, the risk of copyright infringement at the output phase of genAI is heightened when the generated work is objectively similar to the original work. Furthermore, the author noted that AI companies are liable for copyright infringement during the training phase of genAI, while AI companies and end-users are the most suitable candidates to bear infringement liability during the output phase of AI. Most importantly, it has been highlighted that holding companies liable will negatively impact the development of quality AI and harm advancement at large. The dissertation argued that holding the AI users liable may discourage genAI use, harming the market.

Moreover, the author made recommendations that South Africa can implement or adopt to address copyright infringement triggered by genAI. These recommendations include adopting AI legislation, concluding licensing agreements between AI companies and copyright owners, using collecting societies, introducing tax on AI, utilising technological measures that prevent

[13] ISPA "Code of Conduct" available at *https://ispa.org.za/code-of-conduct/,* accessed on 29 April 2024.

copyright infringement, and adopting AI Industry Guidelines and Code of Conduct. These recommendations must be embraced as a joint effort by several stakeholders, such as governmental bodies, businesses, copyright owners, and AI user

BIBLIOGRAPHY

Primary Sources

Constitution

Constitution of the Republic of South Africa, 1996

Case Law

South Africa

Claasen and Another v TEC Novation Solution (Pty) Limited and Another (2017/40521) 2018 (ZAGPPHC).

Dexion Europe Ltd v Universal Storage Systems (Pty) Ltd 2002 JOL 10172 (SCA).

Galago Publishers (Pty) Ltd & Another v Erasmus 1989 (1) All SA 431 (A).

Haupt t/a Softcopy v Brewers Marketing Intelligence 2006 (4) SA 458 (SCA).

King v SA Weather Services 2009 (3) SA 13 (SCA).

K v Minister of Safety and Security 2005 (3) All SA 519 (SCA).

Southern African Music Rights Organisation Ltd v Svenmill Fabrics (Pty) Ltd 1983 (1) SA 608 (C).

Memory Institute SA CC t/a SA Memory Institute v Hansen and Others (253/2002) 2003 (ZASCA).

Moneyweb (Pty) Limited v Media 24 Limited and Another 2016 (3) All SA 193 (GJ).

Payer Components South Africa Ltd v Bovic Gaskins & Others 1995 (2) All SA 600 (A).

Foreign

Andersen et al v. Stability AI Ltd. et al No. 3:23-cv-00201-WHO Document 117 (N. D. Cal. 2023).

Authors Guild v. OpenAI, Inc., No. 1-23-cv-08292 Document 39 (S.D.N.Y. 2023).

Cinar Corporation v Robinson 2012 (3) S.C.R 55 (SCC).

CCH Canadian Ltd. v. Law Society of Upper Canada, 2004 1 S.C.R.

Concord Music Group, Inc v Anthropic PBC, No. 3-23-cv-01092 Document 1 (M.D. Tenn. 2023).

Getty Images (US) Inc v Stability AI Inc No. 1:23-cv-00135 Document 1 (D. Del. 2024).

Google LLC v. Oracle America, Inc. 593 U. S. (2021).

Huckabee et al v. Bloomberg L.P. et al, No. 1:23-cv-09152 Document 1 (S.D.N.Y. 2023).

Metro-Goldwyn-Mayer Studios Inc v Grokster Ltd, 545 U.S. 913 (2005).

New York Times v Microsoft et al, No 1:23-cv-11195 Document 1 (S.D.N.Y. 2023).

Tremblay, et al., v OpenAI, Inc., et al., No. 3:23-cv-03223 Document 1 (N. D. Cal. 2023).

Tremblay et al v. OpenAI, Inc. et al No 3:23-cv-3223-AMO Document 104 (N. D. Cal. 2024).

Sheeran v Chokri 2022 EWHC 827 (CH).

Legislation and Bills

South Africa

Copyright Act 98 of 1978.

Copyright Amendment Bill B 13D—2017.

Foreign

Copyright Act, 17 U.S.C.

Copyright, Designs and Patents Act 1988.

Copyright Law of the People's Republic of China, 2010.

Bill PL 1465/2024, available at *https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2414736&filename=PL%201465/2024,* accessed on 22 May 2024.

Proposal No. 1630 aimed at regulating artificial intelligence through copyright (*Proposition de loi n°1630 visant à encadrer l'intelligence artificielle par le droit d'auteur*) 2023.

Treaties and International Instruments

Agreement on Trade-Related Aspects of Intellectual Property Rights, 1994.

Berne Convention for the Protection of Literary and Artistic Works, 1979.

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 200.

European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), available at *https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf,* accessed on 22 May 2024.

Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled, 2013.

WIPO Copyright Treaty, 1996.

Secondary Sources

Books and Chapters

Dinev *et al* 'Can artificial intelligence infringe copyright? Some reflections' in Abbott & Geffen (eds) *Research Handbook on Intellectual Property and Artificial Intelligence* (2022).

Dariusz Szostek & Mariusz Załucki (eds) *Legal Tech* (2021).

Mike McCnville & Wing Hong Chui *Research Methods for Law* 2 ed (2017).

Jacqueline Heaton *The South African Law of Persons* 5 ed (2017).

