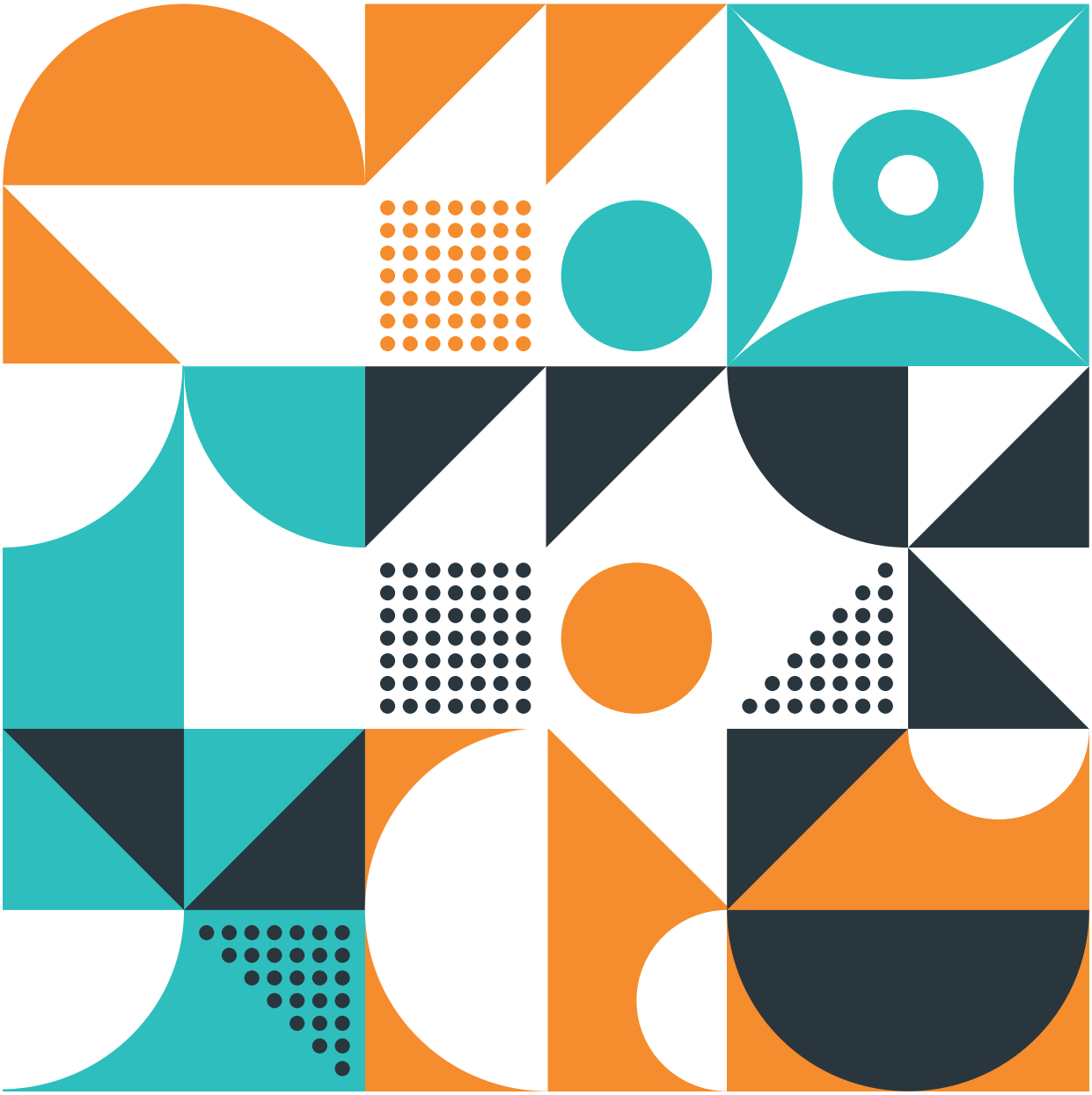
Computers & Law

Journal for the Australian and New Zealand Societies for Computers and the Law

ISSN 0811-7225

2022
Volume 94



Computers & Law

Journal for the Australian and New Zealand Societies for Computers and the Law

This issue may be cited as
(2022) 94 *Computers & Law*

ISSN 0811-7225

Published by AustLII Foundation Press
Printed by UTS Printing Services

Articles in this issue are also available via AustLII at:
<http://www.austlii.edu.au/au/journals/ANZCompuLaw/>.

All enquiries about this publication should be addressed to:

The Editors
Computers & Law
C/- Faculty of Law
University of Technology Sydney
PO Box 123
Ultimo NSW 2007
Australia
<editors@auscl.org>

Computers & Law

Journal for the Australian and New Zealand Societies for Computers and the Law

INFORMATION FOR CONTRIBUTORS

Computers & Law welcomes submissions that relate to contemporary legal and technical developments including but not limited to:

- Computational Law, Legal Informatics and Rules as Code
- Artificial Intelligence, Standards and Ethics
- Future Democracy
- Future Money and Digital Currency
- Future Cities, IOT and Surveillance
- Privacy, Big Data and Competition
- Intellectual Property and Technology Transfer
- Future Medicine
- Quantum Computing
- Cybersecurity

TYPES OF SUBMISSION

The following types of submission are welcomed:

- *Academic Article* submissions are to be approximately 5,000 to 7,000 words in length (excluding footnotes, which should not exceed 1000 words).
- *Thought Leadership Essays* are to be approximately 1,500 to 3000 words in length (excluding footnotes, which should not exceed 1000 words).
- *Industry Reports and Updates* are to be 500 to 1,000 words in length.

SUBMISSION GUIDELINES

Contributors grant to AustLII and AUSCL a non-exclusive, irrevocable, perpetual licence to publish their articles in any format and medium including but not limited to online and in print, as decided by AustLII from time to time.

Contributors, in preparing their submissions for publication in this Journal, are required to note the following:

1. The submission is made by a contributor for publication on the basis that:
   (a) The submission has not been previously published; and
   (b) The submission is an original work of the contributor.
2. The contributor is responsible for the accuracy of the submission including citations and references. Citations are required to follow the *Australian Guide to Legal Citation* ('AGLC') and should be in the form of footnotes. AGLC4 is available at <https://law.unimelb.edu.au/mulr/aglc/about>.
3. The following information must be included with the submission:
   (a) The contributor's name, title, affiliation, email address; and
   (b) An abstract of the submission (80-120 words). However, an abstract is not required for an Industry Report submission.

Please contact the editorial team at <editors@auscl.org> for more information.

Computers & Law

Journal for the Australian and New Zealand Societies for Computers and the Law

2022
VOLUME 94

CONTENTS

INDUSTRY REPORTS

A NOTE FROM THE PRESIDENT

The *Computers & Law* Journal was first published in 1983 as a joint publication of all Australian and New Zealand Societies for Computers and the Law. Many leaders from across the legal and computer science professions contributed to the Journal, including the founders of the New South Wales and Victorian Societies, Professor Graham Greenleaf and Julian Burnside.

Over the years, the content and style of the Journal has undergone a number of iterations, culminating in its relaunch with this volume. It is now a professional, academically recognised publication, contributing to important discourse on the intersection of law, policy and technology. The three tiers of content – Academic Articles, Thought Leadership Essays and Industry Reports – are a reflection of the growing diversity of AUSCL's membership, the critical nature of the issues being discussed and the importance of learned discussion.

Offering a tiered approach, aligns with AUSCL's mission to co-create a sustainable future, as we strongly believe that interdisciplinary and intergenerational engagement is essential for carefully considering and solving some of the most significant challenges faced by modern society.

On behalf of AUSCL and the present and future readership of this Journal, I thank Professor Natalie Stoianoff (Editor-in-Chief), Associate Professor Maria O'Sullivan (Managing Editor – Academic), Rob MacLean (Managing Editor – Industry), the Editorial Board and our publisher, AustLII, for their vision, professionalism and perseverance in bringing this project to fruition for the benefit of all.

Marina Yastreboff
President
Australasian Society for Computers and Law (AUSCL)

A NOTE FROM THE EDITORS

Welcome to the rebirth of *Computers & Law*, the Journal for the Australian and New Zealand Societies for Computers and the Law (now the Australasian Society for Computers and Law). It has been a long time in the coming but it is now here and with a new look and format and a permanent home.

As with the evolution of the Australasian Society for Computers and Law (AUSCL), the journal, *Computers & Law*, was ready for change. Taking advantage of not only technological developments but also the opportunity of open access, the journal provides a range of publication types from fully refereed Academic Articles to Thought Leadership Essays, and Industry Reports.

We think you will agree that the 5 Academic Articles, 3 Thought Leadership Essays and 3 Industry Reports in this Volume 94 provide a variety of opportunities for readers to engage with current issues in the field. The topics covered include the challenge of copyright and NFTs, ethical AI and SMEs, quantum technologies' impact on society, the copyright protection of data and data bases, the Digital Platform Ombud Scheme, explainable AI in healthcare, self-regulation of social media platforms, cyber insurance and more. This journal is an opportunity for each of the AUSCL Workstreams to report on their achievements either in the form of Industry Reports or the submission of a Thought Leadership Essay while the Academic Articles provide an avenue for the excellent research being conducted in our universities and abroad to have a wider readership and real-world impact.

This volume 94 is published by AustLII Foundation Press which is a new publishing endeavour of AustLII. We are grateful to AustLII and especially Associate Professor Philip Chung and Professor Andrew Mowbray (who have joined our Editorial Board) for this partnership and for providing a permanent home for this and subsequent issues of Computers & Law. With the launch of this volume, it is our intention to provide early access to articles online as and when they are accepted and ready for publication, hence you will notice the special page numbering system. Once an appropriate number of articles, essays and reports are gathered an issue will be published, which means we have the flexibility to publish more than one issue per year (that is per volume). We are also entertaining the possibility of publishing special issues as and when they are requested and approved by the editorial board.

We are now gearing up for the 40th Anniversary issue of *Computers & Law* and look forward to receiving many more submissions across the three categories. The deadline for submissions will be 15 December 2023 and must follow the *Information for Contributors* described in this volume including the 4th edition of the *Australian Guide to Legal Citation.*[1] Questions and submissions should be sent to the <editors@auscl.org>.

[1] AGLC4 is available at <https://law.unimelb.edu.au/mulr/aglc/about>.

Meanwhile, we hope you enjoy this volume and we express our gratitude to all those who have contributed.

September 2023

CRYPTOKITTIES, ART TOKENS AND BORED APES IN THE METAVERSE: HOW NON-FUNGIBLE TOKENS (NFTS) CHALLENGE AUSTRALIAN COPYRIGHT LAW DURING AN AGE OF DISRUPTION

WELLETT POTTER*

ABSTRACT

The post-COVID-19 era is an age of disruption, which presents significant social, cultural and technological challenges and opportunities for society at large. There has been substantial wealth generation fuelled from digital currencies, which has led to interest and sales of Non-Fungible Tokens ('NFTs') and their associated assets. This article will examine the growth and hype about artistic NFTs in the context of recent years. It will then examine the application of current Australian copyright laws to such NFTs and their assets to determine subsistence and infringement of these works. The notion of what it means to 'own' an NFT will be examined. When applying traditional proprietary notions of ownership to NFTs, it will be seen that they have the capacity to challenge established norms which have evolved in a material, pre-technological world. Finally, this article will ponder the question as to whether a new type of virtual ownership right is emerging for NFTs and their associated assets.

CONTENTS

* Lecturer, University of New England, Armidale, Australia. This article was developed from a presentation given online at the *Asian Pacific Copyright Association 2022 Conference* (National University of Singapore) 14 November 2022. The author can be contacted at wpotter2@une.edu.au.

I Introduction

The post-COVID-19 era is an age of disruption, which presents significant social, cultural and technological challenges for society at large. On one hand, while there has been supply-chain interruptions, food shortages and inflation in many countries, on the other, there has been substantial opportunities, increased digitalisation[1] and significant wealth generation fuelled by digital economies. One of the implications of this has been the growth, interest and sales ('drops') of Non-Fungible Tokens (hereafter 'NFTs'). NFTs have been in existence for over a decade, but during the last year their popularity and notoriety has exploded. In 2021, global NFT sales topped $24.9 billion, a massive increase from $94.9 million in 2020.[2] Interestingly, 85% of all NFT transactions have been performed by 10% of traders, which indicates that a small minority are engaging in a large number of sales.[3]

As of December 2022 in Australia, there are no specific regulations which have been developed for NFTs. Nor has there been any intellectual property ('IP') litigation involving NFTs. Depending on its use, an NFT might fall under intellectual property, consumer and securities law. An NFT might also qualify under the general definition of a 'financial product', through s 763A of the *Corporations Act 2001* (Cth) and be subject to the regulatory framework of the *Australian Securities & Investment Commission* ('ASIC').

This article seeks to examine how NFTs and their associated assets challenge Australian copyright law and notions of proprietary ownership in the post-COVID era. To achieve this, firstly, this section will examine the growth and hype about NFTs in the context of recent years. There will be discussion about what an NFT is and exploration of some of the most expensive and notable drops throughout 2021. This will lead into analysis in Section II about what it means to own an NFT, which will explore the process of token creation and sale.

Section III will then examine the application of current Australian copyright laws to NFTs to determine how copyright may protect them. This article will then focus upon the copyright protection of NFTs attached to digital art as a case study, using the example of a *Bored Ape Yacht Club* NFT ('*BAYC*'). Then, the notion of NFT copyright infringement will be discussed. The issue as to whether tokens are considered property in their own right will also be considered. As there has not been any copyright litigation involving an NFT in Australia at the time of writing, recent cases from China, the US and UK will be used to analyse these issues.

Finally, section IV will argue that from a legal perspective, NFTs challenge traditional notions of property, which originated in the pre-technological, material world, involving the fundamental concepts of ownership, control and what it means to possess an item. It will be pondered whether a new type of virtual ownership right – a meta-property right – for digital assets is slowly evolving in the NFT marketplace.

---

1 See generally, Daren Tang, 'The Future of Intellectual Property and WIPO in a Time of Crisis and Opportunity' (2022) 32 *Australian Intellectual Property Journal* 204.

2 Elizabeth Howcroft, *Reuters: NFT Sales Hit $25 Billion in 2021, But Growth Shows Signs of Slowing* (Web Page, 12 January 2022) <https://www.reuters.com/markets/europe/nft-sales-hit-25-billion-2021-growth-shows-signs-slowing-2022-01-10/>. Note: data was taken from market tracker DappRadar.

3 Ibid.

A *What is an NFT?*

Although the term 'NFT' is often used, there generally remains confusion about what it is and how it is used. When a good is 'fungible', it means that it is 'easy to exchange or trade for something else of the same type and value'.[4] An example of a digital fungible product is bitcoin – like can be exchanged for like. However, by their very nature, NFTs are *not* fungible, meaning that no two NFTs have the same properties – each is unique and it is this uniqueness that gives them value (emphasis added). It is very important to note that an NFT is the *token* which is attached to an asset, rather than the asset itself (emphasis added). An NFT is defined as 'a unique unit of data (= the only one existing of its type) that links to a particular piece of digital art, music, video, etc. and that can be bought and sold.'[5] An analogy can be drawn to an auction ledger, which outlines the details of what has been auctioned, as distinct to the item which has been sold.

Another way to describe an NFT is a type of digital deed or token on a blockchain platform, which is linked to a unique item. The blockchain acts as a type of a public ledger by representing and authenticating the digital asset (sometimes referred to as the underlying asset), by verifying its ownership and history. Each new transaction represents a new block on the chain. Blockchain registration results in NFT ownership being recorded in a decentralised and transparent way. Another benefit is that registration can assist the creator to receive a resale royalty payment if the NFT is later resold – this instruction is coded into the contract so that it happens automatically upon resale.[6] The information which is recorded on the blockchain includes who created the asset, who linked it to the blockchain and its purchase history. This important and nuanced distinction between the NFT (i.e., the token/data on the blockchain) and the digital asset to which it is linked is one which is not often well understood. For the purposes of this article, where relevant, the terms 'NFT/token' and 'asset/associated asset' will be used to delineate these items.

NFTs originated from the need in digital environments to be able to replicate the properties belonging to physical items, which include uniqueness, scarcity, being rivalrous and proving ownership. This began as a method for visual and digital artists to be able to prove digital ownership of their work on the blockchain and to be able to control their work's value.[7] Tokens have the capacity to enable scarcity for any associated asset.[8] For example, a token can make the ownership of digital art exclusive, and verify the ownership of such assets, which were issues that had previously been problematic. As the blockchain comprises of many computers working on complex algorithms to validate a transaction and all records are kept public, this means that it is very difficult to conduct fraud. The unique identity of the NFT on the blockchain cannot be replicated or replaced and this assists in proving/tracing

---

[4] Cambridge University Press, *Cambridge Advanced Learner's Dictionary and Thesaurus, Fungible* (Web Page, 19 October 2022) <https://dictionary.cambridge.org/dictionary/english/fungible>.

[5] Cambridge University Press, *Cambridge Advanced Learner's Dictionary and Thesaurus, Non-Fungible Token* (Web Page, 19 October 2022) <https://dictionary.cambridge.org/dictionary/english/non-fungible-token>.

[6] Logan Kugler, 'Non-Fungible Tokens and the Future of Art' (2021) 64(9) *Communications of the ACM* 19, 20. Also see generally, Michael D Murray, 'NFTs Rescue Resale Royalties? The Wonderfully Complicated Ability of NFT Smart Contracts to Allow Resale Royalty Rights' (July 15, 2022). Available at SSRN: <https://ssrn.com/abstract=4164029> or <http://dx.doi.org/10.2139/ssrn.4164029>.

[7] See generally, Jessica Bookout, et al, 'A Brief Introduction to Digital Art & Blockchain' (2019) 37(3) *Cardozo Arts & Entertainment Law Journal* 553.

[8] Joshua A T Fairfield, 'Tokenized: The Law of Non-Fungible Tokens and Unique Digital Property' (2022) 97(4) *Indiana Law Journal* 1261, 1262.

ownership. The NFT concept has rapidly expanded to various types of items across many industries, including the gaming industry, where it is now commonplace for in-game purchases to be NFTs. Popular NFT games include 'The Sandbox', 'Cryptokitties', and 'Axie Infinity'.

Broadly, there are two categories of NFTs: those which are associated with digital, intangible works and those which are associated with tangible works. In the first category, many types of digital assets may be attached to NFTs – for example, any type of image/artwork, a video, an audio file, online real-estate, memes, gaming avatars/assets, or even a tweet. When a digital asset is the subject of an NFT, it is important to note that what is usually stored on the blockchain is a link to it, rather than the asset itself (it is unusual to store the asset on the blockchain). Specific examples of digital NFTs include *Cryptokitties*, which were created in 2017 by Dapper Labs, *Cryptopunks,* which were launched in 2017 by Larva Labs and the *BAYC* which was created in 2021 by Yuga Labs.

The other category of NFTs are associated with tangible assets. They work through linking the asset with the NFT via a QR code or link. In this way, these NFTs have similar properties to a certificate of title for real property. When a purchase is made, no physical object changes hands, but the NFT guarantees the authenticity of the original asset and denotes ownership. It is possible for these NFTs to be associated with many types of assets. Examples include tokens for tangible artwork, sound recordings, ticketing for events and certificates of title for real estate[9] – the possibilities are almost limitless.

B *Record-Breaking NFTs*

In the past year, NFTs have gained media attention due to record-breaking sales. In December of 2021, PAK's NFT attached to an artwork titled 'The Merge' sold for a record $91.8 million USD, which comprised of 28,983 collectors purchasing 312,686 NFTs to this work.[10] The way that this drop worked was that over three days, purchasers could buy any number of tokens that they wished – this is known as an 'open edition' drop, where there is no limit on tokens.[11] The tokens began at a price of $75 USD and increased at six-hour intervals by $25 USD.[12] Purchasers received their token to access the art once the drop closed.[13]

Another token attached to artwork which sold in March of 2021 was a digital collage artwork (a digital JPEG file), titled 'Everydays – the First 5000 Days' by Artist Beeple (Mike Winkleman).[14] This sold at *Christie's* for $69 million USD.[15] It comprised of a collage of artworks, which Beeple had been producing each day for fourteen years.[16] Similarly, another

---

[9] Kent Barton, 'New Frontiers, Enter the Metaverse: Challenges and Opportunities in NFTs' (Shapeshift Report, 29 May 2021) Foreword.

[10] Fang Block, *PAK's NFT Artwork 'The Merge' Sells for $91.8 Million – PENTA* (Web Page, 7 December 2021) <https://www.barrons.com/articles/paks-nft-artwork-the-merge-sells-for-91-8-million-01638918205>.

[11] Ibid.

[12] Ibid.

[13] Ibid.

[14] Jacob Kastrenakes, *Beeple Sold a NFT for $69 Million – The Verge* (Web Page, 12 March 2021) <https://www.theverge.com/2021/3/11/22325054/beeple-christies-nft-sale-cost-everydays-69-million>.

[15] Ibid.

[16] Ibid.

sale of distinction in November of 2021 was a token for virtual land in the online game 'The Sandbox', which was purchased by investor Republic Realm for $4.3 million.[17]

However, there appears to be volatility in the NFT marketplace and speculation is that growth has started to slow as of December 2022.[18] An example is Twitter founder Jack Dorsey's first tweet, which was linked to an NFT and sold for $2.9 million in 2021.[19] The tweet was published on 21 March 2006 and stated 'just setting up my twttr'.[20] However, Dorsey's tweet NFT value dropped 99% in a year,[21] which demonstrates that an NFT's value is driven by market demand. Community engagement with tokens plays a critical role with this.[22] There are cycles that expand and contract through time, with expensive NFTs pitched to a niche market. This parallels the situation with tangible collector's assets, such as fine art.