James E Baker, Laurie N Hobart & Matthew Mittelsteadt *An Introduction to Artificial Intelligence for Federal Judges* (2023).

J Vesala & R M Ballardini 'AI and IPR Infringement: A Case Study on Training and Using Neural Networks' in R M Ballardini *et al* (eds) *Regulating Industrial Internet through IPR, Data Protection and Competition Law* (2019).

Owen Dean & Alison Dyer (eds) *Dean & Dyer Introduction to Intellectual Property Law* (2014).

Sylvia Papadopoulos & Sizwe Snail ka Mtuze *Cyberlaw @ SA IV: the law of the Internet in South Africa* 4 ed (2022).

Journal Articles

A M Turing 'Computing Machinery and Intelligence,' (1950) 49 *Mind* 433-460, available at *https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf, accessed 10* February 2024.

Artha Dermawan 'Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?' (2023) *Journal of World Intellectual Property Law.*

Andrés Guadamuz 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs' (2023) Available at SSRN: *http://dx.doi.org/10.2139/ssrn.4371204.*

Bridget Watson 'A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence' (2017) 58 *IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property*.

Cameron J. Hutchison 'Substantial Similarity after Cinar Corp v Robinson' (2015) 31 *Canadian*

Intellectual Property Review.

Daniel J Gervais 'The Machine as Author' (2019) 105 *Iowa Law Review* 2053-2106

Dimas Saputr & Azis Budianto 'Legal Analysis of Artificial Intelligence Responsibility in Copyright Infringement in the Digital Era' (2023) Proceedings of the 3rd *International Conference on Law, Social Science, Economics, and Education.*

Enrico Bonadio & Luke McDonagh 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity' (2020) 2 *Intellectual Property Quarterly.*

European Parliamentary Research Service *Civil liability regime for artificial intelligence* (2020) available at *https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf.*

Karin Calitz and Cecile de Villiers 'Sexual abuse of pupils by teachers in South African schools: the vicarious liability of education authorities' (2020) 137 *SALJ.*

Leonardo Banh & Gero Strobel 'Generative artificial intelligence' (2023) 63 *Electronic Markets.*

Lee-Ann Tong "Copyright Protection for Computer Programs in South Africa: Aspects of Generis Categorisation" (2009) *JWIP.*

Michael Haenlein & Andreas Kaplan 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) *California Management Review* 61(4) at 6 available at *https://journals.sagepub.com/doi/10.1177/0008125619864925.*

Mkhululi Nyathi 'Re-Asserting the Doctrinal Legal Research Methodology in the South African Academy: Navigating the Maze' (2023) 140 *SALJ.*

Ravindra Kumar & Pankaj Kumar 'Training ai and copyright infringement: where does the law stand?' (2022) 2 *Indian Journal of Integrated Research in Law*.

Raju Narayana Swamy 'Strong AI to Super-intelligence: How is AI placed vis-à-vis Intellectual Property Rights' (2021) 3 *International Journal of Computer Communication and Informatics.*

Rex Martinez 'Artificial Intelligence: Distinguishing Between Types & Definitions' (2019) 19 *Nevada Law Journal.*

Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch & Patrick Zschech 'Generative AI' (2023) *Business & Information Systems Engineering.*

Zack Naqvi 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 *Marquette Intellectual Property Law Review.*

Internet Sources

Apple 'Siri' available at *https://www.apple.com/siri/* accessed on 12 February 2024.

Anthropic 'Talk to Claude' available at *https://claude.ai/login?returnTo=%2F* accessed 27-02-2024.

Amazon 'Alexa features' available at *https://www.amazon.com/b?ie=UTF8&node=21576558011* accessed 12 February 2024.

Binder Dijker Otte South Africa "An Introduction to Artificial Intelligence and Machine Learning" available at *https://www.bdo.co.za/getattachment/Industries/Financial-Services/Emerging-Tech-Toolbox/Artificial-Intelligence-and-Machine-Learning/FS-AI-Whitepaper-design-July-2020.pdf* accessed 13 February 2024.

Bill PL 1465/2024, [Portuguese version] available at *https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2414736&filename=PL%201465/2024*, accessed on 22 May 2024.

Carys Craig "The AI Copyright Trap" *iNtaka Centre for Law & Technology* available at *https://www.youtube.com/watch?v=vX-c-YOKs_8,* accessed on 27 April 2024.

Council of the European Union "Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI" available at *https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/,* accessed on 22 May 2024.

Dusan Simic 'Mastering the Essentials of AI Training Data: A Comprehensive Guide' available at *https://www.linkedin.com/pulse/mastering-essentials-ai-training-data-comprehensive-guide-dusan-simic-fclxf/* accessed 12 February 2024.

Ellen Glover 'What Is AI Winter?' available at *https://builtin.com/artificial-intelligence/ai-winter* accessed accessed 10 February 2024.

Forbes "The Best Free AI Music Generators Of 2023" available at *https://www.forbes.com/sites/ariannajohnson/article/the-best-free-ai-music-generators/?sh=6a68e3484ffb,* accessed on 3 July 2023.