C What Makes NFTs Valuable?

There are various economic and socio-cultural factors which prompt people to attribute value to NFTs. Such factors pertain to the digital file linked to the NFT and include utility, aesthetics/appeal to their owner, the uniqueness and scarcity, the fame of the creator/owner and the potential liquidity status of the asset itself. The purchase of NFTs as an investment has been attractive to wealthy individuals such as celebrities and singers, with examples including singer Justin Bieber owing a $625,130 USD portfolio and comedian Dave Chappelle a $262,000 USD portfolio.[23]

Part of the socio-cultural attraction of NFT ownership was explained by celebrity Paris Hilton, when interviewed in January of 2022 on the *Tonight Show* with Jimmy Fallon. Both parties had purchased separate digital cartoon ape NFTs from the popular *BAYC*, with Ms Hilton purchasing her ape for approximately $300,000 USD.[24] Upon showing the audience a hard-copy digital print of her NFT *BAYC #1294* ('ape *#1294*') which was depicted wearing a hat and sunglasses, Ms Hilton explained what attracted her to purchase it. 'That's my ape – it's really cool… the hat, the shades… I was going through a lot of them, I was like, I want something that reminds me of me, but this one, it does. We made, like, another version of it where he takes the hat off and blonde hair comes out… [an] animated version.'[25]

---

[17] Elizabeth Howcroft, *Reuters: NFT Sales Hit $25 Billion in 2021, But Growth Shows Signs of Slowing* (Web Page, 12 January 2022) <https://www.reuters.com/markets/europe/nft-sales-hit-25-billion-2021-growth-shows-signs-slowing-2022-01-10/>.

[18] Alex Wilhelm and Anna Heim, *Are We Entering a NFT Downturn – TechCrunch* (Web Page, 11 March 2022) <https://techcrunch.com/2022/03/10/are-we-entering-an-nft-downturn/>.

[19] Jeff Kauflin, *Why Jack Dorsey's First-Tweet NFT Plummeted 99% In Value in a Year, Forbes* (Web Page, 14 April 2022) <https://www.forbes.com/sites/jeffkauflin/2022/04/14/why-jack-dorseys-first-tweet-nft-plummeted-99-in-value-in-a-year/?sh=525d80ef65cb>.

[20] Jack Dorsey, *Just setting up my twttr - Twitter* (Web Page, 21 March 2006) <https://twitter.com/jack/status/20>.

[21] Kauflin (n 19).

[22] Logan Kugler, 'Non-Fungible Tokens and the Future of Art' (2021) 64(9) *Communications of the ACM* 19, 20.

[23] Jai Singh, *Bored Ape NFT's most popular with celebrities like Madonna and Paris Hilton – Proactive Investor UK* (Web Page, 24 June 2022) <https://www.proactiveinvestors.co.uk/companies/news/985805/bored-ape-nft-s-most-popular-with-celebrities-like-madonna-and-paris-hilton-985805.html>.

[24] Nate Kostar, *Paris Hilton Announces BAYC #1294 Purchase on the Tonight Show – Rarity Sniper News* (Web Page, January 2022) <https://raritysniper.com/news/paris-hilton-announces-bayc-1294-purchase-on-the-tonight-show/>.

[25] The Tonight Show With Jimmy Fallon, 'Paris Hilton Surprises Tonight Show Audience Members By Giving Them Their Own NFTs' (YouTube, 25 January 2022) 00:04:25-00:04:55 <https://www.youtube.com/watch?v=5zi12wrh5So&t=386s>.

Mr Fallon went on to further articulate his attraction to his token, 'We're part of the community. This is my ape. It reminded me of me a little bit because I wear striped shirts, I've worn these heart sunglasses because my daughters, just as a joke, they have them and just as a joke - I put them on, so I've done this and I love yacht-rock and being breezy. And I like the blue.'[26] What can be gleaned from this conversation is that these purchasers have bought their NFT apes for several reasons which will briefly be explored.

Firstly, each of the tokens attached to digital ape images are an economic investment, due to their value, uniqueness and scarcity. Also, the fact that each token has been purchased by a celebrity will likely improve their liquidity status in the future, should they be sold. Secondly, these assets belong to a popular NFT collection, so they are a type of digital collector's item which are highly sought-after. Thirdly, the aesthetics of each individual ape has appealed to their purchasers, with each personally identifying with them. This is similar to the reasons that attract people to buy tangible collector's goods, such as baseball cards or vintage Smurf figurines.

As mentioned earlier in this article, as a token is an intangible good – a record on a digital ledger - this begs the question as to what a buyer is actually paying for when they make a purchase. What does it mean to own a token when the asset exists in an intangible, digital environment or in the metaverse? The following section will unpack these issues in greater detail.

II WHAT DOES IT MEAN TO OWN AN NFT? WHAT ARE BUYERS PAYING FOR?

A *The Minting of NFTs*

In order to understand what it means to own a token, it is necessary to understand a general overview of the minting process. The first step is that a creator must have created some type of asset (for example, an artwork, literary work, etc) which they wish to link to an NFT.[27] A creator can upload the details about this asset to an NFT platform, and execute a smart contract. This involves deciding upon the conditions of the contract, coding this and uploading this data to the appropriate NFT platform. Doing this has given some creators greater freedom in setting their own cost and cutting out middle entities, which has been particularly exciting for the digital art industry.[28] The smart contract is usually programmed to self-execute if particular conditions of sale are met,[29] which are based in contract law.[30] It is also possible to program smart contracts to create new events or execute further contracts or tokens, or to embed cryptocurrencies or further digital assets (although this happens infrequently).[31]

Depending on the complexity and timing of the transaction, a compulsory registration fee is charged. This is known as a 'gas fee' and is often very expensive due to the carbon

---

[26] Ibid 00:04:55-00:05:22.

[27] Rebecca Carroll, 'NFTs: The Latest Technology Challenging Copyright Law's Relevance within a Decentralized System' (2022) 32(4) *Fordham Intellectual Property, Media & Entertainment Law Journal* 979, 986.

[28] Ibid 1005.

[29] Shaan Ray, 'NFTs and Smart Contracts – LinkedIn' (Web Page, 19 May 2021) <https://www.linkedin.com/pulse/nft-smart-contracts-shaan-ray-mba>.

[30] Fairfield, (n 8) 1290.

[31] Ray (n 29).

emission cost. This has meant that in the past, the token market has generally been geared towards the wealthy. This 'gas fee' is usually paid to the network validators and it covers the cost of the blockchain services, including energy costs, validation and securing fees.

There has been controversy associated with the amount of power and therefore carbon emissions that it takes to mint a token. The reason for this is that the underlying process of maintaining certain blockchains known as 'proof of work' is deliberately energy intensive.[32] An example is Ethereum blockchain, where miners must solve an algorithm to add a new block of verified transactions. This is deliberately a high-energy and therefore expensive task, so as to dissuade fraudulent activity. An average NFT on the blockchain has been found to have a carbon footprint which is higher than an EU resident's monthly electricity usage.[33] However, in recent times, some blockchain, such as Solana, are offering the opportunity to be carbon neutral and this results in lower 'gas fees'.

When a creator mints a token, after paying the 'gas fee', they are assigned a 'public key'. This is used to verify that they minted (or authored) the digital item – it is a type of certificate of authenticity. The 'public key' is recorded in the token's metadata and on the blockchain, which gives it transparency. Minting on the decentralised blockchain also means that this data cannot be altered or deleted. The 'public key' has the additional utility of allowing a creator to earn resale royalties, every time that their token is later resold to a new owner, which is a particularly attractive option to artists.[34] This has provided a solution to a long-standing problem of resale royalties, particularly for digital artists. It also fulfils one of the aims of copyright – to incentivise authors to continue producing their works.[35]

Because NFTs are secured on the blockchain, they are minted and traded using various forms of cryptocurrency, including Solana, Flow, Ethereum and Wax. Upon buying a token, a new owner is assigned a 'private key', which is stored in a digital wallet. This 'private key' provides security and verifies original ownership, proving that the new owner's digital token is the original.[36] When a purchaser meets the conditions outlined in the smart contract (usually through purchasing the token with cryptocurrency) the smart contract is automatically executed.[37] It will distribute the NFT to the purchaser and record the transaction on the blockchain.[38]

A smart contract contains terms of sale – it provides a license, which outlines the permitted use of the associated digital file. When the purchaser executes the contract, the terms are typically grounded in pre-existing IP rights and can considerably vary. For example, a purchaser is usually not granted commercial rights to the digital file attached to the token. Instead, the IP rights to the associated asset are likely retained by the copyright owner, who may be the creator. In a similar situation to other creative works which are the subject of IP rights, it is only the copyright owner who has the right to exercise those rights, including the right of reproduction.

Alternatively, in a rarer number of situations, a token purchaser may be granted partial or full commercial rights to their associated asset, such as the right to exploit derivative works.

---

[32] Justine Calma, *The Climate Controversy Swirling Around NFTs – the Verge* (Web Page, 16 March 2021) <https://www.theverge.com/2021/3/15/22328203/nft-cryptoart-ethereum-blockchain-climate-change>.

[33] Ibid.

[34] Ethereum Inc, *Non-Fungible Tokens (NFT)* (Web page, 25 October 2022) <https://ethereum.org/en/nft/>.

[35] John Locke, *Second Treatise on Government* (1680); *Statute of Anne 1710* (UK) 8 Anne, c 19; Hettinger, 'Justifying Intellectual Property' (1989) 18(1) *Philosophy & Public Affairs* 31, 36–7.

[36] Ethereum Inc (n 34).

[37] Ray (n 29).

[38] Ibid.

The licensing of the *BAYC* is one such example which will be examined as a case study. The next section will unpack these issues by starting with a discussion about subsistence in tokens and their linked digital assets under Australian law.

III Australian Copyright Subsistence, Implications & Challenges

When considering copyright subsistence in NFTs under Australian law, there are several issues to unpack. Again, it is important to delineate between copyright of the associated asset and copyright of the token itself.

A *Subsistence in Associated Assets*

When considering whether copyright subsists in an associated asset, there are a number of subsistence criteria that must be met and if they are satisfied, then copyright automatically vests. Firstly, it is necessary to consider how the associated asset would be classified as subject matter under the *Copyright Act 1968* (Cth) (hereafter '*the Act*'). Under the Act, works are divided into Part III and Part IV works. Part III works are those which are historically more creative endeavours and fall within the scope of the *Berne Convention*.[39] This incorporates literary, dramatic, musical and artistic works.[40] Part IV works include more modern works, including sound recordings,[41] films,[42] television/sound broadcasts[43] and published editions of works.[44] If an associated asset fell under one of these categories, then it may qualify for protection, as long as the other subsistence criteria are met.

Although the scope of subject matter that may be protected through copyright is broad, there are some associated assets which would likely fall outside of copyright protection. For example, the associated assets to recent NFTs have included perfume or customisable *Nike* sneakers.[45] Such subject matter falls outside of copyright, but may be covered under other IP protection. Brands are increasingly using NFTs as another form of innovative revenue, with companies such as *Adidas*, *Dolce & Gabbana*, *Nike* and the *US NBA* entering the metaverse and/or selling collectable items as tokens.

For copyright to subsist under Australian law, it is important that an author is identifiable, they must have a territorial connection to Australia[46] and they must demonstrate sufficient originality in the creation of the work.[47] These issues will be unpacked in greater detail in the next section. As long as the associated asset falls within the scope of copyright subject matter and all of the subsistence criteria are met, copyright would vest separately in the asset. Of note is that unless the author of the associated asset is also the token creator, this means that authorship would vest in two different people for two different works – (1) for

---

[39] *Berne Convention for the Protection of* Literary *and Artistic Works*, opened for signature 9 July 1886, 943 UNTS, 178 ('*Berne*'. The UK signed *Berne* on behalf of its dominions (including Australia) on 5 December 1887. *Berne* formally entered into force in Australia on 1 March 1978 and the US on 1 March 1989.

[40] *The Act* s 32.

[41] Ibid ss 85 and 89.

[42] Ibid ss 86 and 90.

[43] Ibid ss 87 and 91.

[44] Ibid ss 88 and 92.

[45] Chris Williams, *Nike Bought RTFKT. Now Its NFTs Are Trading at a Premium* (Web Page, 15 December 2021) <https://cryptobriefing.com/nike-bought-nft-now-its-nfts-are-trading-premium/>.

[46] *The Act* ss 32(1)(a), (c).

[47] *IceTV Pty Limited v Nine Network Australia Pty Limited* (2009) 239 CLR 458, 478–9 [47]–[48] (Gummow, Hayne and Heydon JJ) ('*IceTV*').

the creation of the associated asset; and (2) for the creation of the token. This fact is often not well understood by the public at large.

Another issue is whether the token creator has the right to mint a token from an associated asset. Sometimes this can be straightforward, particularly if the associated asset is created or owned by the same person who mints the NFT. However, this issue can become complex if a copyright-protected work has already been licensed for particular uses, or where some rights have already been transferred, but a license has been retained for a particular use/s. It is debatable as to whether the minting of an NFT of an associated work would be covered under pre-existing licenses - terms would need to be closely interpreted.

B *Subsistence in Tokens*

As a token primarily comprises of text on the blockchain and can be reduced to zeros and ones in its most basic form, its subject matter may qualify as a literary work under Part III of the Act.[48] However, the arrangement of the data underlying the token may be found to lack sufficient arrangement to qualify for copyright protection – this would require evaluation on a case-by-case basis.

The other subsistence criteria denotes that a literary work must be an original work of human authorship,[49] which is reduced to a tangible form by an author (i.e., the token creator or author). This reduction to tangible form is the process of being made; of being written.[50] When a token creator mints their token, they are engaging in a process of reduction to tangible form, through the arrangement of the code for the smart contract.

Also, the token creator must have a territorial link to Australia - at the time the token is made, as the author, they must be a qualified person,[51] or the first publication (i.e., executing the smart contract/placing the token on the blockchain) must occur in Australia.[52] Under the Act, a qualified person is an Australian citizen or resident.[53]

If a token became the subject of a judicial subsistence enquiry, it is likely that the process of the input of the creator (author) in expressing the data in a tangible form would be rigorously examined.[54] There would be a focus upon whether the process contained sufficient original authorial 'independent intellectual effort' through the expression of the arrangement of the token's data.[55] This is because copyright will only protect the *expression* of data, but not the data itself (i.e., the idea/expression dichotomy).[56] It is possible that some tokens may fail to meet this criterion due to a lack of established independent intellectual effort in the expression of the arrangement of data.

---

48 *The Act* s 32.

49 *Sands & McDougall Pty Ltd v Robinson* (1917) 23 CLR 49, 55 (Issacs J); *Desktop Marketing* (2002) 119 FCR 491 532 [160(2)] (Lindgren J), 593 [409] (Sackville J) ('*Desktop Marketing*'); *IceTV* (n 47) 474 [33]–[34] (French CJ, Crennan and Kiefel JJ); *Telstra Corporation Limited v Phone Directories Company Pty Ltd* (2010) 194 FCR 142, 172 [100] (Perram J) ('*Telstra Appeal*').

50 *The Act* s 22(1).

51 Ibid s 32(1)(a).

52 Ibid s 32(2)(c).

53 Ibid s 32(4).

54 Neal F Burstyn, 'Creative Sparks: Works of Nature, Selection, and the Human Author' (2015) 39(2) *Columbia Journal of Law & the Arts* 281, 299-303.

55 *IceTV* (n 47) 478–9 [47]–[48] (Gummow, Hayne and Heydon JJ).

56 *Agreement on Trade-Related Aspects of Intellectual Property Rights* opened for signature 15 April 1994, 1869 UNTS 299 (entered into force 1 January 1995) art 9 § 2 ('*TRIPS*'); Omri Rachum-Twaig, 'A Genre Theory of Copyright' (2016) 33(1) *Santa Clara High Technology Law Journal* 34, 78–82.

The future use of artificial intelligence ('AI') in token creation must also be considered. Unlike the UK's copyright legislation, for any type of computer-generated work, the Act does not make provision for the author to be considered the person by whom the arrangements necessary for the creation are undertaken.[57] During the past ten years, there has been Australian jurisprudence which suggests that, if applied to NFTs which are created through AI processes, the expression of token data may be found to be too far removed from the actions of a person/people who arranged those processes, due to a lack of human 'independent intellectual effort'.[58] In this situation, the result is that subsistence will fail for lack of establishing sufficient human authorship. As AI continues to advance at an astounding pace,[59] this is likely to be a relevant future issue. If the Act remains the same regarding authorship of computer generated works, then this matter requires judicial evaluation and is heavily fact dependant.

Under copyright, when the subsistence criteria are sufficiently met, as the author of the work, the token creator is conferred a bundle of rights. These include the right to (1) reproduce the token in material form;[60] (2) publish the token;[61] (3) communicate the token to the public;[62] and (4) make an adaptation of the token,[63] which is defined as 'an arrangement or transcription.'[64] Hypothetically, if any of these rights are infringed, then copyright litigation may commence. However, in executing the smart contract, the token creator decides which of these rights to retain and which to transfer to the purchaser. The next section will discuss licensing in further detail and will use the *BAYC* as a case study.

C *Licensing of Tokens & the BAYC Case Study*

As previously mentioned, the terms of token licenses vary. Most involve a centralised collaboration model, where there is no transfer of commercial rights for the associated asset. The result is that a purchaser is not permitted to commercially exploit the associated asset or a derivative work.[65] They can prove ownership of their token on the blockchain, but there are no intellectual or property rights for the associated asset.[66]

Under such conditions, in essence, what a token purchaser is buying is the right to be recorded on the blockchain as the official token owner, the right to access/use the associated asset non-commercially and the right to sell the token in the future. Interestingly, this directly contrasts with the way that NFT sales are marketed, which often suggest that tokens are personal property sold as a digital asset. The use of language in marketing often indicates that

---

[57] *Copyright, Designs and Patents Act 1988* (UK) (ch 48) s 9(3).

[58] *IceTV* (n 47) 474 [33], 479 [48] and 494–5 [99] (French CJ, Crennan and Kiefel JJ); *Telstra Corporation Limited v Phone Directories Company Pty Ltd* (2010) 264 ALR 617, 685 [344] (Gordon J); *Dynamic Supplies Pty Limited v Tonnex International Pty Limited* (2011) 91 IPR 488, 500 [49] (Yates J); *Sports Data Pty Ltd v Prozone Sports Australia Pty Ltd Sports Data Pty Ltd v Prozone Sports Australia Pty* (2014) 107 IPR 1, 13 [74], 14 [76] (Wigney J).

[59] Courtney White and Rita Matulionyte, 'Artificial Intelligence: Painting the Bigger Picture for Copyright Ownership' (2020) 30 *Australian Intellectual Property Journal* 224, 224-228.

[60] *The Act* s 31(1)(a)(i).

[61] Ibid s 31(1)(a)(ii).

[62] Ibid s 31(1)(a)(iv).

[63] Ibid s 31(1)(a)(vi).

[64] Ibid s 10.

[65] Fairfield (n 8) 1298.

[66] Farah Mukaddam, *NFTs and Intellectual Property Rights – Norton Rose Fulbright* (Web Page, October 2021) <https://www.nortonrosefulbright.com/en/knowledge/publications/1a1abb9f/nfts-and-intellectual-property-rights>.

token ownership equates to full proprietary and/or IP ownership of the associated asset.[67] The issue of token marketing will be discussed in the next section.

On the other end of the licensing spectrum, few token smart contracts use a decentralised collaboration license, which allow full commercial IP rights to derivative works of the associated asset. Such terms are unusual. It means that the token purchaser can commercially exploit derivative works based on the associated asset. The *BAYC* is an example. The license grants the *BAYC* purchaser an unlimited worldwide license to use, copy and display the art for the purpose of creating derivative works, without requiring permission from Yuga Labs.[68] The commercial success of such a business model seems promising. Firstly, the initial sale of the *BAYC* has exceeded $1 billion in a year and secondly, several ape owners have engaged in lucrative commercial projects involving derivative works.[69] Examples include a partnership between music producer Timbaland for a *BAYC* hip-hop metaverse band, the use of the *BAYC* to promote NBA basketball shoes, and a contract with Universal Music's 10.22PM label to form a four-ape *BAYC* band.[70] Whether decentralised collaborative licenses become popular for future NFT sales remains to be seen.

Whenever any commercial exploitation is involved, terms will always be closely scrutinised. A current example is from the US, involving Miramax and Quentin Tarantino. In January of 2022, Tarantino sold an NFT linked to a digital image of his original hand-written script of movie 'Pulp Fiction' for 1.1 million USD.[71] Although Tarantino owned the rights to the original screenplay and Miramax owned the rights to the screenplay, Tarantino was sued by Miramax for copyright and trademark infringement, breach of contract and unfair competition.[72]

Miramax argued that Tarantino did not have the right to mint NFTs, because they were captured under the term 'emerging technology' and had been assigned to them under the screenplay contract. Tarantino argued that the tokens were not captured under those terms and therefore it was within his rights to mint the tokens from his hand-written screenplay. On 8 September 2022, the parties settled the case,[73] so it is uncertain what would have been ruled about the classification of NFTs had the case proceeded to trial. This case demonstrates that license disputes pertaining to terms of IP linked to NFTs will likely be fiercely litigated in the future.

D *Infringement*

1 *Token and Associated Assets*

As of December 2022, there has not been Australian litigation about the copyright infringement of a token or an associated asset. Currently, it seems unclear as to whether token infringement will become a prevalent future issue. It must be remembered that infringement

---

67 Ibid.

68 Edward Lee, *The Bored Ape Business Model: De-centralized Collaboration via Blockchain and NFTs*, (PDF, 16 November 2021) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3963881>.

69 Ibid.

70 Ibid.

71 Niel Elan, *Pulp Fiction NFT Lawsuit (Miramax V. Tarantino, Et Al.): A Preview Of Coming* Attractions – *Forbes* (Web Page, 25 July 2022) <https://www.forbes.com/sites/legalentertainment/2022/07/25/pulp-fiction-nft-lawsuit-miramax-v-tarantino-et-al-a-preview-of-coming-attractions/?sh=9921fc6ca24d>.

72 *Miramax LLC v Quentin Tarantino and Visiona Romantica Inc* (United States District Court, Central District of California, Case No. 2:21-cv-08979).

73 Ibid, Notice of Settlement.

can only be found if copyright subsists in a work to begin with. When considering subsistence in a token, as has been discussed, it may be found that this fails, due to insufficient originality and/or authorship of the data.

A more pertinent issue is infringement of an associated asset when minting an NFT. Taking an artistic work as an example, it is hypothetically possible that a digital artist could have an artwork downloaded by an unauthorised person and minted on an NFT platform without their consent. This highlights a difficult issue – proof of authorship/ownership of the associated asset. How do NFT platforms know whether the associated asset is truly authored or owned by the person identified as the author or owner during the minting process? Companies are concerned about the fact that their IP can be minted into NFTs by unauthorised people. DC and Marvel have issued warnings that NFT platforms should not mint tokens from their IP.[74]

An asset which has been reproduced and communicated to the public without permission through minting on an NFT platform could amount to copyright infringement, because these actions would infringe upon an artist/owner's rights.[75] In this situation, the artist/owner could litigate for infringement. If an artist's moral rights were also somehow infringed, then this could also be litigated.

Taking the *BAYC* as an example, the terms of licence include the right for an ape owner to engage in secondary uses and to create derivative works if they wish. It is, however, hypothetically possible for anyone to run a Google search, locate a JPEG image of a *BAYC* ape on the internet and make a copy of a cartoon ape via screenshot, or via downloading the image to their computer as a JPEG file. This JPEG file could also be disseminated online via social media. Perpetrating these acts would technically infringe upon a *BAYC* owner's copyright. However, it appears as though such infringing acts are often tolerated by token owners, as long as no commercial aspect is involved. This hypothetical example also highlights how easy the medium of the internet makes it to infringe copyright. It also raises the issue of the potential liability of the role of NFT platforms.

2 *Liability of NFT Platforms*

As an NFT platform is the means through which a token is displayed, published, communicated, disseminated and sold, a relevant issue is whether the platform could be jointly liable if the rights to an associated asset is infringed by a third party user.

On 23 March 2022, China's first NFT platform infringement case was heard in the Hangzhou Internet Court (a specialised court for internet litigation).[76] The associated asset was an artistic work - a cartoon from the popular 'Fat Tiger' series. It depicted a tiger receiving a vaccination, which had been linked to an NFT and sold on a platform by a third party.[77] The court found that the NFT platform had contributorily infringed the plaintiff's rights to the cartoon through dissemination.[78] It was ruled that the defendant had to destroy the NFT by sending it to an inaccessible address (it is not possible to remove NFTs from the

---

[74] Sunny Kumar, Georgina Rigg and Kira Green, *The NFT Collection: The Rise of NFTs – Copyright Strikes Back? (Part 3) – K & L Gates* (Web Page, 7 July 2022) <https://www.natlawreview.com/article/nft-collection-rise-nfts-copyright-strikes-back-part-3>.

[75] Adarsh Vjayakumaran, 'NFTs and Copyright Quandary' (2021) 12(5) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 402, 410 [34].

[76] *Shenzhen Qicedie Cultural Creativity Company Ltd v Hangzhou Yuanyuzhou Technology Company Limited* (2022) Zhe 0192 Minchu No. 1008. Translation provided by TaylorWessing.

[77] Ibid.

[78] Ibid.

blockchain) and damages of RMB 4,000 (USD $600) were awarded.[79] The court emphasised the importance of NFT platforms implementing policies to check ownership of IP works on their platforms.[80]

In Australia, the current situation regarding online content platforms suggests that they could be held liable for NFTs which infringe copyright, or NFTs which are linked to infringed associated assets. The 2017 case of *Pokèmon Company International Inc v Redbubble Ltd*[81] ('*Pokèmon*') considered the liability of Redbubble, an online marketplace, in relation to copyright infringement of an artistic work, as well as violation of Australian Consumer Law. Although this case did not pertain to NFTs, the underlying principles would likely be applicable in an NFT context.

Redbubble allowed third parties to upload images to their website, which could then be printed onto products (such as mugs, clothing etc) when ordered by customers.[82] An unauthorised original artistic image of Pikachu belonging to the Pokèmon Company was uploaded to the Redbubble site and *Pokèmon* sued for infringement.[83]

The Federal Court examined Redbubble's actions in attempting to mitigate copyright infringement.[84] This included user agreements declaring that the user owned or had permission to upload the works, a system of notification of infringing content, a take-down procedure and a content-monitoring team.[85] After examining these methods, the Federal Court found that this was insufficient to mitigate Redbubble's liability. The court stated 'In each case the originator was the artist [user] who had placed the image on the Redbubble website. Redbubble, however, was responsible for determining that content through its processes, protocols and arrangements with the artists'.[86]

After analysis, the Federal Court found that Redbubble could have taken other steps to have prevented the infringement from occurring, but chose not to.[87] Redbubble were found to have committed three forms of infringement, including (1) direct infringement through communication of the infringing works to the public;[88] (2) knowingly exhibiting infringing works publicly to engage in trade[89] and (3) indirectly infringing through the authorisation of the reproduction of the copyright material to manufacture and sell infringing articles.[90]

The Federal Court rejected the awarding of an injunction and nominal damages were set at $1.[91] This reflected the court's finding that the *Pokèmon* Company had not sufficiently demonstrated that they had truly suffered loss, while taking into consideration the methods Redbubble used to mitigate infringement. In 2018, an appeal and cross appeal was filed, but this has been stayed until after a July 2022 decision was handed down from a trademark infringement case involving Redbubble.[92]

---

79 Ibid.

80 Ibid.

81 (2017) 351 ALR 676 (Pagone J).

82 Ibid 679 [5].

83 Ibid 691-693 [32]-[35].

84 Ibid 679 [6].

85 Ibid 682-3 [15], 710-714 [59]-[66].

86 Ibid 701 [48].

87 Ibid 714 [67].

88 Ibid 699-702 [45](a)-[49], contravening s 36 of the Act.

89 Ibid 699 [45](b), 702-704 [50]-[54] contravening s 132AG of the Act.

90 Ibid 699 [45](c), 704-709 [55]-[57].

91 Ibid 719, Orders.

92 *Hells Angels Motorcycle Corp (Australia) Pty Ltd v Redbubble Ltd (No 5)* [2022] FCA 837 (Greenwood J).

The court's analysis and ruling in *Pokèmon* suggest that it will be difficult for NFT platforms to argue ignorance of potential infringement by users. In Australia, there are no safe harbours for commercial online content platforms for when their users infringe. As seen in *Pokèmon*, having users agree to terms of service, including an agreement not to upload infringing material and/or providing indemnity to the NFT platform for infringement or illegal activity, will not necessarily allow online platforms to escape liability.[93] However, *Pokèmon* suggests it is important for NFT platforms to be able to prove to the court what measures they implement to mitigate copyright infringement. The next section will consider whether tokens are considered property in their own right.

E *Are Tokens Considered to be Property In Their Own Right?*

Having explored the implications of infringement, another issue to ponder is whether a token may be considered to be property in its own right. As of December 2022, there has been no Australian cases directly on-point. However, on 10 March 2022, the High Court of England and Wales passed judgement regarding two stolen NFTs.[94] The proceedings were for a restraining order to prevent the dissipation of the tokens and for a Bankers Trust Disclosure Order.[95] The Bankers Trust Disclosure Order was against Ozone Networks Inc, a peer-to-peer NFT marketplace based in the US, for information which would help to identify the unknown perpetrators.[96] The two tokens were owned by the claimant, Laviniah Osbourne ('LO'), a British citizen, and were stolen by unknown person/s.[97]

The background was that on 24 September 2021, in exchange for work, LO had been gifted some tokens which she stored in her crypto wallet.[98] Two of these tokens represented artworks from the 'Boss Beauties' collection, depicting career women from a range of backgrounds.[99] Without LO's consent or knowledge, on 17 January 2022, unknown persons illegally accessed LO's wallet and stole these tokens, transferring them to two other Ozone accounts.[100] On 27 February 2022, LO discovered they were missing, traced them to the new accounts and commenced legal proceedings to freeze the accounts and to prevent any further transactions of these tokens.[101]

The English High Court found that the claim was a good cause of action, despite being lodged against unknown persons – clearly, the tokens had been defrauded from LO's account.[102] Because the order was directed at unknown persons, an issue that was material to this claim was the location of the tokens at the time they were stolen.[103] The court described an NFT as 'a stream of electrons resulting in a credit item to a crypto account.'[104] It was found that the physical manifestation of a token would likely be where their servers were located.[105]

---

93 *Pokèmon* 679 [6].

94 *Lavinia Deborah Osbourne v Persons Unknown, Ozone* [2022] EWHC 1021 (Comm) (Pelling J).

95 Ibid [3]-[4].

96 Ibid [4].

97 Ibid [8].

98 Ibid [7].

99 Ibid.

100 Ibid [8].

101 Ibid [8]-[10].

102 Ibid [12]-[13].

103 Ibid [12].

104 Ibid [14].

105 Ibid.

However, in this instance it would have been impossible to litigate on that basis.[106] Instead it was found appropriate to follow other recent UK decisions involving crypto currency fraud, where the location (and therefore jurisdiction) was determined not by the location of the server, but of the location of the owner's domicile, which was Britain.[107] Then, the High Court turned attention to the Bankers Trust Disclosure Order against Ozone, analysed the requisite tests and found this appropriate to execute.[108]

When considering whether the tokens constituted property, his Honour Pelling J stated, 'I am satisfied on the basis of the submissions made on behalf of the claimant that there is at least a realistically arguable case that such tokens are to be treated as property as a matter of English law.'[109] The injunction against unknown person/s was granted,[110] with the acknowledgement that if it was not, there was a 'very real risk that these assets will be transferred through multiple different accounts at great speed, and in a way that will make it practically either very difficult, or possibly even impossible, for the claimant to trace and retrieve her assets.'[111]

Therefore the English High Court acknowledged that such tokens might be treated as legal property – a chose in action. This means that tokens are capable of being owned in their own right and have associated proprietary remedies, which allow their owner to enforce their rights. If a similar situation were to occur in Australia, the outcome may be the same.

Now that the treatment of tokens as legal property has been analysed, the next section will explore the tension between tangible ownership rights and the licensing of digital goods. It will be seen that the digital environment presents interesting tensions when traditional proprietary ownership is considered in the context of intangible, licensed tokens.

IV NFT Challenges and Opportunities for the Future

A *The Tension Between Tangible Ownership and the Licensing of Digital Goods*

Historically, the notion of tangible property ownership spans centuries and is a foundation of capitalistic societies. Looking at artistic works and real estate as examples, it can be seen that the associated rights provide the owners/proprietors with a monopoly over their work through control. An owner's monopoly rights include the right to use the work, to transfer or destroy it, or to exclude others from using it.

Monopoly rights stem from the idea that the owner ought to be incentivised for investing in/creating that property, by receiving benefits from exploiting their rights. These rights have their origins in a non-technological era through Lockean justification,[112] where an owner exercises control over their work through exclusive possession and the right to sell that item, make a copy of it, destroy it etc. Because of the nature of tangible property, this involves a physical asset, which is controllable due to its exclusivity.

---

[106] Ibid.

[107] Ibid [15], following *Ion Science Limited v Persons Unknown & Others (Unreported)* [2020] (Comm) [15] (Butcher J).

[108] Ibid [32]-[55].

[109] Ibid [13].

[110] Ibid [17].

[111] Ibid [20].

[112] Justin Hughes, 'The Philosophy of Intellectual Property' (1988 – 1989) 77 *Georgetown Law Journal* 287.

When the asset is acquired and possessed, others are prevented from exercising control over that asset. A one-to-one relationship exists between the owner and the item, although there may be different categories of co-ownership involved. An analogy here can be made to a famous work of modern art, which hangs in a gallery – there is only one, authentic original, the monopoly of which is exercised through possession. Ownership of the original is clearly delineated through physical possession. If reproductions of the art are made, despite looking like the artwork, they are not the original. Nor may the owner who buys a reproduction claim ownership of the original, authentic work.

When considering a tangible good, there are two different forms of property rights that might be controlled: (1) the property rights in the tangible item itself (i.e., the form of physical possession and control that can be exerted over the item) and (2) the IP rights that exist if the item is a work which falls under copyright. Taking the modern art example, (1) the property rights exist in the art itself as it hangs in the gallery. The art is possessed by a gallery and they exercise physical possession and control over it. (2) Secondly, there is a separate bundle of IP rights which vest in the owner of the work. Depending on the conditions of sale, this might be the gallery, or the artist. The artist would also possess moral rights attached to the art, which remain a personal, personhood right for as long as copyright subsists and which cannot be sold. These rights are for attribution of authorship;[113] against false attribution[114] and for integrity of authorship.[115]

However, for any intangible, digital items which tokens are attached to, the tangible property rights of ownership do not exist in these works to begin with. Digital items cannot be possessed or appropriated in the same way as tangible items are, because by their very nature, they are intangible. Such items exist in an online environment and in their most basic format, they can be reduced to zeros and ones – to data, or information. Under Australian law, a tangible property right simply cannot vest in information.

This means that the only rights that intangible items are likely to possess are IP rights, as has been explored in Section III. Whether such items exist as a digital artwork connected to a token, or as a piece of virtual property in the metaverse, it is fascinating that many people who purchase such items feel a sense of tangible possessory ownership over the item. This feeling of possession and ownership is akin to physically possessing a tangible good in the physical world. However, the reality is that this type of digital ownership can never truly be the same as physical ownership, due to the differences in medium. A person simply cannot possess an intangible work in an identical way to a tangible work due to its very nature. Although there is constant innovation in technology, it is currently impossible for digital items to replicate exactly the same property rights as tangible goods.

B *Digital Kudos and the Emergence of a New Type of Meta-Property Right*

The expectations of token purchasers and the reality of what they purchase are very different. The reality is that, under some smart contracts, when a purchaser believes that they 'own' an NFT, they do not 'own' or possess the item at all.[116] Instead, there are restrictive conditions, which, at the most, allow access to and personal use of the digital asset associated with the

[113] *The Act* pt IX div 2.

[114] Ibid pt IX div 3.

[115] Ibid pt IX div 4.

[116] *Capitol Records v Redigi*, 910 F 3d 649, 659-660 (2nd Cir, 2018).

token. In most situations, the purchaser will not be granted any commercial IP rights to the associated asset – such rights are likely to be retained by the copyright owner.

Of note here is that the marketing and branding of NFTs, often use words which denote physical ownership akin to a chose in action, rather than admitting that limited rights are instead licensed. However, many token creators market their tokens as personal property, when in reality, tokens are licensed (and would be litigated) under IP laws. When tokens are sold, what is often deliberately not emphasised is the fact that the digital items associated with the token cannot ever be truly possessed in the same way as a physical item. Instead, this myth of NFT ownership is perpetuated as being identical to physical possession, when in reality, often what a purchaser is granted are limited rights of access and personal use. The price that purchasers are willing to pay is usually for the opportunity to proclaim to the world at large that they 'own' the token, as affirmed by the blockchain, which gives them access to the associated file. This is a type of kudos – a type of 'digital kudos'– and the price that some have been willing to pay for this has been staggering. What should also be noted is that when a person's identity is usually logged on the blockchain, it is not by their identifiable name, but by a string of unique numbers or letters, known as the TokenID, so it is not usually easy to identify a purchaser.[117]

The reason for the perpetuation of this proprietary ownership myth of tokens is that sales can be highly profitable. Purchasers are more likely to buy a token if they can claim that they 'own' it, similar to owning and possessing a rare vintage Smurf or fine artwork, rather than conceding that all they have paid for is a license, often for limited, non-commercial rights.

V Conclusion

Having explored what constitutes creation, ownership and infringement of an NFT, this begs the question as to whether new forms of ownership will eventually be developed, particularly as humanity increasingly transitions into virtual environments such as the metaverse. The digital environment is already a place of entertainment, socialisation, work and commerce.

NFT usage is representative of the shift in global economies from the tangible to the intangible.[118] When global NFT productivity began in earnest, it was speculated that this would promote a substantial shift in the application of copyright.[119] To-date, this has not occurred. Instead, what has happened is that tokens have become another method to promote the commercialisation of digital items through pre-existing IP paradigms.[120] However, the fact that the business of NFTs has boomed around the world appears to suggest that there has been a subtle shift in purchasers' thinking. There appears to have been an acceptance of the differences in the concept of ownership regarding digital items and the differences in property rights that digital ownership entails, as compared to tangible items.

Instead of needing to display a monopoly right over a work through physical possession and control, many NFT purchasers appear satisfied with claiming prized ownership on the blockchain by paying for what amounts to limited non-commercial rights under license. The fact is, that in many situations, token owners do not physically possess and control the digital item attached to their token in the same way as they could a tangible good. As global NFT

---

[117] Andrew Guadamuz, 'The Treachery of Images: Non-Fungible Tokens and Copyright' (2021) 16(12) *Journal of Intellectual Property Law & Practice* 1367, 1370.

[118] Tang (n 2) 205.

[119] See generally, Ifeanyi E Okonkwo, 'NFT, Copyright and Intellectual Property Commercialization' (2021) 29 *International Journal of Law and Information Technology* 296.

[120] Ibid.

sales continue to soar and the metaverse expands, it appears as though this virtual ownership right – what could be termed a type of meta-property right – is being heavily utilised and accepted in the marketing of NFT and digital asset sales.

Concurrently, what is fascinating is that there appears to be a shift in collective consciousness about the notion of copying digital items such as tokens. Acts which technically infringe upon a token owner's copyright are being tolerated, as long as no commercial aspect is involved. Activities such as reproduction and dissemination, that were once viewed as a major threat to innovation are often viewed as concepts that are acceptable as part of the creative activities of humanity.[121] This collective shift in thinking has resulted in some NFTs adopting Creative Common licences, therefore by-passing copyright altogether.[122] Although it is beyond the scope of this article, the development of tokens through Creative Commons will be an issue to watch with interest in the future, as will other licensing initiatives that promote collaboration, such as 'can't be evil' licenses.[123]

Recent academic commentary has suggested that in marketing tokens, there ought to be a characterisation of tokens as choses in action (as occurred in the UK) rather than relying upon intellectual property and terms of licensing.[124] This would result in the NFT purchaser being permitted a full bundle of property rights - to 'use, benefit from, capture the rise in value from, and otherwise benefit from the social value of being the owner of the item.'[125] However, this prospect will likely generate extensive debate by those whose interests are at stake.

In Australia, it will be interesting to observe whether, in the future, licensing will continue to primarily be used to determine the rights relating to ownership and use of tokens and their associated assets, or whether a new type of legal framework or regulatory regime will emerge. As the technology underlying tokens continues to develop, there is tremendous potential for new uses and innovation. As this article has explored, NFTs have the capacity to challenge and shift traditional notions of ownership and the limits of this trend are currently unknown. As to whether and how copyright responds remains to be seen.

---

[121] Edward Lee, 'NFTs As Decentralized Intellectual Property' (Forthcoming) 2023 *University of Illinois Law Review*, 3.

[122] Ibid 46. Also see generally, Molly Marias, 'I Want My NFT!: How a NFT Creative Commons Parallel Would Promote NFT Viability and Decrease Transaction Costs in NFT Sales' (Forthcoming) 12(1) *NYU Journal of Intellectual Property & Entertainment Law.*

[123] Lee (n 121) 47.

[124] Fairfield (n 8) 1299-1312.

[125] Lawrence J Trautman, 'Virtual Art and Non-Fungible Tokens' (2022) 50(2) *Hofstra Law Review* 361, 424-425.

SMES AND EXPLAINABLE AI: AUSTRALIAN CASE STUDIES

EVANA WRIGHT,[*] JIANLONG ZHOU,[†] DAVID LINDSAY,[‡]
LINDA PRZHEDETSKY,[§] FANG CHEN[**] AND ALAN DAVISON[††]

ABSTRACT

There is little understanding of the difficulties small to medium enterprises (SMEs) encounter in ensuring that the AI systems they develop, or use, are ethical. SMEs' are less likely than larger businesses to have the resources or time to familiarise themselves with ethical AI principles or how those principles should be applied in practical contexts. This paper reports the results of qualitative research conducted with Australian SMEs and start-ups that design and/or utilise AI technologies as part of their core business practices with a focus on the principle of explainability. The study identified a high level of inconsistency in both attitudes to ethical AI and to practices for implementing ethical AI within businesses in the interviewed SMEs. The paper identifies initiatives that may be implemented to promote greater understanding by SMEs of ethical AI principles, in particular, the principle of explainability.