Forbes "Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The

Creative Economy" available at *https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=65ab1ad82d2b* accessed 20-02-2024.

Forbes 'China Rules AI Firm Committed Copyright Infringement' available at *https://www.forbes.com/sites/johannacostigan/2024/02/29/china-rules-ai-firm-committed-copyright-infringement/?sh=1b93132a454d,* accessed on 26 March 2024.

Gemini 'Gemini' available at *https://gemini.google.com/?hl=en* accessed 13 February 2024.

GitHub "The world's most widely adopted AI developer tool" available at *https://github.com/features/copilot* accessed 13 February 2024.

Glaze "About The Glaze Project" available at *https://glaze.cs.uchicago.edu/aboutus.html*, accessed on 27 April 2024.

IT Law Wiki 'Computational speed' available at *https://itlaw.fandom.com/wiki/Computational_speed#:~:text=Computational%20speed%20is%20the%20speed,Millions%20of%20Instructions%20Per%20Second* accessed 12 February 2024.

IBM "IBM Watson Assistant" available at https://www.ibm.com/docs/en/cloud-private/3.1.0?topic=services-watson-assistant accessed 12 February 2024.

ISPA "Code of Conduct" available at *https://ispa.org.za/code-of-conduct/,* accessed on 29 April 2024.

John Donegan 'The US should look at Japan's unique approach to generative AI copyright law' available at *https://insights.manageengine.com/artificial-intelligence/the-us-should-look-at-japans-unique-approach-to-generative-ai-copyright-law/#:~:text=Unlike%20the%20U.S.%20and%20E.U.,the%20gen%20AI%20arms%20race,* accessed accessed 14 March 2024.

Kathy Berry 'UK fails to agree AI/copyright code of practice,' available at *https://www.linklaters.com/en/insights/blogs/digilinks/2024/february/uk-fails-to-agree-ai---copyright-code-of-practice#:~:text=Under%20the%20current%20legal%20framework,their%20works%20by%20AI%20developers,* accessed 15 March 2024.

Midjourney 'About' available at *https://www.midjourney.com/home* accessed 13 February 2024.

Microsoft Azure 'What is cloud computing?' available at *https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing#:~:text=Simply%20put%2C%20cloud%20computing%20is,resources%2C%20and%20economies%20of%20scale* accessed 12 February 2024.

Michihiro Nishi 'Japanese Law Issues Surrounding Generative AI: ChatGPT, Bard and Beyond' available at *https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2023/10/Japanese-Law-Issues-Surrounding-Generative-AI.html*, accessed 14 March 2024.

NVIDIA "AI with the Heart of a Composer - Aiva | Season 1 Episode 1 | I AM AI Docuseries" note 2:48 available at *https://www.youtube.com/watch?v=CPh0bKcXgLo,* accessed on 27 April 2024.

OpenAI 'Introducing ChatGPT' available at *https://openai.com/blog/chatgpt* accessed 13 February 2024.

OpenAI 'Introducing ChatGPT' available at *https://openai.com/blog/chatgpt* accessed 13 February 2024.

OpenAI 'Pricing' available at *https://openai.com/chatgpt/pricing/* accessed 15 November 2024.

Sara Brown 'Machine learning, explained' available at *https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained* accessed 12 February 2024.

Sky News 'AI Drake and The Weeknd: Song called Heart On My Sleeve - made with cloned voices - removed from streaming services' available at *https://news.sky.com/story/ai-drake-and-the-weeknd-song-called-heart-on-my-sleeve-made-with-cloned-voices-removed-from-streaming-services-12859951*, accessed on 3 July 2023.

Super AI 'Where AI meets the world' available at *https://www.superai.com/* accessed 13 February 2024.

Shinnosuke Fukuoka, Tomonobu Murata and Atsuki Mizuguchi 'Legal Issues in Generative AI under Japanese Law-Copyright' available at *https://www.nishimura.com/sites/default/files/newsletters/file/robotics_ai_230711_en.pdf*, accessed 14 March 2024.

Scott Warren & Joseph Grasser 'Japan's New Draft Guidelines on AI and Copyright: Is It Really OK to Train AI Using Pirated Materials?' available at *https://www.privacyworld.blog/2024/03/japans-new-draft-guidelines-on-ai-and-copyright-*

is-it-really-ok-to-train-ai-using-pirated-materials/ accessed 14 March 2024.

Rachel Montagnon & Sungmin Cho 'UK withdraws plans for broader Text and Data Mining (TDM) copyright and database right exception' available at *https://hsfnotes.com/ip/2023/03/01/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception/,* accessed 15 March 2024.

Rockwell Anyoha 'The History of Artificial Intelligence' *Harvard Special Edition on Artificial Intelligence* 28 August 2017 available at *https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/* accessed 10 February 2024.

Ultraman Connection "Who is Ultraman" available at *https://www.ultramanconnection.com/pages/who-is-ultraman/,* accessed on 24 May 2024.

Vijay Kanade 'What Is Super Artificial Intelligence (AI)? Definition, Threats, and Trends' available at *https://www.spiceworks.com/tech/artificial-intelligence/articles/super-artificial-intelligence/* accessed 13 February 2024.