CONTENTS

[*] Senior Lecturer, Faculty of Law, University of Technology Sydney. The authors acknowledge the valuable contribution of Dr Althea Gibson to the research project in her capacity as Research Assistant. The research outlined in this paper was conducted according to UTS ethics approval UTS HREC ETH21-6172.

[†] Associate Professor, Data Science Institute, Faculty of Engineering and IT, University of Technology Sydney.

[‡] Professor, Faculty of Law, University of Technology Sydney.

[§] PhD Candidate, University of Technology Sydney.

[**] Distinguished Professor, Data Science Institute, Faculty of Engineering and IT, University of Technology Sydney.

[††] Professor, Faculty of Arts and Social Sciences, University of Technology Sydney.

I INTRODUCTION

Micro, small and medium-sized enterprises ('SMEs') are the 'economic backbone' of many countries and economies.[1] For instance, SMEs represent 99 per cent of all business in the European Union[2] and create nearly two-thirds of new private sector jobs in the USA.[3] However, due to fewer financial resources and the prevalence of economies of scale, SMEs are typically slower to adapt to information and communication technologies than larger companies.[4] Consequently, in the context of significant recent advances in Artificial Intelligence (AI), and its wide-scale deployment, governments and international bodies have recognised the need for special measures to support the adoption and use of AI systems by SMEs.[5]

For example, in its 2020 *Recommendation on Artificial Intelligence,* the OECD emphasised the importance of national AI policies and international cooperation paying 'special attention' to SMEs,[6] including support to facilitate the ethical and trustworthy development, implementation and use of AI.[7] The European Commission's 2021 proposal for an *AI Act* points to the importance of removing barriers to the adoption of AI by SMEs and the need for national governments to develop initiatives targeted at small-scale providers and users of AI systems, including awareness-raising initiatives.[8] Moreover, in Europe, several states have established 'Digital Innovation Hubs' where SMEs can access technical expertise and experiment with AI technologies, and the EU and member states have committed to investing € 1.5 billion to roll out the hubs further.[9] Similarly, the Australian *AI Action Plan,* released in June 2021, incorporates the establishment of a National AI Centre,

---

1 Emil Blixt Hansen and Simon Bøgh, 'Artificial intelligence and internet of things in small and medium-sized enterprises: A survey' (2021) 58 *Journal of Manufacturing Systems* 362, 362.

2 European Commission, 'Internal Market, Industry, Entrepreneurship and SMEs: SME definition', (Web Page) <https://ec.europa.eu/growth/smes/sme-definition_en>.

3 Office of the United States Trade Representative, 'Small and Medium-Sized Enterprises' (Web Page) <https://ustr.gov/trade-agreements/free-trade-agreements/trans-pacific-partnership/tpp-chapter-chapter-negotiating-8>.

4 Organisation for Economic Cooperation and Development, *ICT, E-Business and Small and Medium Enterprises* (OECD Digital Economy Papers No 86, 2004).

5 See, e.g, House of Lords, Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and able?* (Technical Report, 2018).

6 Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence* (2019) OECD/LEGAL/0449, rec V.

7 Ibid.

8 European Commission, *Proposal for a Regulation of the European Parliament and of the Council: Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, Doc No COM(2021) 206 final, 21 April 2021 *('Proposed AI Act')*, Recitals (72) – (73).

9 European Commission, *Fostering a European approach to Artificial Intelligence,* COM(2021) 201 5 final, 21 April 2021, 22.

which will specifically address barriers facing SMEs in adopting and developing AI, as well as four AI and Digital Capability Centres, which are intended to assist SMEs in accessing AI technologies and expertise.[10]

Despite an understandable focus on the problems facing SMEs in adopting and using AI systems, this has yet to be matched by a similar level of attention to the particular difficulties SMEs encounter in ensuring that the AI systems they develop or use are ethical. Given SMEs' constraints, they are less likely than larger businesses to have the resources or time to familiarise themselves with ethical AI principles or how those principles should be applied in practical contexts. If, however, SMEs are to deploy AI systems successfully, they need to be adequately equipped to address the challenges of ensuring that the systems, and the way they are used, comply with ethical principles.

A pre-requisite for promoting ethical AI for SMEs is building an understanding of the current level of knowledge, and engagement with, ethical AI by SMEs. This paper reports the results of qualitative research conducted with Australian SMEs and start-ups that design and/or utilise AI technologies as part of their core business practices. The research was specifically directed at investigating the understanding among SMEs of ethical issues relating to the explainability of AI systems, which is one of the foundation issues in ethical AI. The overall objective of the research was to provide a baseline of information relating to the approaches and attitudes of SMEs to ethical AI to be used as part of the overall project of translating ethical AI principles into practice. The study identified a high level of inconsistency in both attitudes to ethical AI and to practices for implementing ethical AI within businesses in the interviewed SMEs. Ethical considerations are often viewed as secondary to other business priorities, and additional resources are required to support the adoption of ethical AI principles by SMEs in Australia.

II The Ethical Principle of 'Explainability'

One of the fundamental problems posed by the significant advances in non-symbolic or statistical AI systems is that – due to reliance on complex 'black box' functions - it can be difficult or impossible to determine how an output is produced.[11] It is, therefore, unsurprising that the majority of the many statements of principles of ethical AI incorporate a version of the principle that, in certain contexts, it must be possible for AI systems to be satisfactorily explained to humans. For example, in one of the most commonly cited surveys of AI ethical principles, Jobin et al. concluded that transparency was the 'most prevalent principle'.[12]

There are, however, many variations in how this principle is expressed; and seemingly intractable terminological confusion. For example, the European Commission's High Level Expert Group (HLEG)'s *Ethics Guidelines for Trustworthy AI* incorporates the principle of 'explicability' as one of four fundamental ethical AI principles, but when operationalising the principle, it effectively translates it into the practical requirement of 'transparency'.[13] According to the HLEG, the principle of 'explicability' means 'that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions

---

[10] Australian Government, *Australia's AI Action Plan* (June 2021) 12.

[11] See, e.g., Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez and Javier Del Ser et al., 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI' (2020) 58 *Information Fusion* 82.

[12] Anna Jobin, Marcello Ienca and Effy Vayena, 'The global landscape of AI ethics guidelines' (2019) *Nature Machine Intelligence* 389, 391.

[13] European Commission High-Level Expert Group on Artificial Intelligence, *Ethical Guidelines for Trustworthy AI* (2019).

– to the extent possible – explainable to those directly and indirectly affected'.[14] The HLEG Guidelines further divide the requirement of transparency into the following three elements:

(i) traceability—the ability to 'trace back' the data, model, rules and recommendations of an AI system;
(ii) explainability—the ability to 'explain both the technical processes of the AI system and the reasoning behind the decisions or predictions that the AI system makes'; and
(iii) open communication about the limitations of the AI system—this includes advising users that they are interacting with an AI system and informing them of the purpose, criteria and limitations of the decisions generated by the system.[15]

Australia's *Artificial Intelligence Ethics Framework,* on the other hand, treats the principles of transparency and explainability as effectively co-extensive, providing that:

> There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.[16]

In effect, the Australian framework reduces the principles to a requirement of 'responsible disclosures', which 'should be provided in a timely manner, and provide reasonable justifications for AI system outcomes'.[17]

The term 'explainability' (or sometimes 'explicability') is part of a cluster of related terms, including 'transparency', 'interpretability', 'understandability' and 'intelligibility', which are sometimes used interchangeably and sometimes distinguished. The terms are commonly organised hierarchically, with 'transparency' often appearing as an umbrella term. For example, commenting on the approach taken in statements of ethical AI principles, Jobin et al. observed that '[r]eferences to transparency comprise efforts to increase explainability, interpretability or other acts of communication and disclosure'.[18] This usage is reflected in the European Commission's proposed *AI Act,* which uses 'transparency' as an umbrella term, requiring that '[h]igh risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately'.[19]

However, terms other than transparency also appear as umbrella terms. After noting the inconsistent use of terminology, a 2019 UK House of Lords committee report adopted the general term 'intelligibility' to apply to both 'technical transparency' and 'explainability'.[20] According to the report, 'technical transparency' means ensuring that experts understand an AI system, including how and why outputs are produced. On the other hand, the report confined 'explainability' to ensuring that AI systems 'are developed in such a way that they can explain the information and logic used to arrive at their decisions'.[21] In a 2018 paper, however, Floridi et al. had a different take on 'intelligibility': the paper applies the catch-all term 'explicability' in both the epistemological sense of 'intelligibility', relating to how an AI

---

[14] Ibid 13.

[15] Ibid 14–15.

[16] Australian Government, *Australia's AI Ethics Principles* (7 November 2019).

[17] Ibid.

[18] Jobin Ienca and Vayena (n 12).

[19] *Proposed AI Act* (n 8) art 13(1).

[20] House of Lords, Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and able?* (Technical Report, 2018) 36.

[21] Ibid 39.

system works, and the ethical sense of 'accountability', relating to responsibility for how an AI system works.[22]

This distinction between epistemological and ethical usages partially explains the terminological confusion: it is one thing to focus on the technical features of an AI system and quite another to focus on human understanding of, or responsibility for, technical systems. Another source of the confusion is the extent to which the problem is approached from different disciplinary perspectives, including computer science, human-computer interaction, psychology and law.[23] This paper returns to the problem of how to more precisely formulate the principle of explainability after introducing three specific issues arising from ethical obligations to provide an explanation: the circumstances in which an explanation may be required; the context-dependent nature of explanations; and the potential for trade-offs between explainability and accuracy. In these sections, we use the term 'explainability' indiscriminately to encompass other related terms, such as 'interpretability'.

A *When is an Explanation Required?*

An important threshold question when considering the explainability principle is: when should an explanation be required? Different approaches may be taken to specify the circumstances in which an explanation is needed. Most approaches apply a framework based on the assumption that the higher the risk posed by an AI system, the greater the need for an explanation. There are, however, differences in conceptualising risks. It has, for example, been argued that the obligation to provide an explanation should depend on the domain in which the AI system is used. For instance, the UNESCO Ad Hoc Expert Group on the Ethics of AI states that the principle of explainability is more important when an AI system is used in a high-risk 'domain' such as law enforcement, security, education, recruitment or health care.[24]

Other approaches define risks by reference to the impacts on affected persons, and especially impacts that affect an individual's rights or interests. For example, the European Commission's proposed *AI Act* defines 'high risk' AI systems as 'systems that pose significant risks to the health and safety or fundamental rights of persons'.[25] Similarly, the Australian Human Rights Commission (AHRC) has made the point that 'it is good practice to provide reasons for decisions that affect a person's legal or similarly significant rights, regardless of the status of the decision maker and even where there is no legal requirement to provide reasons'.[26]

The considerable difficulties in specifying the circumstances in which an explanation may be required, which extends to difficulties in predicting risks, emphasises the need for most AI systems to be potentially explainable. That said, counter-veiling considerations must be considered in both imposing obligations to provide an explanation and determining the

---

[22] Luciano Floridi, Josh Cowls, Monica Beltrametti et al, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28(4) *Minds and Machines* 689, 700.

[23] Amina Adadi and Mohammed Berrada, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' (2018) 6 *IEEE access* 52138; Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 *Artificial Intelligence* 1.

[24] UNESCO Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, *First Draft of the Recommendation on the Ethics of Artificial Intelligence* SHS/BIO/AHEG-AI/2020/4 REV.2 (7 September 2020) [71], 14.

[25] *Proposed AI Act* (n 8) 1.1.

[26] Australian Human Rights Commission, *Human Rights and Technology* (Final Report 2021) ('*Human Rights and Technology*') 81.

form an explanation should take. First, as explained further below, there may be trade-offs in design decisions in developing AI systems between explainability and other values, such as efficiency or accuracy. Secondly, both designing explainable systems and explaining AI systems imposes costs.[27] Thirdly, complete transparency may be undesirable to the extent to which it may result in the release of trade secrets or personal data.[28]

This indicates that considerable care is needed in determining the circumstances in which an explanation should be provided and the form an explanation should take.

B *Explanations Necessarily Depend upon Context*

It is widely accepted that explanations of AI systems are highly contextual. For example, in their guidance on explaining decisions made with AI, the UK Information Commissioner's Office (ICO) and the Alan Turing Institute note that several factors, such as the application of the AI system, the type of data involved, the setting, and the individual recipient of the explanation all 'affect what information an individual expects or finds useful'.[29] The guidance further points out that organisations should tailor explanations to their audience so that they 'avoid creating explanation fatigue … (by saying too much) and at the same time allow … [organisations] to protect … [their] intellectual property and safeguard [their] system from being gamed'.[30]

Similarly, Preece et al. argue that the question of whether AI is explainable cannot be answered before answering the question 'explainable to whom?' Accordingly, they point out that explainability means different things to system creators, system operators, those making decisions based on AI systems, those affected by AI decisions, and those whose data has been used in AI systems and system regulators.[31] Dawson et al. take this further by pointing to the different purposes of different audiences. Accordingly, they observe that while explanations for users of AI systems may focus on what the system is doing and why, explanations for creators may aim to explain how the system is working for validation or certification, and explanations for the general public may aim to build user trust and confidence in AI systems.[32]

Similarly, in a detailed discussion, Zhou and Danks distinguish the goals of different groups, which they label 'engineers', 'users' and 'affectees'.[33] For example, they argue that an 'affectee' (such as a person whose face is recognised by a facial recognition system) may simply require a non-technical 'difference-based intelligibility', namely 'an input-output characterization of the decision processes embodied by the algorithm, thereby reducing uncertainty and demonstrating the reliability of the system'.[34] On the other hand, they claim that 'users' of an AI system require 'function-based intelligibility', which entails more

---

[27] Adadi and Berrada (n 23).

[28] *Human Rights and Technology* (n 21) 66.

[29] Information Commissioner's Office and The Alan Turing Institute, *Explaining decisions made with AI* (20 May 2020), 21.

[30] Ibid 49.

[31] Alun Preece, Dan Harborne, Dave Braines et al, 'Stakeholders in Explainable AI' (Conference Paper, AAAI FSS-18: Artificial Intelligence in Government and Public Sector Conference, 2018) <https://arxiv.org/abs/1810.00184v1>.

[32] D Dawson, E Schleiger, J Horton et al, *Artificial Intelligence: Australia's Ethics Frameworks* (Data61 CSIRO, 2019).

[33] Yishan Zhou and David Danks, 'Different "Intelligibility" for Different Folks' (Conference Paper, AAAI/ACM Conference on AI, Ethics and Society, 7-8 February 2020).

[34] Ibid 196.

information about the operation of the AI system, including information about the inputs it requires to operate and 'appropriate conditions for its use and adaptation'.[35] This information can be supplied by AI developers in design documents. Finally, AI engineers require 'causal-process intelligibility', which is the type of intelligibility that has been the focus of much of the scholarly literature.[36] This involves information about 'computational architecture, specific models, parameter values, internal states and their relationships … [and] hardware or user interface constraints'.[37]

Apart from distinctions based on the purposes of the recipients of an explanation, distinctions have been drawn between different types of explanation. For example, the ICO and the Alan Turing Institute's guidance on explainability notes that explanations can be either 'process-based' - that is, they can provide information demonstrating responsible design and deployment of an AI system - or 'outcome based'. The guidance identifies six main types of explanation: rationale explanations; responsibility explanations; data explanations; fairness explanations; safety and performance explanations; and impact explanations.[38] According to the guidance, the particular type of explanation that may be required depends on the use of the AI system. For instance, if an AI system was used to process applications for a job, an unsuccessful applicant may wish to know that they have not been discriminated against (i.e., they may require a 'fairness explanation'). Alternatively, a patient who has received a medical diagnosis generated by an AI system will wish to know that the diagnosis is accurate (a 'safety and accuracy explanation').[39] In yet another taxonomy, computer science scholars Vilone and Longo note that the existing research on the explainability of AI systems has identified the following different types of explanations intended to fulfil different purposes:[40]

- Traced-based explanations—for system designers
- Reconstructive explanations—for end users
- Mechanistic explanations—how does it work
- Operational explanations—how do I use it
- Ontological explanations—describe the structural properties of the system
- Teaching explanations
- Introspective tracing explanations
- Introspective informative explanations
- Post hoc explanations
- Execution explanations.

Regardless of the 'type' of explanation or the recipient's purpose, the quality of the decision will dictate the extent to which a decision is 'explainable'. Meske et al. observe that consideration of the quality of an explanation involves consideration of multiple factors, such as plausibility, comprehensibility, interpretability, fairness, and privacy.[41] Gilpin et al. have

---

[35] Ibid 197.

[36] Ibid.

[37] Ibid.

[38] Information Commissioner's Office and The Alan Turing Institute, *Explaining decisions made with AI* (20 May 2020) 20.

[39] Ibid 52-54.

[40] Guilia Vilone and Luca Longo, 'Explainable Artificial Intelligence: A Systematic Review' (2020) <https://arxiv.org/abs/2006.00093>.

[41] Christian Meske, Enrico Bunde, Johannes Schneider and Martin Gersch, 'Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities' (2022) 39(1) *Information Systems Management* 53, 57-58.

linked the quality of an explanation to the level of understanding of the recipient of such an explanation, claiming that a good explanation has been provided when the recipient 'can no longer keep asking why'.[42]

Using examples and making relationships or causal links explicit will also improve the quality of an explanation.[43] Vilone and Longo observe that 'it is part of human nature to assign causal attribution of events' and, as such, explanations of AI systems 'must make the causal relationships between the inputs and the model's predictions explicit'.[44] Similarly, Graaf and Malle argue that, as people often regard AI systems as operating with human-like intention, it is important that explanations fall within 'the bounds of the conceptual and linguistic framework' used to explain human behaviours.[45] After reviewing over 250 social science publications on explanations, Miller concluded that explainable AI researchers should consider that effective explanations are generally selective, 'contrastive' (that is, they should explain why one event happened instead of another) and focus more on causal links than on probabilities.[46]

In their guidance, the ICO and the Alan Turing Institute advise that organisations can 'layer' their explanations by first providing individuals with priority explanations, and then making additional explanations available in further layers.[47] The guidance also advises that explanations should be conceptualised as a two-way conversation and use visual aids such as 'visualisation media, graphical representations, [or] summary tables' where appropriate.[48]

C *The Trade-off Between Explanation and Accuracy*

It is commonly argued that there is a correlation between the accuracy and complexity of AI systems and that this has a negative impact on the 'explainability' of a decision. As Burrell observes, '[m]achine learning models that prove useful (specifically, in terms of the 'accuracy' of classification) possess a degree of unavoidable complexity'.[49] For this reason, many commentators observe that there is an unavoidable trade-off between performance and explainability - '[o]ften, the highest performing methods (e.g., DL [deep learning]) are the least explainable, and the most explainable (e.g., decision trees) are the least accurate'.[50] The design of an AI system will dictate the extent to which a system is explainable, and 'trade-offs

---

42 Leilani H. Gilpin, David Bau, Ben Z. Yuan et al 'Explaining Explanations: An Overview of Interpretability of Machine Learning' in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics* (IEEE, 2018) 80–89.

43 Katie Atkinson, Trevor Bench-Capon and Danushka Bollegala, 'Explanation in AI and law: Past, present and future' (2020) 289 *Artificial Intelligence* 103387.

44 Vilone and Longo (n 40) 8.

45 Maartje M. A. de Graaf and Bertram F Malle 'How People Explain Action (and Autonomous Intelligent Systems Should Too)' in *2017 AAAI Fall Symposia* (AAAI Press, 2017) 19–26.

46 Tim Miller, 'Explanation in artificial intelligence: Insights from the social sciences' (2019) 267 *Artificial Intelligence* 1.

47 Information Commissioner's Office and The Alan Turing Institute, *Explaining decisions made with AI* <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/>.

48 Ibid.

49 Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016) 3(1) *Big Data & Society* 1, 5.

50 David Gunning, Mark Stefik, Jaesik Choi et al 'XAI – Explainable artificial intelligence' (2019) 4(37) *Science Robotics* <doi: 10.1126/scirobotics.aay7120>; See also, Philipp Hacker, Ralf Krestel, Stefan Grundmann and Felix Naumann, 'Explainable AI under contract and tort law: legal incentives and technical challenges' (2020) 28(4) *Artificial Intelligence and Law* 415.

might have to be made between enhancing a system's explainability (which may reduce its accuracy) or increasing its accuracy (at the cost of explainability).'[51]

However, not all scholars agree that such a trade-off is inevitable. Increased explainability may, in fact, lead to increased accuracy to the extent that it helps lead to the correction of deficiencies in AI systems.[52] Cynthia Rudin goes as far as to state that the existence of any trade-off between explainability and accuracy is a myth.[53] Furthermore, she argues that the reliance on the post-facto explanation of high-stakes decisions by complex systems may be inadequate and, in fact, 'perpetuate bad practice and … potentially cause great harm to society.'[54] Instead, the focus should shift to the design of interpretable systems that will 'provide their own explanations, which are faithful to what the model actually computes.'[55] Using the surprising outcome of the 2018 Explainable Machine Learning Challenge as an example, Rubin and Radin argue that simple, interpretable models may be as accurate as more complex models while also avoiding some of the other issues that arise in relation to black box systems.[56]

III SMEs and Explainability

The literature on how SMEs approach the issue of AI ethics and explainability is limited. A recent study by Ayling found that surveyed SMEs did not view explainability as necessary beyond 'communicating with their customers about their products' as part of their sales process.[57] In contrast, Bessen et al., following a survey of 225 AI start-ups, found that 58% of surveyed companies had established a set of codified firm-level ethical AI principles but that many of these companies 'have never invoked their ethical AI principles in a costly way, such as firing an employee, dropping training data, or turning down a sale.'[58] The study established that resources are critical to the adoption and implementation of AI ethics principles. It also found that larger start-ups, companies that had collaborated with high-technology firms and companies with prior experience in implementing GDPR obligations, were more likely to have established ethical AI frameworks.[59]

In the absence of specific literature considering how SMEs approach the issue of explainability, research on general business responses to AI may provide further insights into SMEs and the adoption of AI ethics principles. In a small study of nine executive managers of businesses (of varying sizes and in a range of sectors) across Germany, Austria and Scandinavia, interviews revealed that 77.77% of interviewed managers believed that AI ethics

---

[51] European Commission High-Level Expert Group on Artificial Intelligence, *Ethical Guidelines for Trustworthy AI* (2019) 18.

[52] Arrieta et al (n 11) 83.

[53] Cynthia Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead' (2019) 1(5) *Nature Machine Intelligence* 206, 207.

[54] Ibid 206.

[55] Ibid.

[56] Cynthia Rudin and Joanne Radin, 'Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition' (22 November 2019) 1(2) *Harvard Data Science Review* <https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/8>.

[57] Jacqueline Ayling, 'Putting AI ethics to work: Are the tools fit for purpose for SMEs?' (PhD Thesis, University of Southampton, 2021) 78, 101.

[58] James Bessen, Stephen M. Impink, Lydia Reichensperger and Robert Seamans, *Ethical AI Development: Evidence from AI Startups* (Working Paper, March 2022) <https://scholarship.law.bu.edu/faculty_scholarship/1188> 4, 5.

[59] Ibid 14, 17.

should be a high priority in their business.[60] However, while interviewees indicated that transparency was an important value in their business practices, some of these managers were concerned that revealing information could compromise intellectual property rights and competitive advantage.[61]

Literature suggests that the successful adoption of ethical AI by businesses requires organisational transformation. For example, an international survey of 1580 executives in 510 large companies conducted by Capgemini Research Institute showed that 77% of executives are uncertain about the ethics and transparency of their AI systems.[62] The same survey showed that concern about using AI systems influenced strategic business decisions: '41% of senior executives report that they have abandoned an AI system altogether when ethics concerns were raised; 55% implemented a "watered-down" version of the system'.[63] The Capgemini report is important for this paper for two reasons: one, it is likely that executives in SMEs are also uncertain about AI ethics and explainability, and two, the adoption of AI systems and thus purchasing decisions by potential customers, may be influenced by the extent to which AI systems are explainable.

Explainability is especially important to SMEs for many reasons. As discussed above, at a macro level, explainability helps ensure that AI systems remain accountable – that is, they can be audited and their accuracy assessed. As argued by Bauer et al., explainability also assists humans to 'widen their horizons of reasoning and understanding'.[64] From a business perspective, explainability helps to ensure customer trust and satisfaction.[65] KPMG notes that 'organisations must think about the governance of algorithms to build trust in outcomes and achieve the full potential of artificial intelligence'.[66] Failure to ensure AI systems are explainable may expose SMEs to 'financial, reputational, and regulatory risks'.[67]

Customer satisfaction is particularly vital to SMEs, which depend on repeat business far more than multinational enterprises, which tend to have large customers.[68] Further, SMEs have fewer resources to invest in technology and governance systems. By ensuring that AI systems are explainable from the outset, SMEs can minimise the risk of having to update or upgrade systems to provide explainability. The lack of financial resources also makes it imperative that SMEs do not unnecessarily expose themselves to reputational loss or legal liability. For example, the OECD has flagged 'reputational and legal risks' as one of several

---

[60] Joseph Baker-Brunnbauer, 'Management perspective of ethics in artificial intelligence' (2021) 1 *AI and Ethics* 173, 177.

[61] Ibid 180.

[62] Capgemini Research Institute, 'Why addressing ethical questions in AI will benefit organisations' (4 July 2019) 9.

[63] Ibid 2.

[64] Kevin Bauer, Oliver Hinz, Wil van der Aalst and Christof Weinhardt, 'Expl(AI)n It To Me – Explainable AI and Information Systems Research' (2021) 63(2) *Business and Information Systems Engineering* 79, 80.

[65] See, e.g., IBM, 'Transparency and trust in the cognitive era' *IBM THINK* (Blog Post, 17 January 2017) <https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/>; Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky et al, 'The role of trust in automation reliance' (2003) 58(6) *International Journal of Human Computer Studies* 697.

[66] KPMG, *Uncover the full potential of artificial intelligence* (2019) <https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/uncover-the-full-potential-of-artificial-intelligence.pdf> 2.

[67] Ilana Golbin, Anand S. Rao, Ali Hadjaran et al, 'Responsible AI: A Primer for the Legal Community' in *2020 IEEE International Conference on Big Data* (IEEE, 2020) 2121, 2123.

[68] Dóra_Horvath and Roland Zs. Szabo, 'Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?' (2019) 146 *Technological Forecasting and Social Change* 119.

barriers to SMEs' use of data analytics and implementation of data solutions.[69] Accountability of AI systems, including explainability, may also be an important consideration in any potential future investment by venture capital or other external investors with ethical AI 'seen as one of the important drivers of portfolio risk and return.'[70] Potential collaborators may also require compliance with AI ethics principles, particularly with established '[h]igh technology firms, which often share their data sources with startups'.[71]

As discussed above, the OECD acknowledges that SMEs require special support to facilitate the ethical and trustworthy development, implementation and use of AI.[72] The question remains as to how SMEs may implement ethical principles in practice and how SMEs should be supported to do so. The remainder of this paper outlines the methodology and findings of qualitative research conducted with Australian SMEs and start-ups investigating the understanding among SMEs of ethical issues relating to the explainability of AI systems.

IV METHODOLOGY

The study was designed to elicit in depth qualitative information about the views, challenges and expectations of Australian SMEs[73] concerning AI ethics, with a focus on the explainability of AI systems. It was conducted in two main stages: semi-structured interviews with selected participants and a follow-up survey questionnaire.

The participants in the study were chosen based on pre-selection criteria, which ensured that they were well-placed to answer questions on the use of AI systems in their respective businesses. Most participants interviewed were involved in key business decisions relating to the selection, design, implementation, use and/or evaluation of AI technologies in their businesses. Although every participant either designed or used AI in their business, not every business intended to use AI when the business began; in several cases, businesses started using AI after they had already commenced providing a product or service.

Due to the small size of the businesses, which included start-ups, interviewees often noted that team roles and responsibilities overlapped prior to the business expanding sufficiently to engage specialists responsible for technology-related decisions. Business founders and key decision-makers came from a range of backgrounds, including business, finance, law, computer science and academia; however, every participant interviewed had, at a minimum, a baseline understanding of how AI worked in their business. The selection of a

---

[69] Organization for Economic Cooperation and Development, *The Digital Transformation of SMEs* (OECD Studies on SMEs and Entrepreneurship, 2021) ch 5.

[70] AI Asia Pacific Institute, *Transforming Ethics in AI Through Investment* (Web Page, 7 February 2021) <https://aiasiapacific.org/2020/08/20/transforming-ethics-in-ai-through-investment/>.

[71] Bessen et al (n58) 4.

[72] Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 22 May 2019, s 2.

[73] There is no universal agreement as to what constitutes an SME, with definitions varying between countries and even industries. The European Commission defines an SME as an organisation with less than 250 employees and either an annual turnover of less than €50 million or a balance sheet of less than €43 million. See *Commission recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises* [2003] OJ L 124/36, art 2(1). By contrast, in the US the definition of an SME varies according to the industry in which the enterprise operates. See US Small Business Administration, 'Table of Small Business Size Standards Matched to North American Industry Classification Codes' (Web Page, 14 July 2022) <https://www.sba.gov/sites/default/files/2022-07/Table%20of%20Size%20Standards_Effective%20July%2014%202022_Final-508.pdf>. In Australia, federal legislation contains a number of different definitions of 'small businesses'. See e.g., Australian Government, *Australian Securities and Investments Commission Act 2001* (Cth), ss 12BC, 12BF; *Corporations Act 2001* (Cth) s 761G.

range of professional backgrounds was designed to ensure the broadest possible understanding of AI systems, including the benefits and risks of the technologies.

Eight participants took part in the open-ended interviews, while five completed the follow-up survey. Although this was admittedly a small sample, the study resulted in a rich collection of qualitative material concerning attitudes to ethical AI and the understanding of explainability among Australian SMEs.

A *Semi-structured Interviews*

The semi-structured interviews, based on a series of open-ended questions, consisted of Zoom sessions of approximately one hour. The interviews were subsequently transcribed using a transcription service.

The open-ended questions, which were designed to form part of a free-flowing discussion, investigated the following issues:

- the professional background of the interviewee, including their role in the business
- the nature of the business, including its clients or customers
- the use of AI in the business, including how AI is developed and/or used in products or services
- the participant's understanding of 'explainability' in the context of the use of AI in the business
- whether the 'explainability' of AI was regarded as important by the business
- whether, in what circumstances, and how, the business explained its AI systems
- what was entailed by an explanation
- the use of data in AI systems and the safeguards, such as documentation, applied by the business to the use of data
- the implications of the use of data for explainability
- the use of the AI model (or algorithm) by the business and the safeguards, such as documentation, applied by the business to the model
- understanding of any potential trade-offs between explainability and accuracy
- the business risks entailed in providing an explanation, including potential disclosure of trade secrets
- responsibility within the organisation for AI governance, including responsibility for ensuring ethical AI and addressing potential problems
- organisational understanding of ethical AI, especially as applied to SMEs
- particular areas of uncertainty in complying with ethical AI
- whether there is a specific need for guidance about ethical AI for SMEs.

B *Follow-up Survey*

Analysis of the responses to the semi-structured interviews revealed areas that required clarification or further exploration. Participants were therefore asked to complete a follow-up online survey approximately four weeks following the initial interviews. The follow-up survey canvassed the following issues:

- the risks associated with providing explanations of AI systems, including the potential for revealing trade secrets
- views about the potential role of government in ensuring ethical AI
- the most effective tools or methodologies for promoting ethical AI
- how the participant/SME interpreted 'explainability'
- where the business would be likely to seek guidance about ethical AI

- if an ethical AI framework were to be mandated, who should be responsible for creating the framework?
- which of the existing ethical AI frameworks would be preferred?
- the techniques or methodologies preferred by the business for ensuring explainability and ethical AI
- tolerance for risk associated with practices that may fail to ensure explainable AI

The responses to the follow-up survey added valuable detail to the information collected from the qualitative interviews.

V FINDINGS

Following analysis of the qualitative interviews and the follow-up survey, the responses were grouped into the following overarching themes:

- organisational knowledge and awareness of AI ethics
- problems with data quality
- approaches to selecting and implementing AI models
- attitudes to AI assurance processes, including documentation and auditing
- organisational AI governance mechanisms
- approaches to understanding 'explainability' and implementing explainable systems
- understanding of potential trade-offs between 'explainability' and accuracy
- concerns about risks of explaining AI systems
- approaches to 'explainability for SMEs'
- approaches to the role of government in ensuring ethical AI

Each of these themes is expanded upon immediately below.

A *Knowledge and Awareness of AI Ethics*

The interviews revealed considerable variation in the level of knowledge and understanding the SMEs expected employees to have concerning ethical AI. While some businesses suggested that their staff inherently knew about AI ethics ('the machine learning guys all know the consequences of getting some of the stuff wrong'), some expected staff who join the business to have learnt about AI ethics through formal study ('people who are involved on the AI part have at least a master's of data science and AI from one of the good Australian universities, so it's part of their curriculum in any case. So we are not training them; we expect them to have that insight and training before they come and work with us').

There were, however, notable differences in the level of in-house training provided on ethical AI: while some proactively trained staff in AI ethics ('so as part of the onboarding we have a section on talking about data and we … [consider that] … the team we have should treat that data as our own data'), some were completely unaware of ethics training ('I've haven't had or seen any ethics training'), while others failed to see the value of AI ethics training ('obviously, no amount of training can teach someone to be ethical').

B *Data Quality*

Statistical AI systems, including machine learning systems, are highly dependent on the quality of the data sets used to train the systems. Yet leveraging data was critical to most of the businesses that participated in this study. For example, the business model adopted by a number of the SMEs is based on deriving value from previously underutilised data. As one

interviewee put it, 'data is a pure margin product', adding that much data is effectively costless to collect.

Nevertheless, a strong theme emerging from the interviews is that SMEs implementing AI systems face considerable challenges with data quality. The difficulties encountered are illustrated by the following comments:

- 'the data is notoriously bad quality'.
- 'we had quite a lot of issues with just dealing with the varying quality of imagery that we had'.
- '[the data was] poor quality so we had to kind of get rid of the whole data set'.
- 'it's the only dataset we've got, so it's either going to work or it isn't'.

Overall, the responses indicated more focus on the practical difficulties of 'wrangling' data than on ethical dilemmas relating to low quality data. That said, there was a general awareness of the need for exercising caution in working with data, with one interviewee observing, in relation to the selection of data for a machine learning system, that 'you just got to be careful how you put that into a model because it re-weights everything in its own way.'

C *Selecting and Implementing AI Models*

The SMEs that formed part of this study applied a variety of approaches to selecting and implementing AI models, with some building their own models, others modifying existing models, and yet others using 'off the shelf' solutions. The choice taken depended upon the accessibility of a suitable model for the task at hand. For example, as one interviewee explained the choice of models:

- 'We had [Microsoft] Azure Cognitive Services available to us [for this project], [so] we just let Microsoft build the detailed models. So in this case, we have not built the AI models, but in previous cases where we had done other work around deep learning algorithms, that's where we had a decision, and primarily we used an ensemble learning methodology [which combines multiple models]. So we just ran multiple algorithms and then checked out what results were most commensurate with what we wanted'.

The extent to which a business uses an 'off the shelf' model or build their own has important implications for the explainability of the AI system. For those businesses with the in-house technical skills to build their own models, model development was strongly influenced by literature reviews. The comments on model development included the following:

- 'We go and do a scan of I guess, literature - a bit like … a research project - do a mini literature review and come back and say, "Well, look. This is what from a research perspective the latest practice looks like or the latest research tells us is relevant."'
- 'We didn't build off others. We did look at literature reviews…'.
- 'I guess, [our model is] based on research about what sort of literature review of what models are being used in this general area trying to pull together combinations of natural language and existing building research'.

D *AI Assurance Processes*

The explainability of AI systems, not to mention the safety and reliability of systems, depends upon the implementation of assurance processes, such as appropriate documentation and system auditing. The study was therefore interested in the attitudes of SMEs to implementing

internal assurance processes. There was a high degree of inconsistency in the approaches taken to documentation of AI systems, with considerable differences in the interpretation of what might be required and how documentation was implemented:

- 'We document everything internally. We use Confluence, so just like a Wiki page type of tool'.
- 'There's just one document, generally. There's a project write-up at the end of each project'.
- 'We have more like system descriptions and architecture design documentation that says these are the steps that we went through, the data cleaning steps'.

Apart from the inconsistent approaches to documentation, some SMEs did not separately document processes, with one expressing scepticism about the value of documentation. These attitudes were reflected in comments such as the following:

- 'All of the documentation as far as the model itself is concerned would be contained in the code'.
- 'We didn't document stuff, because who are we documenting it for? It's kind of a backend system'.

Among the interviewees, there was general scepticism about the value of auditing AI systems, at least in the context of current AI auditing practices:

- 'I've not met an AI auditor. I've met lots of people who make up these checklists, that haven't got a clue what's going on inside the technology'.
- 'I think the AI audit fraternity needs to upskill before it can start actually … providing a valid service around what they're looking into, because the guys I've spoken to so far just don't have a clue what's under the hood. They're just really making questions off feature lists'.
- 'At the end of the day, the audit is only as good as … the person doing it and the day that they did it on'.

These comments appear to reflect concerns that the practice of AI auditing is not well-established and that third party auditors do not have sufficient expertise to adequately assess a business's AI system. That said, some interviewees indicated that they incorporated independent expert evaluation in their system development, with one reporting that 'we have an industry expert that looks at the results and works way back'. Another interviewee reported that they had engaged one of the big four consulting firms to audit their system, which included providing the firm with access to the code.

E *AI Governance Mechanisms*

The interviews explored a range of issues relating to the internal governance mechanisms of SMEs. Understandably, the SMEs forming part of the study had flat structures, which conditioned their governance mechanisms. The majority of the interviews were conducted with the founders of the SMEs, with some being conducted with data scientists. Typically, the founders had considerable technical expertise and were initially either closely involved with, or intimately aware of, the governance of AI systems used in their business.

In most cases, given the flat structures, the founders or CEOs were regarded as ultimately responsible for the AI systems deployed by the business. The following comment was typical: '(t)he buck kind of stops with me in terms of the ownership of whether or not it's ethical'. However, where data scientists were employed in-house, they commonly shared responsibility, with one data scientist commenting, 'I'm responsible ultimately for any changes or code that gets deployed into a production environment'. Referring to the shared

responsibility, another data scientist observed, '(t)hey [ie. management] shouldn't be asking for something that's unethical and I shouldn't be giving it to them'.

Regardless of perceptions of ultimate responsibility, the scale of the businesses meant that there was considerable overlap in roles and responsibilities, with a range of people assuming roles in using or modifying algorithms.

F *Approaches to Explanations*

There were significant variations among interviewees in the approaches taken to the 'explainability' of AI systems. A number of interviewees accepted that, in the context of their business, it was important to provide users with an explanation. The following are examples of comments that emphasised the importance of explanations:

- '… if you make a decision based on an algorithm's recommendation, that decision … has to be explained'.
- '… from our perspective, we have to make sure that … the way those models are working, if a human were to look at them, that it would make sense and is interpretable'.
- '…we have to really think about what we are giving the users and [about] the explanation. So we add an additional explanation…'.
- 'I think everyone should be given an explanation: everyone who is part of that journey or interacts with [the AI system] or is affected by it'.

On the other hand, some interviewees questioned whether it was always possible or desirable to provide an explanation. Comments to this effect included the following:

- '… black box algorithms like neural networks are very difficult because … [from] … the internal architecture … it is really difficult to explain how we came to that conclusion'.
- 'But does it matter? Does it matter that you can say why it came up with a certain result? I don't think it does that much'.
- 'We actually didn't do explainability. We had a black box opaque AI system …'.

The interviewees generally regarded explainability as subservient to overarching business goals. Therefore, explainability tended to be valued if it could build trust in the business. As one interviewee put it, 'explaining… why they have been recommended what they have been recommended, what the logic behind that is… [can make users] feel more assured'. Similarly, another interviewee commented:

- '… we have the human element in there, which is actually our competitive advantage. So our algorithm may not be superior but the way we present information is far superior [to] anyone else.'

Against this, as illustrated by the following comments, several interviewees considered that users or customers do not care about explainability:

- 'Query how much a client cares once a [practitioner says], 'This is good to go''.
- 'You've just got to tell them what it does. No one really cares about what's underneath it all, they care about what's the output and why that output came to be'.

Two strong themes to emerge were the perceptions that, first, providing an explanation would increase the information burdens placed on users or customers and, secondly, that users or customers are more interested in the accuracy of systems than in understanding precisely how they work. The following comments illustrate these themes:

- 'I mean, I do have the sense, more and more actually over time, that people are just experiencing tremendous information overload. And an AI system being explicable, I mean, it's just more data to grok'
- 'I think maybe in the B2C market, you can just say that you have AI and people just generally accept … [that] … because it's all about value'
- 'So I don't think just dumping them with a lot of technical jargon is going to help'.

Concerns about how users might respond to explanations also influenced the approaches taken by interviewees to the form in which explanations should be given. The various approaches taken by interviewees to how explanations should be given appear from the following comments:

- '… it would be just a human explaining the things that are clearly marked as automated. We are quite careful to mark clearly what is and isn't automated in our document selection. We have a few different categories of decision, default decisions, automated decisions, user decisions, system decisions. And we make very clear distinctions … between those'.
- 'We give a popup saying very specifically this is what it does. It … [asks]… do you still want to use it? So [we] give the users the choice of using the algorithm and so it's [an] instant thing on an app that they can yes say, "Yes, I know it's very clear." [In addition] … we give explanation videos so that gives them [some] context'.
- 'It needs to be short and sweet'. If it is pictorial, just one scan and … [they] … get it. Brilliant'.
- 'What we did instead was we showed … what … was recommended. We didn't explain why. And we also showed them how they performed on those recommendations'.
- 'They kind of just want a human being who knows the area to give them a simple explanation that's germane to them specifically. I think that's a lot of what people are paying for'.
- '… I don't know whether we need to explain the whole AI algorithm, how it works. I think it just needs to be … because you said this, we are recommending this. … Explaining the frequency and why we [are] saying it, but at a high level, rather than saying here is our algorithm and this is how it works, because not everyone needs to know and not everyone will follow. And they don't care, to be honest. They just want to know that there is some science in … this'.
- 'Look, what I try to do is I use a lot of analogies [that are used] at the moment because it's difficult for people to understand what the impact [is]'.
- '… typically what we try to do is we actually try to verbally explain [the system] and if there's any publicly available … [information] … we try to give that away and then in summarised format …'.
- 'Generally we'd like to … [provide an explanation] … face to face and talk people through, say, a presentation which then we would be happy to leave with them'.

G *Trade-offs Between 'Explainability' and Accuracy*

As the study was concerned with the value SMEs place on explainability compared with potentially competing objectives, it questioned interviewees about their approaches in the event of a trade-off between explainability and accuracy. As explained immediately above, one of the themes to emerge from the study was that interviewees generally tended to

consider that their customers valued accuracy over explainability. The following comment captures this general attitude:

- '… as long as the model is proven accurate and is working the way it wants, then generally you don't care how it came to that decision. It's similar to … [how] … a lot of people driving a car don't really care how the car works, as long as it works'.

The overarching concern among interviewees, as reflected in the following comment, seemed to be the need to assure customers of the accuracy of a system rather than trouble them with the details of how it works:

- 'The reality of it is, I think what the market wants from us is real accuracy. I have been surprised that it's like once people are paying any kind of money, the expectation of accuracy is very, very high'.

That said, some interviewees considered that providing an explanation can assist in assuring customers that the system is reliable and accurate. As one interviewee put it:

- 'And … [what] … I always just ask clients is, "is it better to have a model that you know predicts extremely well but I can't explain to you or is it better to have a model I can explain to you but doesn't predict as well?" And … they always get torn on that. ... The way I see it is that we want to try and get the most predictive model that we can, … [while] … understanding that sometimes to get people comfortable with the idea that … [the system] … is doing what is expected, we need to give them some level of explainability'.

H *Risks of Explaining AI Systems*

As reported above, most of the interviewees regarded explainability as subsidiary to the wider commercial goals of their business. This was particularly evident in general concerns that trade secrets would be compromised if too much information were to be provided about the algorithm. Comments reflecting this concern included the following:

- 'I would have concerns about sharing trade secrets, especially if it's a competitive advantage. If it's a means to an end and everyone is doing it and I'm borrowing from the world and giving it back to the world, I will have no problem. But if it's something that's proprietary, it's something that we have done and it gives us a competitive advantage, then that's something we don't want to erode'.
- 'It's very tough from an industry competition standpoint to make many of these decision-making processes comprehensible to your competitors by publicly disclosing them, or even your customers who may well be a mystery shopper from your competitor. I have real doubts, just from a purely commercial standpoint, about explaining too much how we do stuff, because I'm not really interested in competitors taking off with the ideas'.
- 'I think actually, the transparency required for people to understand it, you would have to disclose commercially sensitive trade secrets. It's a very tough problem'.

At the extreme end, one interviewee went so far as to say, that disclosing too much information to a client may mean they have 'designed a business for doomsday'.

While a minority of businesses dealt with the risks by not revealing any information about the AI system, the majority adopted strategies aimed at minimising the risks. One strategy, for example, was to provide a verbal summary ('we actually try to verbally explain it and if there's any publicly available things we try to give that away and then in summarized format…'). Another strategy was to limit the amount of information provided by the customer interface ('… we are seriously considering taking it … [explainability information]

… off the interface, because it gives away a lot to potential competitors without providing a great deal of value to people who just don't seem to want to really know'). The majority of SMEs interviewed therefore tended to regard the disclosure of information about their AI systems as a balance between revealing enough to assure customers while not revealing information that could advantage competitors.

I *Explainability for SMEs*

The interviews aimed to identify the specific issues faced by SMEs. To begin, most interviewees agreed that SMEs should comply with ethical principles. A good example of this was the observation that:

> '… there should be a requirement to comply, because you can't just say, "Well, it applies to large enterprises and then small businesses can do without it," because then that's not appropriate'.

That said, there was general agreement that SMEs face particular challenges in complying with ethical AI principles, arising principally from resource constraints. This led to suggestions that a degree of flexibility is required in how SMEs comply with ethical principles, including the principle of explainability. The following comments were typical of these attitudes:

- '… there needs to be some freedom for smaller companies, because they would die otherwise. They would not be able to innovate, because we don't have the resources to do what big companies can do'.
- 'It's just, how do you make … compliance workable for a small business?'
- '… I think the reality of startups is that startups are there to scale and make money and have a massive valuation. So they'll try to get to the quickest spot to POC [proof of concept] or MVP [minimum viable product], all right? So …[this] … means that ethics and all of those things may not be the priority…'
- '… being an SME, short of resources all the time … there was a million competing priorities. Being transparent wasn't the high on the list'.
- 'I think the small business don't have the support and the finance and the team and the experience to deal with it and to get around it. So large companies can get around it'.

While there was agreement both that SMEs should comply with ethical principles and that they face particular challenges in doing so, there was considerable uncertainty about the flexibilities that could assist SMEs in complying. As one interviewee put it, 'it's … hard to know what the compliance would look like'. One suggestion was that compliance thresholds might be flexible ('I am a big fan of thresholds for complicated compliance'). There was, however, general agreement among interviewees about the potential benefits of greater guidance on how SMEs can comply with ethical AI principles. As illustrated by the following comments, those supporting guidance emphasised the importance of a sufficient level of detail:

- 'I'd love to see … guidance because below the concepts of AI is all the detail, and that's where it all goes wrong'.
- '… to the extent that guidance is provided to small businesses, it should be super, super, super example heavy. And I can say that … the best thing you can show a client from any kind of regulator are just case study after case study after case study'.

Although guidance was supported, it was also generally regarded as less important than other measures ('I think … guidance [is] definitely helpful but it's more [important for other]

support'). The most common response to the challenges facing SMEs was to suggest mechanisms for supporting them to comply, especially by providing financial incentives or support, such as tax breaks. In this respect, the following comments were typical:

- 'Take, for example, tax. Everyone has the same tax rules, whether you are a small company or big companies. But then the government makes exceptions for small companies'.
- 'There's a lot of information but other than saying that you could go to jail for data handling, data privacy and data breach - the big sledgehammer - but there's nothing else to support. Where is the program? I don't even know if there's a program for a startup to say, "Oh, how can I be AI ethics compliant?"'
- 'There's so many people giving advice. There're so many people giving information sessions but what are you actually doing? Are you providing funding? Is the federal government giving a tax break…?'
- '…through the incubators and startup [subsidies] have things in there saying, "Here's the program, we will provide you an independent assessment to do that for free of charge or a tax break or we'll give you and help you to get down and look at your tools and your practices and help you to develop a data governance strategy and implement it for you free of charge or a tax break."'

J *Role of Government*

The interviewees held some common attitudes to the potential role of government in promoting or ensuring ethical AI. As explained above, there was a view that government could provide more assistance to SMEs in ensuring ethical AI. As one interviewee put it:

> 'If the ethics is important to the governments around AI, because it's the future, … and someone cares deeply enough that it needs to be done in the right way, then I think [there is] some policy benefit [in greater] … access to resources'.

The follow-up survey indicated general support for government involvement in developing ethical AI principles. There was, however, far less agreement about government involvement in more proactive regulation. From the interviews, one interviewee raised the possibility of government providing a form of 'certification that could be used in marketing', while another raised the prospect of voluntary regulation. However, none of the businesses that participated in the follow-up survey believed that government should develop AI-specific laws or regulation.

VI ANALYSIS OF FINDINGS

SMEs are increasingly using AI systems, promising socially beneficial innovation. However, this raises significant questions about how SMEs can comply with ethical AI principles, especially in the face of resource constraints. This study examined the approaches and attitudes of selected Australian SMEs to ethical AI, focusing on the ethical principle of explainability. In general, there was a high level of inconsistency in both attitudes to ethical AI and to practices for implementing ethical AI within businesses. The overall impression is that SMEs and start-ups are largely charting their own path with very limited assistance; understandably, they regard ethical considerations as secondary to other business priorities.

While the interviewees displayed a general grasp of ethical AI issues, such as problems relating to data quality, there was a lack of detailed knowledge of ethical principles. This extended to limited familiarity with specific statements of ethical principles, such as the principles promoted by the Australian government. This appears to be associated with the considerable variations in the approaches to implementing ethical AI within the businesses.

For example, there were no consistent approaches to in-house training or where to seek guidance on ethical AI, whether within the business or externally.

The inconsistencies were particularly apparent in the approaches to ethical AI safeguards, such as documentation and third party auditing. As a generalisation, documentation practices appeared to vary widely, while scepticism was expressed about the value of third party auditing. Given the flat structures of SMEs, internal accountability mechanisms tended to be informal, with ultimate responsibility for systems residing with founders, CEOs and/or data scientists. In general, there was no formal allocation of responsibility for ensuring ethical AI to a particular officer or organisational unit.

Even though most interviewees saw a need for some degree of explainability, there was considerable variation in the approaches taken to providing explanations: some businesses invested in providing additional information about the AI, while others avoided providing explanations. Generally, an explanation was considered important solely to the extent it was believed to deliver value to the business by, for example, building trust. On the other hand, where a business believed that customers are more interested in results than in how a system operates, explanations were not provided. Overall, the interviewees were cautious about burdening customers with too much information and believed their customers were more interested in accuracy than explainability. Therefore, where explanations were provided, some care was taken with the form of explanation, with simple or concise summaries being preferred and often provided face-to-face. This general approach also reflected concerns that providing too much information could reveal trade secrets or otherwise benefit competitors.

The interviewees generally considered that SMEs should comply with ethical AI principles, but were acutely aware of the particular challenges facing SMEs in implementing ethical AI. This led to calls for some leeway in how SMEs might be required to comply with ethical principles, but also for government assistance to aid SMEs with compliance. While the interviewees supported greater guidance on ethical AI principles, such as using case studies, they generally considered that financial assistance, such as tax relief or financial support for assessing AI systems, would likely be more critical. In general, the SMEs considered that there was a legitimate role for government in developing ethical AI principles and providing guidance. On the other hand, responses indicated a concern that other forms of government involvement, such as through regulation, could impose unnecessary costs on SMEs.

While some of the inconsistencies in the approaches and attitudes of the SMEs may reflect the different industries and AI systems the interviewees are involved with, it seems likely that there are more generally applicable explanations. As it remains early in the adoption of AI systems by SMEs, there is still a lot of trial and error in how AI is being implemented, with considerable variations between businesses. For many start-ups and SMEs, the priorities of the business will largely reflect those of the founder or founders. More importantly, start-ups and SMEs are overwhelmingly concerned with building and maintaining their business, often in highly competitive markets, with other concerns being subsidiary. While larger enterprises have the resources to address broader concerns, SMEs must direct their resources to ensuring financial viability. These considerations strongly suggest that, unaided, it is likely that the attitudes and practices of SMEs will continue to exhibit an undesirably high degree of inconsistency.

VII Conclusion

Acknowledging the limited nature of this study, we consider it reasonable to conclude that there is an unarguable case for governments to pay greater attention to promoting ethical AI

among SMEs, including in promoting greater understanding of the principle of explainability. In particular, we believe there is a case for initiatives such as the following:

- Greater efforts are required to educate SMEs about ethical AI, including proactively disseminating information about ethical AI principles through SME networks, as well as developing detailed case studies illustrating how SMEs can apply the principles in practice;
- In particular, detailed guidance is needed to assist SMEs in understanding what is required for explaining AI systems, and how this can be done without exposing trade secrets or alienating customers;
- While there is a general need for more resources to be allocated for training businesses in ethical AI, there is a specific (and we think pressing) need for training to be targeted to the particular challenges faced by SMEs. Given the resource constraints faced by SMEs, there may be a case for financial aid to support SME participation in training initiatives;
- More guidance is needed in relation to the standard business safeguards reasonably expected from SMEs using AI systems, especially those systems that may significantly impact people. These safeguards should include standardised expectations relating to documentation, as well as standardised processes for assessing AI systems;
- As a step towards greater consistency in the practices adopted by SMEs, some consideration could be given to adopting a government-backed certification scheme for ethical AI. Such a scheme, while not a magic bullet, could assist with transparency and standardisation of practices, including practices relating to explaining AI systems; and
- Greater standardisation also seems to be required in relation to the qualifications and practices of third party auditors of ethical AI. This may require some form of certification.

In drawing these conclusions, we stress that the sorts of measures contemplated will require significant investment by government including funding to develop and deliver resources and support to SMEs. The findings of this research should inform the program of activities of the National AI Centre and the associated AI and Digital Capability Centres to be established by the Australian Government under the Australian AI Action Plan.

THE HACKER STRIKES BACK: EXAMINING THE LAWFULNESS OF "OFFENSIVE CYBER" UNDER THE LAWS OF AUSTRALIA

BRENDAN WALKER-MUNRO,* RUBY IOANNOU† AND DAVID MOUNT‡

ABSTRACT

Over the past ten years, criminal offending utilising or involving computers and information systems has risen to become one of the most prevalent global security threats. With a low cost to entry and high potential benefits, cybercrime is a significant challenge for traditional forms of law enforcement investigation. In response, many Western democracies have passed laws permitting officers of policing and intelligence agencies to "hack back" – that is, to use computers to attack, infiltrate, damage or disrupt the information systems of criminal offenders. Yet the contours and boundaries of those laws are underexamined in the literature. How do these agencies pursue criminals operating extraterritorially? On what legal basis can police, intelligence services or even the military attack the computers of a criminal group? This paper seeks to chart the legal parameters of the use of cyber capabilities by Australian national security agencies on a domestic basis.

CONTENTS

In August 2020, *The Australian* newspaper published details of a discussion paper authored by the Department of Home Affairs.[1] According to the report, the Department was proposing legislative amendments to the *Security of Critical Infrastructure Act 2018* (Cth) ("the SOCI Act") which would have permitted the Australian Signals Directorate (ASD) – Australia's top secret signals intelligence agency – to intervene in times of emergency. In those terms, the

* Senior Research Fellow, T.C. Beirne School of Law, University of Queensland.
† Research Assistant, University of Queensland.
‡ Lecturer in Cyber Criminology, School of Social Science, University of Queensland.

1 Simon Benson, Geoff Chambers, '"Hack back" powers to repel cyber attack', *The Australian* (online, 12 August 2020).

emergency was deemed any 'immediate and serious cyber threat' to Australia's 'economy, security or sovereignty, including threat to life'.[2] Once such an emergency situation arose, ASD would then be empowered to 'take direct action to actively deny, disrupt and respond to malicious activity with corresponding powers and immunities'.[3] More recently, ASD was named in a 'joint standing operation' with the Australian Federal Police to target cybercriminals and foreign hackers.[4] Though the methodology of ASD remained highly classified, it was apparent from the report that Australia was contemplating the ability of ASD to "hack back"; that is, to use ASD's own cyber capabilities against the would-be criminals.

But what does hacking back mean in legal terms? The exact circumstances in which hacking back might be contemplated, which types of emergency situations qualified, and exactly what those powers and immunities authorise is not well understood in the literature. Even more concerningly, since many digital connections are transnational, there exists significant capacity for hacking back to cross jurisdictional boundaries with greater ease. Australia possesses a unique geostrategic position, with an abundance of natural wealth, prosperous citizens, and high performing economies – all of which are rich targets for cybercrime. Notwithstanding, this topic is under-described in the current legal discourse.

The hack back phenomenon is not solely limited to Australia. In 2021, the International Institute for Strategic Studies (IISS) published its report into the cyber capabilities of 15 countries. The IISS report claimed that the United States (US) 'capability for offensive cyber operations is probably more developed than that of any other country, although its full potential remains largely undemonstrated'.[5] Following closely behind the US in cyber offensive capability were other members of the Five Eyes intelligence alliance – including the United Kingdom (UK), Australia and Canada – but also ally States such as Israel, France and Japan. States with traditionally counter-Anglo interests were also included such as China, Iran and Russia.[6]

The concerns here are far from academic. Not only do individual officers of the Executive need to be properly empowered for undertaking what are illegal acts if conducted by a member of the public, but Australia must maintain proper respect for the rules-based global order. The legislative underpinnings of hack-backs are also important from the perspective of paying respect to international sovereignty. In other words, how might it be legal for one country (such as Australia) to attack the software or hardware owned by a criminal group located in a second country, but passing through the digital connectivity infrastructure of a third country? Opportunities for major diplomatic or political incidents are manifest.

This paper will therefore seek to make some inroads into understanding and mapping the legality of "hacking back" under Australian law – in particular, its position as a tool of state intervention in the same vein as covert operations or spying. The focus of that analysis will be solely upon the domestic legislation of Australia that creates or permits opportunities for hacking back. Beyond the necessary discussion of aspects of international law that might

---

2 Ibid.

3 Ibid.

4 Ry Crozier, 'Australia sets up 100-strong permanent "operation" to target hackers', *IT News* (online, 12 November 2022) <https://www.itnews.com.au/news/australia-sets-up-100-strong-permanent-operation-to-target-hackers-587691>.

5 International Institute for Strategic Studies, *Cyber Capabilities and National Power: A Net Assessment* (Final report, 28 June 2021) <https://www.iiss.org/blogs/research-paper/2021/06/cyber-capabilities-national-power>, 15.

6 Ibid, 10-12.

create domestic obligations, this paper will not deal with any instrument or custom of international law.

Part 1 will introduce the idea of "hacking back" as a conceptual vehicle, exposing some of the issues which pose challenges for various legislation in Australia. Section 2 will examine the legal provisions which would enable deployment of cyber offensive capabilities inside Australia's territorial boundaries. A brief history of the Acts permitting calling-out of cyber offensive assets will also be explored. Section 3 will then discuss the circumstances under which our national security agencies should be able to utilise their capabilities domestically. Concepts of emergency and necessity will be examined to formulate principles under which it could be considered permissible for Australia to hack back, as well as highlighting a significant deficit in the existing research literature around this topic. Finally, the paper will close with some summarising observations in Section 4.

I WHAT IS HACKING BACK?

In order to understand the legalities of hacking back in context, it is important to establish exactly what the paradigm refers to. Unsurprisingly, though States might have been quick to admit the existence of cyber units with these types of capabilities, they have been far more circumspect about exactly what those capabilities could allow or permit.[7] Early examinations of cyber capabilities generally grouped together by intention, distinguishing those that were considered "offensive" (being in an attacking or proactive capacity) from those that were "defensive" (being in a responsive or reactive capacity).[8] Pragmatically though, the definition between offensive and defensive has been obliterated. The joint US and Israeli cyber operation codenamed "Olympic Games" – which allegedly included the deployment of the Stuxnet worm against Iranian nuclear centrifuges – and the Russian WannaCry and NotPetya ransomware attacks were really both offensive and defensive in nature (at least according to their instigators).[9]

This difficulty is compounded by a lack of common language. Both US and UK military doctrine treat offensive capabilities in a similar way as defensive ones, by considering 'cyber operations' as the projection of power into the online, digital or virtual environments through

---

[7] Allie Coyne, 'Australia has created a cyber warfare unit', *ITNews* (online, 30 June 2017) <https://www.itnews.com.au/news/australia-has-created-a-cyber-warfare-unit-467115>; Jeremy Fleming, 'Director's speech at Cyber UK 2018' (Speech, CyberUK18 Conference, 12 April 2018), <https://www.gchq.gov.uk/pdfs/speech/director-cyber-uk-speech-2018.pdf>, 7; Michael S. Rogers, Evidence to Senate Committee on Armed Services, United States, 27 February 2018, <https://www.armedservices.senate.gov/imo/media/doc/Rogers_02-27-18.pdf>, 1-3; National Defence (Canada), 'Strong, Secure, Engaged: Canada's Defence Policy' (online), <https://www.canada.ca/content/dam/dndmdn/documents/reports/2018/strong-secure-engaged/canada-defence-policy-report.pdf>, 41; Ed Caesar, 'The Incredible Rise of North Korea's Hacking Army', The New Yorker (online, 19 April 2021) <https://www.newyorker.com/magazine/2021/04/26/the-incredible-rise-of-north-koreas-hacking-army>.

[8] Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Report, RAND Corporation, 2009); Kenneth Lieberthal, Peter W. Singer, *Cybersecurity and U.S.-China Relations* (Washington DC, Brookings Institution, 2012); Adam P. Liff, 'Cyberwar: A New 'Absolute Weapon'? The Proliferation of Cyberwarfare Capabilities and Interstate War' (2012) 35(3) *Journal of Strategic Studies* 401; Lucas Kello, 'The Meaning of the Cyber Revolution: Perils to Theory and Statecraft' (2013) 38(2) *International Security* 7; Timothy J. Junio, 'How Probable Is Cyber War? Bringing IR Theory Back in to the Cyber Conflict Debate' (2013) 36(1) *Journal of Strategic Studies* 125.

[9] James P. Farwell, Rafal Rohozinski, 'The New Reality of Cyber War' (2012) 54(4) *Survival* 107, 109; Mary Ellen O'Connell, 'Attribution and Other Conditions of Lawful Countermeasures to Cyber Misconduct' (2020) 10(1) *Notre Dame Journal of International Comparative Law* 1.

the use of computer or information systems.[10] Canada legalised aspects of offensive cyber in 2018 under its counterterrorism laws but used the more euphemistic "active cyber" – the 'prevent[ion of] threats before they reach Canadian (and possibly allied) targets' – to describe their activities.[11] New Zealand remains staunchly in the defensive camp, despite possessing what some have called offensive capabilities.[12] Australia's cyber security strategy does not define offensive or defensive actions,[13] nor did Australia's international cyber engagement strategy.[14] However, both the-then Prime Minister of Australia[15] and Australian Signals Directorate[16] both made public statements in which they announced the existence of Australia's offensive capabilities in this domain.

The predominant problem with both differentiating by intention or language appears to be trying to view cyber operations through the same lens as traditional kinetic warfare where, instead of two sides exchanging bullets or artillery, warfare in the cyber domain has tried to incorporate the idea of a back-and-forth exchange of viruses, worms and malware.[17] Much of the literature acknowledges that in the online world the traditional doctrines of warfare break down, non-State and State actors are indistinguishable and the lines between crime, terrorism and open war are impossible to define.[18]

"Hacking back" exemplifies these difficulties in both intention and language. The term generally covers those 'proactive steps a victim of a cyberattack takes against their assailant in order to retaliate against their attacker',[19] thus incorporating dimensions that are both defensive (retaliating against an attacker) and offensive (employing methods to manipulate, damage or destroy the attacker's systems). By referring to the steps taken by "victims", the literature also acknowledges that "hacking back" applies to non-State actors where 'government departments and law enforcement agencies are unable or unwilling to

---

[10] Joint Chiefs of Staff, *Cyberspace Operations* (JP 3-12, Department of Defense, 8 June 2018) <https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_12.pdf>; Chiefs of Staff, *Cyber Primer* (Third edition, October 2022) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1115061/Cyber_Primer_Edition_3.pdf>.

[11] Stephanie Carvin, 'Zero D'Eh: Canada Takes a Bold Step Towards Offensive Cyber Operations', *Lawfare* (blog, 27 April 2018) <https://www.lawfareblog.com/zero-deh-canada-takes-bold-step-towards-offensive-cyber-operations>.

[12] Tom Pullar-Strecker, '"Open secret" NZ has offensive cyber capability, security firm says', *Stuff* (online, 1 December 2021) <https://www.stuff.co.nz/business/127156155/open-secret-nz-has-offensive-cyber-capability-security-firm-says>.

[13] Commonwealth of Australia, *Australia's Cyber Security Strategy 2020* (P-20-02344, 2020).

[14] Department of Foreign Affairs and Trade (Cth), *Australia's International Cyber Engagement Strategy* (October 2017).

[15] Malcolm Turnbull, 'Launch of Australia's Cyber Security Strategy Sydney' (Press Statement, 21 April 2016).

[16] Jackson Graham, '"Uncomfortable" debate about offensive cyber attacks increasingly public as security environment shifts', *The Mandarin* (online, 22 November 2021) <https://www.themandarin.com.au/175558-uncomfortable-debate-about-offensive-cyber-attacks-increasingly-public-as-security-environment-shifts/>.

[17] James P. Farwell, Rafal Rohozinski, 'The New Reality of Cyber War' (2012) 54(4) *Survival* 107, 113.

[18] Thomas Rid, 'Cyber War Will Not Take Place' (2012) 35(1) *Journal of Strategic Studies* 5; John Stone, 'Cyber War Will Take Place!' (2013) 36(1) *Journal of Strategic Studies* 101; Richard A. Clarke, 'The risk of cyber war and cyber terrorism' (2016) 70(1) *Journal of International Affairs* 179; Tarah Wheeler, 'In cyberwar, there are no rules' (2018) 230 *Foreign Policy* 34; Christopher J. Finlay, 'Just war, cyber war, and the concept of violence' (2018) 31(3) *Philosophy & Technology* 357.

[19] Valeska Bloch, Sophie Peach, Lachlan Peake, 'The Hack Back: The Legality of Retaliatory Hacking' (2018) 37.4 *Communication Law Bulletin* 8.

effectively respond to cybercrime'.[20] Hacking back is also equally plagued by different terms, with 'active defence', 'retaliatory hacking' and 'counter hacking' being used interchangeably with "hacking back" in the literature.[21]

This paper will take a broad view of the term "hacking back", with a view that it describes the actions taken by both State and non-State actors in response to a cybersecurity threat or incident to 'manipulate, deny, disrupt, degrade, or destroy targeted computers, information systems or networks'[22] belonging to or supporting the actions of an attacker. There are several reasons for taking this approach. The first is that this definition incorporates those aspects of previous definitions on which scholars agree.[23] Secondly, this definition recognises the growing role of non-State actors in responding to cybersecurity threats[24] (and permits examining their call for greater legitimacy and legal protection, which will be dealt with later). Thirdly, it focuses the inquiry solely on responsive acts and thus excludes proactive or "first strike" capabilities.[25] Fourthly, this definition places the concept of "intercepting communications" (using interception warrants issued under the *Telecommunications (Interception and Access) Act 1979* (Cth) or its international equivalents) outside of the scope of the paper, as the defined act of hacking back involves solely retaliatory or reactive responses.

These definitions now having been established, we wish to examine how Australia currently legitimises the use of hacking back by its organs of national security, intelligence agencies, police, and the Australian Defence Force (ADF). These agencies are not only the most likely first responders to the types of national security threats in the cyber domain – such as adversarial States, cybercriminal groups, politically motivated extremists and "lone wolf" actors[26] – but also those traditionally funded and empowered to undertake such actions.

II LEGISLATIVE AUTHORITY FOR HACKING BACK IN AUSTRALIA

The national security apparatus of Australia is, at least by reference to its enabling legislation, notoriously fragmented: one recent legislative reviewer called the oversight of law

---

[20] Gavin Smith, Valeska Bloch, *The hack back: The legality of retaliatory hacking* (Blog, Allens Linklaters, 17 October 2018) <https://www.allens.com.au/insights-news/insights/2018/10/pulse-the-hack-back-the-legality-of-retaliatory-hacking/>.

[21] Jay P. Kesan, Ruperto P. Majuca, 'Hacking Back: Optimal Use of Self-Defense in Cyberspace' (2010) 84(3) *Chicago-Kent Law Review* 1; Benjamin Baker, 'Considering the Potential Deterrence Value of Legislation Allowing Hacking Back' (2018) SSRN: https://ssrn.com/abstract=3319530; Gabriel Martinez, *Hacking Back: Self-Defense, Self-Preservation or Vigilantinism* (PhD thesis, Utica College, 2020); Hannah Gallagher, 'Recognising a Right to Hack Back-Tom and Jerry in Cyberspace?' (2022) 25(1) *Trinity College of Law Review* 56.

[22] Tom Uren, Bart Hogeveen, Fergus Hanson, *Defining Offensive Cyber Capabilities* (Final report, ASPI, 4 July 2018) <https://www.aspi.org.au/report/defining-offensive-cyber-capabilities>. We also use the word "computer" broadly, as it can include mobile devices and Internet of Things (IoT) enabled systems.

[23] Ibid; Bloch, Peach & Peake, n 16; Gallagher, n 18; Josh Gold, *The Five Eyes and Offensive Cyber Capabilities: Building a 'Cyber Deterrence Initiative* (Report, NATO Cooperative Cyber Defence Centre of Excellence, 2020); Brendan Walker-Munro, 'White Hat, Black Hat, Slouch Hat: Could Australia's Military Cyber Capability be Deployed against Threats Inside Australia?' (2023) *Federal Law Review*, in proof.

[24] Tom Kulik, 'Why the Active Cyber Defense Certainty Act Is a Bad Idea', *Above the Law* (blog, 29 January 2018) <https://abovethelaw.com/2018/01/why-the-active-cyber-defense-certainty-act-is-a-bad-idea/>.

[25] Leonard Spector, 'Cyber Offense and a Changing Strategic Paradigm' (2022) 45(1) *Washington Quarterly* 38.

[26] Helen Wong, *Cyber Security: Law and Guidance* (Bloomsburg Professional, New York, 2018) [12.01].

enforcement and intelligence agencies 'a dog's breakfast'.[27] Intelligence agencies – being those constituting the National Intelligence Community (NIC)[28] – are predominantly covered by the provisions of the *Intelligence Services Act 2001* (Cth) ("IS Act"). The only exception for this paper for NIC agencies is ASIO, which draws its powers from the *Australian Security Intelligence Organisation Act 1979* (Cth) ("the ASIO Act"). The AFP are also excepted from the IS Act, which are also given powers as constables under the *Crimes Act 1914* (Cth) ("Crimes Act"), but are still considered part of the NIC.

Australia's military forces – the Army, Navy and Air Force operating under the unified banner of the ADF – can draw their operating legitimacy from two locations. The first is in statute, more specifically the powers accruing to Defence members participating in a "call out order" issued by the Governor-General under Part IIIAAA of the *Defence Act 1903* (Cth) ("the Defence Act"). Part IIIAAA actions are reserved for responses to "domestic violence" where the 'use, or potential use, of force (including intrusive or coercive acts) is required by Defence members' in situations of threats to Commonwealth interests or at the behest of the States or Territories.[29] The other source of operational legitimacy for deployment of the ADF comes from the Crown prerogative which permits the Governor-General to direct the ADF to 'be anywhere' and thereby undertake such operational activities as he or she sees fit.[30]

Finally, the policing forces at the State and Territory level are usually dependent entirely upon the enabling legislation in each of their jurisdictions. However, amendments to the *Surveillance Devices Act 2004* (Cth) (SDA) in 2021[31] introduced a series of warrants capable of supporting hacking back for State and Territory police. Generally speaking, such warrants permit police officers to hack into a computer – whether located in Australia or elsewhere – and modify, alter, delete or destroy information on any such computer or network.[32] Whether or not those provisions enable the populist concept of "hacking back" to occur is a rather more nuanced question.

A *Legal Authority for NIC Hacking Back*

The NIC operates on what can largely be considered a geographical divide – agencies such as the Australian Secret Intelligence Service (ASIS), Australian Signals Directorate (ASD), Australian Geospatial-Intelligence Organisation (AGO) and the Defence Intelligence Organisation (DIO) have a foreign remit, to collect and process intelligence and to conduct operations involving threats not located within Australia. Given these agencies are largely bound by the limitations in the IS Act,[33] their utilization is predominantly connected to preventing or disrupting foreign cybercrime[34] or assisting the ADF with military operations

---

[27] Dennis Richardson, *Comprehensive Review of the Legal Framework of the National Intelligence Community* ("the Richardson Review") (Volume 1, December 2019) [3.71].

[28] Australian Security Intelligence Organisation (ASIO), Australian Secret Intelligence Service (ASIS), Australian Signals Directorate (ASD), Australian Geospatial-Intelligence Organisation (AGO), Defence Intelligence Organisation (DIO), Office of National Intelligence (ONI), Australian Federal Police (AFP), Australian Criminal Intelligence Commission (ACIC), Australian Transaction Reports and Analysis Centre (AUSTRAC), and the Department of Home Affairs.

[29] D. L. Johnston, *Defence Assistance to the Civil Community Policy* (Department of Defence, 31 August 2021) 5.

[30] Cameron Moore, *Crown and Sword: Executive power and the use of force by the Australian Defence Force* (ANU Press, 2017) 169.

[31] As a result of the passing of the *Surveillance Legislation Amendment (Identify and Disrupt) Act 2021* (Cth).

[32] SDA, Pt 2, Divs 4-6.

[33] IS Act, s 7.

[34] Ibid, s 7(1)(c).

overseas.[35] Two recent examples include the targeting of Islamic State assets during "Glowing Symphony"[36] and assisting Australian Federal Police investigate the hackers behind the theft of health insurance information from Medibank Private.[37]

Conversely, the Office of National Intelligence (ONI) is constituted under a separate Commonwealth Act[38] and has only leadership and coordination roles within the NIC ecosystem. ONI is limited by statute to the leadership and 'evaluation of matters' involving the NIC, but not the NIC agencies themselves, and to advise the Prime Minister on the results of such evaluations. These evaluations are curtailed to the extent that they may or do 'inappropriately impact on, or encroach on the functions, powers and responsibilities' of members of the NIC.[39]

A strong prospect for the authorisation of NIC agencies to conduct "hacking back" under a legitimate framework is contained in the ASIO Act.[40] Effectively, the ASIO Act empowers the Director-General of ASIO to seek a warrant from the Attorney-General if:

> …he or she is satisfied that there are reasonable grounds for believing that access by the Organisation to data held in a computer (the target computer) will substantially assist the collection of intelligence in accordance with this Act in respect of a matter (the security matter) that is important in relation to security.[41]

The necessary test for the Attorney-General to issue such a warrant is three-fold. Firstly, the warrant must specify a target computer by reference to some defining characteristic that separates the target computer from unrelated, nearby or connected devices.[42] Secondly, there must exist a security matter upon which the Organisation has collected, or is proposing to collect, intelligence. Thirdly, the Director-General must hold a reasonable belief based on reasonable grounds that the Organisation gaining access to the target computer will be 'substantially assist' the collection of intelligence in respect of that security matter.

These tests place important boundaries around the exercise of ASIO's hacking back power. The concept that a warrant cannot authorise a "fishing expedition" – usually described as an exercise of statutory power in the absence of reasonable grounds, conducted in the hope of finding some evidence which justifies the intrusion – has been long established in the common law.[43] In this case, a computer access warrant must be directed towards a target computer, defined with some precision in order to comfort the Attorney-General that the Organisation's access to that computer is both possible and reasonable to achieve the objectives of the warrant.

Further, the use of a computer access warrant under the ASIO Act requires a connection to a security matter both generally but also in relation to the collection of intelligence from the target computer. This must be a matter related to security as it is defined in the Act,[44]

---

[35] Ibid, s 7(1)(d).

[36] Australian Signals Directorate, *Annual Report 2019-20* (Final report, Canberra, 12 October 2020) 29.

[37] Mamoun Alazab, 'A new cyber taskforce will supposedly 'hack the hackers' behind the Medibank breach. It could put a target on Australia's back', *The Conversation* (online, 16 November 2022) <https://theconversation.com/a-new-cyber-taskforce-will-supposedly-hack-the-hackers-behind-the-medibank-breach-it-could-put-a-target-on-australias-back-194532>.

[38] *Office of National Intelligence Act 2018* (Cth).

[39] Ibid, ss 8-10.

[40] ASIO Act, s 25A.

[41] Ibid, s 25A(2).

[42] Ibid, s 25A(3A)(c)-(e).

[43] Warwick McKean, 'Searches and Sandwiches' (1978) 37(2) *The Cambridge Law Journal* 200-202 <https://www.jstor.org/stable/4506084>.

[44] ASIO Act, s 4.

requiring that ASIO by necessity limits the circumstances in which a warrant of this type might be sought or granted (although ASIO does have a derivative use provision enabling them to copy any data relating to a security matter which was *not* mentioned in the warrant[45]).

Finally, any "hacking back" conducted under a computer access warrant may only alter, damage or destroy data on that computer for the purpose of obtaining access to the data the subject of the warrant and/or concealing the activities of ASIO.[46] A computer access warrant does not authorise ASIO or its officers or agents to alter, damage or destroy data for any other purpose (i.e., disrupting or preventing crime), though the Act curiously only protects *lawful* use of a computer or telecommunications network by persons involved with the target computer and remains silent about *unlawful* use.[47]

For completeness, there are two other forms of warrant which might also be relied upon by ASIO to conduct "hacking back" activities. An 'identified person warrant' may be issued by the Attorney-General on the application of the Director-General of ASIO, and where the Attorney-General is satisfied of two conditions: firstly, that a named person is involved in activities prejudicial to security; and secondly that the issuing of the warrant will, or is likely to, 'substantially assist the collection of intelligence relevant to security'.[48] Once issued, the warrant may be directed to that named person's electronic devices if, and only if, ASIO can satisfy the Director-General or Attorney-General at a later time that the data sought 'will substantially assist the collection of intelligence relevant to the prejudicial activities of the identified person'.[49] ASIO may then 'if necessary to achieve that purpose, add, copy, delete or alter other data in the target computer',[50] but are bound by the same prohibitions on loss or damage as computer access warrants.[51]

The second form of warrant under which a computer might be hacked back is under a foreign intelligence warrant.[52] Rather than a separate class of warrants, foreign intelligence warrants are those issued under the same provisions above (sections 25A and 27C) but are directed not to "security" matters but "foreign intelligence" matters which arise within Australia.[53] These warrants, again requested by the Director-General and granted by the Attorney-General, carry an additional threshold test. Prior to issuing a foreign intelligence warrant, the Attorney-General must receive advice from the Defence Minister or Foreign Affairs Minister to the effect 'that the collection of foreign intelligence relating to that matter is in the interests of Australia's national security, Australia's foreign relations or Australia's national economic well-being'.[54] Foreign intelligence warrants also do not authorise the damaging of computers which would interfere with lawful use.[55]

Having thus exhausted all the hacking provisions under the ASIO Act, it remains to consider whether any of the other NIC agencies might be able to conduct cyber offensive

---

45 If the data 'appears to be relevant to the collection of intelligence by the Organisation in accordance with this Act'; ibid s 25A(3A)(b).

46 Ibid, s 25A(4)(a) and (c).

47 Ibid, s 25A(5).

48 Ibid, s 27C(2).

49 Ibid, ss 27C(3)(c)(ii) and 27E(4). Authorisations must be requested and granted in writing: ibid, s 27J(1) and (3).

50 Ibid, s 27E(2)(c).

51 Ibid, s 27E(5)

52 Ibid, s 27A.

53 And would therefore be the remit of other NIC agencies: IS Act, s 6, 6B and 7.

54 ASIO Act, s 27A(1)(b).

55 Ibid, s 27A(3D).

activities within the domestic territory of Australia. IS Act agencies – including ASIS, AGO and ASD – function under the framework of Ministerial authorisation and directions.[56] A direction imposes obligations on an NIC agency to refrain from conducting certain activities in respect of Australian persons, whether those activities are the production of intelligence,[57] assistance to military operations by the ADF[58] or the prevention or disruption of cybercrime[59] unless the Minister authorises those activities.

To provide such an authorisation, the Minister must be satisfied of numerous matters;[60] however, this is not the greatest difficulty which faces IS Act agencies in attempting to "hack back" within Australia's boundaries. The limitation placed upon these agencies by the IS Act clearly excludes much of their capabilities from operating inside Australia on their own initiative. For example, ASIS may only collect intelligence about 'the capabilities, intentions or activities of people or organisations **outside Australia**', or 'undertake such other activities as the responsible Minister directs relating to the capabilities, intentions or activities of people or organisations **outside Australia**'[61] (emphasis added in all cases).

AGO and ASD have similar limitations placed upon them. AGO for example is limited to producing various forms of geospatial intelligence on persons and organisations 'outside Australia',[62] whilst ASD is limited to preventing or disrupting cybercrime 'undertaken by people or organisations outside Australia'.[63] Neither agency's core remit relates to activities undertaken within the territorial boundaries of Australia and in fact excludes those activities by operation of statute. There are however, several interesting loopholes which might permit IS Act agencies participating in the use of "hacking back" without offending the limitations imposed on them.

The first is where these agencies have functions bestowed upon them by the IS Act which do not carry the "outside Australia" caveat. In the case of ASIS this is counter-intelligence activities[64] (i.e., 'the identification and neutralization of the threat posed by foreign intelligence services, and the manipulation of those services for the manipulator's benefit'[65]), and for ASD this is their technological protection function (i.e., to 'protect specialised technologies acquired in connection with the performance' of any of ASD's other functions under the IS Act[66]).

Broadly speaking, both ASIS and ASD might be permitted to conduct hacking into an offender's computer where that offender was geographically located inside Australia if, and only if:

- In the case of ASIS, they reasonably believed that the offender was working for, or acting on behalf of, a 'foreign principal'[67] and such hacking back activities were reasonably necessary to fulfil ASIS' counterintelligence functions; or

---

[56] IS Act, s 8 and 9.

[57] Ibid, ss 8(1)(a)(i), (iaa) and (ii).

[58] Ibid, ss 8(1)(a)(ia) and (ib).

[59] Ibid, s 8(1)(a)(iii).

[60] Broadly, see ibid, ss 9(1), (1A) and (1AAA).

[61] Ibid, s 6(1)(a) and (e).

[62] Ibid, s 6B(1)(a).

[63] Ibid, s 7(1)(c).

[64] Ibid, s 6(1)(c).

[65] Carl Anthony Wege, 'Hizballah's Counterintelligence Apparatus' (2012) 25(4) *International Journal of Intelligence and Counterintelligence* 771.

[66] IS Act, s 7(1)(da).

[67] *Criminal Code*, s 90.2.

- In the case of ASD, the offender was attempting to steal, subvert, damage or destroy some particular specialised technology which ASD had acquired in the performance of its other IS Act functions, i.e., specialised cryptographic, communication or computer technologies.[68]

The vagueness of these terms should give IS agencies cause for caution; however, there are more pressing issues. In both cases, the conduct of either or both of ASIS or ASD in hacking back a domestic Australian offender would require both a Ministerial direction and authorisation, and a Ministerial direction *cannot* be issued in respect of activities which would require a warrant under the ASIO Act (such as a computer access warrant).[69] Even in the event that a direction could be issued, the Minister would need to be satisfied, as a precondition to the giving of an authorisation that 'there are satisfactory arrangements in place to ensure that the nature and consequences of acts done in reliance on the authorisation will be reasonable, having regard to the purposes for which they are carried out'.[70] The reasonableness of the acts done will be difficult to justify in circumstances whereby the same activities justifying ASIS or ASD's involvement would also more reasonably justify ASIO's jurisdiction as a matter affecting Australia's security.[71]

The second pathway under which hacking back might be contemplated is under the assistance functions accruing to IS Act agencies,[72] not only as discrete functions of their agencies but also under sections 13 and 13A of the IS Act. These activities do not automatically mandate Ministerial directions or authorisations, though they may be covered by other directions or authorisations depending on the activities being contemplated. Whether providing assistance to the AFP, a State or Territory Police Force, a specialist regulator or other Department, there is the possibility that the Minister could issue an authorisation for activities involving the deployment of such capabilities inside Australia for the purpose of its assistance functions.

The necessary test for the Minister is – hinging on use of the word "may" in section 9(2) of the IS Act – dependent solely on whether he or she was reasonably satisfied that an authorisation would permit activities necessary for the proper performance of one of the relevant IS agency's functions, there are satisfactory control arrangements in place to not exceed the authorisation, and those arrangements can ensure that the 'nature and consequences of acts done in reliance on the authorisation will be reasonable'.[73]

This construction too should be treated with some scepticism. The cooperation provisions in the IS Act explicitly limit the boundaries of cooperation with a government agency to the performance of the cooperating agency's functions. For example, the AFP has policing and investigative functions to deal with cybercrime.[74] ASD has a function to 'provide material, advice and other assistance… on matters relating to the security and integrity of information that is processed, stored or communicated by electronic or similar means',[75] and

---

[68] IS Act, s 7(1)(e)(i).

[69] Ibid, s 8(1B).

[70] Ibid, s 9(1)(c).

[71] In fact, s 4 of the ASIO Act includes 'acts of foreign interference', 'espionage' and 'attacks on Australia's defence system' as matters of national security.

[72] Ibid, ss 6(1)(ba) (ASIS); 6B(1)(b), (c), (e) and 6B(2) (AGO); s 7(1)(ca), (d), (e) and 7(2) (ASD).

[73] Ibid, s 9(1)(a)-(c). Section 9(1)(d) does not apply, as the authorisation does not relate to activities of ASIO: ibid, ss 8(1)(a)(ia) and (ib).

[74] *Australian Federal Police Act 1979* (Cth), s 8(1)(b)(i) and 8(1)(bf)(i).

[75] Ibid, s 7(1)(ca).

the IS Act permits ASD to assist the AFP with the performance of the investigation function.[76] Yet the exact practical assistance which ASD can provide is circumscribed by application only to persons and organisations located outside Australia.[77] Though this limitation is specifically carved out when IS Act agencies cooperate either with ASIO or each other, the nature of such support is specifically limited to what is asked by the requesting agency, i.e., IS Act agencies cannot venture outside the parameters of their lawful functions where this has not been specifically requested.[78] ASIO activities would need to be authorised by a warrant under the ASIO Act, and IS Act agencies – even those cooperating with one another – would be limited by the application to persons and organisations outside Australia.

In summary, the above analysis shows that the deployment of Australian "hack backs" by the IS Act agencies (ASIS, AGO and ASD) would not be lawful if it was deployed inside Australia, both by reference to the powers and functions of those agencies. Instead, ASIO appears to retain the primary jurisdiction for the conduct of hack backs in Australia under the authority of computer access warrants; however, this capability is limited to only the collection of intelligence about the activities of the offender and would not extend to disrupting or destroying the offender's data or systems in response to an incident. The question of whether ASIO or the IS Act agencies could assist either of the ADF or Police forces to do so will be answered in the subsequent sections.

B *Legal Authority for ADF Hacking Back*

The ADF first established an Information Warfare Division in 2017 as part of its Joint Capabilities Group, with a specific remit to conduct cyber operations alongside traditional forms of military operations and armed conflict.[79] A year after its establishment, in 2018 the Australian Strategic Policy Institute then examined the publicly available information regarding how ADF capability worked alongside ASD, concluding that '[a]ny offensive cyber operation in support of the ADF is planned and executed under the direction of the Chief of Joint Operations and, as with any other military capability, is governed by ADF rules of engagement'.[80] Since that time, ADF personnel and materiel have focused on the domain of cyber in extensive training and capability development works.[81] In other words, deployment of offensive cyber capabilities – "hacking back" – by the ADF would be conducted under the imprimatur and legality of a military operation and not an intelligence operation.

So, what are the legal parameters of a military deployment in which hack backs might be used? And could Australia ostensibly be a location in which such a deployment could occur?

---

[76] Ibid, s 13(1)(a).

[77] Ibid, s 11(1).

[78] Ibid, s 13A(1)(b) and (2).

[79] Greg Austin, '"Cyber revolution" in Australian Defence Force demands rethink of staff, training and policy', *The Conversation* (online, 4 July 2017) <https://theconversation.com/cyber-revolution-in-australian-defence-force-demands-rethink-of-staff-training-and-policy-80317>.

[80] ASPI, n 19, 7.

[81] Marcus Thompson, 'The ADF and Cyber Warfare' (2016) 200 *Australian Defence Force Journal* 43; Jonathon C. Ladewig, 'Australia's Readiness for a Complex Cyber Catastrophe' (2018) 14(2) *Australian Army Journal* 57; Ben Johanson, 'Asymmetric Advantage in the Information Age: An Australian Concept for Cyber-Enabled "Special Information Warfare"' (2018) 14(2) *Australian Army Journal* 79; Linda Reynolds, 'Stronger cyber defences for deployed ADF networks' (Media release, 12 August 2020) <https://www.minister.defence.gov.au/media-releases/2020-08-12/stronger-cyber-defences-deployed-adf-networks>; Australian Cybersecurity Magazine, 'Australian Defence Force pilots cyber training program' (online, 8 September 2020) <https://australiancybersecuritymagazine.com.au/australian-defence-force-pilots-cyber-training-program/>.

The answer is not straightforward. Firstly, Australia has a very complex system relating to domestic deployment of its military force under call out orders contained in Part IIIAAA of the Defence Act. These deployments are either to protect Commonwealth interests from real, perceived, or anticipated threat[s82] or in response to a request from a State or Territory.[83] In both cases, there is also likely to be a need for the Governor-General to be satisfied that a state of 'domestic violence' exists – a term that lacks definition anywhere on Australia's statute books[84] – but would likely require substantial threat of or actual loss of life or Commonwealth assets.[85]

Secondly, the ADF is commanded by the Governor-General[86] through the Chief of the Defence Force.[87] As the King's representative in Australia, the Governor-General may also choose to deploy the ADF under the prerogative of the Crown or under doctrine of necessity;[88] neither of which would require authorisation by an Act of Parliament nor a decision by a Minister.[89] The ADF could also undertake certain "pre-deployment" actions and functions under its own initiative and without vitiating the exercise of either executive power[90] or the Crown prerogative.[91]

Lastly, whilst the High Court of Australia might have been willing to impose boundaries on the executive prerogative in *CPCF*,[92] these boundaries are not as persuasive to military deployments. This is generally because the call out order regime contains a caveat provision in section 51ZD of the Defence Act, essentially severing any need for the ADF to rely upon call out orders in every situation in which they proposed to deploy their capabilities: '[t]his Part does not affect any utilisation of the Defence Force that would be permitted or required, or any powers that the Defence Force would have, if this Part were disregarded'.[93]

Whether the activating condition is the making of a call out order or the ADF's inherent capability to deploy in response to an emergency situation will in turn determine the legality of ADF use of "hacking back" capabilities. For example, assuming that a call out order has been made by the Governor-General – and it matters not whether the situation involves a State request or the protection of Commonwealth interests – the deployment of ADF cyber capabilities is a command decision for the Chief of the Defence Force. If the hacking back is proposed in the context of a broader Police investigation or response, the use of ADF assets to do so would not only need to be 'reasonable and necessary' to achieve a purpose consistent

---

82 Defence Act, s 33.

83 Ibid, s 35.

84 Including the partner provision in the *Australian Constitution,* s 119.

85 Samuel White, Andrew Butler, 'Reviewing a Decision to Call out the Troops' (2020) 99 *AIAL Forum* 1, 58; Samuel White, 'Keeping the Peace of the iRealm' (2021) 42 *Adelaide Law Review* 1, 113

86 *Australian Constitution*, s 68.

87 Defence Act, s 9(1).

88 *R v Home Secretary of State for the Home Department, Ex parte Northumbria Police Authority* [1989] 1 QB 26.

89 Moore, n 29, 169.

90 Being 'a capacity to engage in enterprises and activities peculiarly adapted the government of the nation and which cannot otherwise be carried on for the benefit of the nation': *Victoria v Commonwealth and Hayden* (1975) 134 CLR 338; *Pape v Commissioner of Taxation* (2009) 239 CLR 1.

91 David Letts, Rob McLaughlin, 'Military Aid to the Civil Power', in Robin Creyke, Dale Stephens, Peter Sutherland (Eds.), *Military Law in Australia* (The Federation Press, Alexandria, 2019) 117-128.

92 *CPCF v Minister for Immigration and Border Protection & Anor* (2015) 255 CLR 514.

93 Cf. *CPCF*, where the High Court did not believe the Commonwealth could 'having failed to enter through the front door...enter through the back door and in effect achieve the same result by that means of entry': ibid, [141].

with the originating call out order,[94] but also at the request of the Police Force of the hosting State or Territory.[95]

Assuming it remained within the scope of the call out order, the ADF's use of hacking back capabilities could either be explicitly authorised by the Minister in an authorisation under the Defence Act[96], or apply by virtue of the Minister's declaration of a specified area[97] or in defence of critical infrastructure.[98] Hacking back could then be used by the ADF under a call out order if they needed to support more traditional security operations such as achieving the capture or recapture of an asset, in response to a hostage situation or to put an end to the state of domestic violence.[99]

On the other hand, the use of ADF assets to perform more traditional policing roles such as search and seizure[100] for evidentiary purposes is ultimately unsupportable in law, as the search powers available under a call out order must be observed at all times.[101] Nor do search powers under call out orders generally permit 'remote' or 'online' search and seizure, whether conducted under a search determination[102] or the power to search vehicles or aircraft.[103] The exercise of hacking back powers to search an offender's computer by military forces also butts up uncomfortably against the common law privilege against self-incrimination.[104]

Alternately, if the ADF deploys its cyber assets domestically under Crown or executive prerogative, it could operate without the constraints and requirements of Ministerial authorisations and simply rely on the balance of Crown power to support it 'responding to emergencies or keeping the peace'.[105] There are two significant problems with such a suggestion, especially in the contentious field of cyberattacks and military forces more generally.

The first issue is that the deployment of military forces without a statutory footing offends the 'very old common law proposition that the Royal prerogative does not extend to entering private property for the purposes of keeping the peace'.[106] Though the ADF has an absolute ability to protect and maintain the laws of the Commonwealth, this ability flounders in the absence of a clear and present danger to the Commonwealth itself – and, if such a danger existed, would justify the making of a call out order in the first place.

The second issue arising from the use of prerogative to deploy military forces is that the ADF loses the power and protection of the Defence Act in doing so. Not only are its officers

---

[94] Ibid, ss 33(3), 34(3), 35(3) and 36(3). The obligation is imposed by s 39(2), and subject to ss 39(3) and 40.

[95] Ibid, ss 40(1)(a)(ii) and 40(1)(b).

[96] Ibid, s 46(7)(i).

[97] Ibid, s 51D(2)(j).

[98] Ibid, s 51L(3)(h).

[99] Ibid, ss 46(1)(a), 46(5) and 46(7).

[100] Ibid, s 46(7)(d) and (e) permits an ADF member to search persons, locations or things for things that may be seized, or persons who may be detained, in relation to the call out order – the Act does not appear to limit such searches to the physical world.

[101] Ibid, s 51S(1).

[102] Defence Act, s 51C(1).

[103] Ibid, s 51E(2).

[104] See for example *X7 v National Crime Commission* (2013) 248 CLR 92.

[105] *Burmah Oil Co Ltd v Lord Advocate* [1965] AC 75; *R v Secretary of State for the Home Department; Ex parte Northumbria Police Authority* [1989] 1 QB 26.

[106] *Entick v Carrington* (1765) 19 St Tr 1029; cited in Samuel White, 'To Kill The Queen's Enemies (And Keep the Peace As Well)', *Australian Public Law* (blog, 23 February 2022) <https://www.auspublaw.org/blog/2022/02/to-kill-the-queens-enemies-and-keep-the-peace-as-well>.

and personnel open to criminal or civil liability in the pursuit of their duties,[107] but they lose all manner of statutory powers that could arise under the issue of a call out order or Ministerial authorisation. Deployments of ADF personnel during the management of the COVID-19 epidemic in Australia are instructive in this regard: they derived their powers from the various biosecurity and emergency management declarations, not from the provisions of the Defence Act.[108]

Finally (and for completeness) it is worth examining the assistance provisions of the IS Act previously covered as they might apply to domestic military operations. The position of ASD is complicated – they are not strictly part of the ADF even if they may employ or second 'any officer, sailor, soldier or airman'[109] and thus cannot be given orders by the Governor-General.[110] However, ADF may request ASD assistance either pursuant to the exercise of Defence powers[111] or as a request under the IS Act.[112] Both have their difficulties. It is possible to conceive a scenario where the Police make a request of the ADF who then make a request of ASD who might then collaborate with ASIO… hardly an efficient use of time or resources. It might also be seen by the Inspector-General of Intelligence and Security – the oversight body of IS Act agencies – as a way of sidestepping legitimate regulatory controls built into the law.[113]

Having concluded this analysis of the ADF's hacking back capabilities, it seems unlikely that deployment of military cyber capability inside Australia's territory would be *prima facie* lawful under the *Defence Act 1903*. Even accounting for the support provisions enabling IS Act agencies, without the very clear authority granted to the ADF by the Governor-General in the form of a call-out order, the use of hacking back capabilities by military personnel inside Australian territory seems to be extremely vexed.

This of course neglects collective effects of the residual powers of statehood available to Australia,[114] the Governor-General's command privileges[115] and the common law doctrine of necessity *viz* where a State can act in its own self-defence. Under that collection of legal powers and immunities, the ADF can respond to a foreign threat to Australia's national security as:

> [t]he conduct of war appears to be a prerogative power of the Crown and military forces exist primarily to execute this prerogative on the sovereign's behalf… the exercise of martial law in the factual circumstances of a war or insurrection is an aspect of prerogative power and not the common-law doctrine of necessity available to any person. This is a preferable view because it is not for any person to exercise the military power of the Crown or to claim to do so on its behalf.[116]

In those circumstances where the ADF is required to deploy domestically to confront a foreign threat, it would be permitted to utilise the full range of its warfighting capabilities (including "hacking back" and similar information warfare style methodologies).

---

[107] Defence Act, s 51S(2) and 123AA; cf. Defence Act, s 51N(1).

[108] Zoe Lippis, 'The Defence Act 1903 (Cth): A guide for responding to Australia's large-scale domestic emergencies' (2022) 45(2) *Melbourne University Law Review* 597, 648.

[109] Defence Act, s 4.

[110] Ibid, ss 4 and 17.

[111] Ibid, ss 46(7)(i), 51D(2)(j) and 51L(3)(h).

[112] IS Act, s 13(1)(a).

[113] Ibid, Note to s 13A(1).

[114] Including nationhood powers exercisable by Australia as a sovereign nation.

[115] Vested by the *Constitution*, s 68.

[116] Moore, n 29, 144-145.

We therefore turn finally to the legal status of policing hack backs.

C *Legal Authority for Police Hacking Back*

The final pathway for examining Australian legislative support for hacking back falls now to the status of policing forces under the *Surveillance Devices Act 2004* (Cth) ("SDA"), but also those bodies charged with various forms of integrity investigation such as the Australian Commission for Law Enforcement Integrity, Australian Criminal Intelligence Commission (ACIC), and the independent standing anti-corruption commissions of the States and Territories.[117] Each of the State and Territory police forces are also recognised as 'law enforcement agencies' and their members (including secondees) as 'law enforcement officers'.[118]

The SDA creates and maintains a warrant scheme for both Commonwealth law enforcement as well as State and Territory law enforcement agencies investigating State offences with a Federal aspect.[119] This is highly relevant to discussions around cybercrime and online malfeasance, as one of the core provisions of these types of offences are those which have 'involved an electronic communication'.[120] Divisions 4 (computer access warrants), 5 (data disruption warrants) and 6 (network activity warrants) of the SDA create the framework of most relevance to the discussion in this paper.

Computer access warrants under the SDA bear some similarity to their ASIO Act counterparts; however, the applicant is a 'law enforcement officer'[121] (not the Director-General of ASIO) and the issuing authority is an 'eligible Judge'[122] or 'nominated AAT [Administrative Appeals Tribunal] member'[123] (not the Attorney-General). The same test for reasonable grounds applies to the consideration of the application, but the criteria upon which those reasonable grounds are based changes significantly under the SDA (however, only the 'relevant offence' grounds are in scope here[124]).

The issuing authority must also consider numerous criteria in deciding whether to make a computer access warrant including the nature and gravity of the alleged offence,[125] the privacy of any person the warrant may affect,[126] whether alternative methods might obtain the same evidence or information,[127] and the likely value of evidence or intelligence obtained.[128] Like an ASIO warrant, if the issuing authority is satisfied then the warrant must sufficiently particularise both the alleged offences and the target computer which is the subject of the warrant.[129]

But there are difficulties with the use of computer access warrants as a vehicle to hacking back offending targets; again, like their ASIO equivalents the warrant only authorises access to the computer, not damage or disruption. Law enforcement officers are permitted to access

---

[117] SDA, s 6(6) and (7).
[118] Ibid, s 6(7).
[119] Ibid, s 7.
[120] Ibid, s 7(a); *Australian Federal Police Act 1979* (Cth), ss 4AA(1)(c) and 4AA(3)(e).
[121] Ibid, s 6 and 27A.
[122] Ibid, ss 12 and 27C.
[123] Ibid, ss 13 and 27C.
[124] Ibid, s 27C(1).
[125] Ibid, s 27C(2)(a).
[126] Ibid, s 27C(2)(c).
[127] Ibid, s 27C(2)(d).
[128] Ibid, s 27C(2)(e).
[129] Ibid, s 27D(1)(a), (b)(ii), (b)(vii)-(ix) and 27D(4).

and obtain the data subject to the warrant only – this does not authorise deletion, modification, destruction or damage to any data or system.[130] Concealing the access also does not authorise the use of disruptive, destructive, or damaging methodologies beyond the bare minimum necessary to achieve the purposes of the warrant.[131]

Nor do network activity warrants permit any broader forms of "hacking back" in the terms described in this paper.[132] Like computer access warrants, the purpose of network activity warrants is to map out and obtain evidentiary information relevant to a 'criminal network of individuals'.[133] This in turn is defined as a 'electronically linked group of individuals' who use, communicate or facilitate the commission of a relevant offence using online or digital methods.[134] These warrants are also more restricted in their application, in that only the chief officer of the AFP or ACIC may apply for a network activity warrant,[135] and that data on the target computer relates to the network of individuals and is 'relevant to the prevention, detection or frustration of one or more kinds of relevant offences'.[136] Destruction or damage to data or systems is generally not permitted.[137]

The final form of warrant – data disruption warrants – are also limited to law enforcement officers of the AFP or Australian Crime Commission.[138] In order to bring an application, relevant offences must have been or are about to be committed, and the disruption of data in a target computer must achieve either the frustration of the commission of the relevant offence relating to that data (data-based warrant[139]), or the frustration of offences of a similar kind (offence-based warrant[140]). Data disruption warrants also requiring an endorsement from a more senior officer within the agency who has 'relevant skills, knowledge, and experience to endorse the making of applications for the issue of data disruption warrants'.[141] Again, the list of matters to be considered by an eligible Judge or nominated AAT member is daunting and comprehensive.[142]

Quite unlike computer access warrants and network activity warrants, data disruption warrants clearly fall within the parameters of "hacking back" as this paper describes it. Data disruption warrants may be applied for unsworn, permitting law enforcement to respond quickly to incidents as and when they occur (noting the requirement to provide a sworn application within 72 hours of the making of the application[143]). The warrants precisely authorise and permit the damage, deletion and destruction of data which is subject to the warrant as well as authorising the alteration, copying or deletion of *any other data* on the target computer whilst the warrant is in force – so long as the disruption of offending named in the warrant is achieved.[144] Although the usual caveat of restricting damage that would

---

[130] Ibid, s 27E(2)(c), (d) and (g).
[131] Ibid, s 27E(7) and (8).
[132] Ibid, s 27KP(1)(c), (d), and (e)(ii).
[133] Ibid, s 27KK(1).
[134] Ibid, s 7A.
[135] Ibid, s 27KK(1).
[136] Ibid, s 27KK(1)(b)(ii).
[137] Ibid, s 27KP(6).
[138] Ibid, s 27KA(1).
[139] Ibid, s 27KA(1)(c)(i).
[140] Ibid, s 27KA(1)(c)(ii).
[141] Ibid, ss 27KBA(4)(b) and (5)(b), 27KBB(4)(b) and (5)(b).
[142] Ibid, s 27KC(1)-(3).
[143] Ibid, s 27KA(4) and (5).
[144] Ibid, s 27KE(2)(c)(vi), (d) and (e)

impinge on lawful uses of the target computer applies,[145] contextually the bar is much lower given that the purpose of the warrant is disruption and not evidence gathering.

It is also apposite to note that the provision of assistance to law enforcement – either by IS Act agencies or by ASIO – is much cleaner and easier under the imprimatur of a data disruption warrant. Both ASD capabilities (leveraging the technical expertise and advice function[146]) and ASIO (who have a specific law enforcement assistance function[147]) could be brought to bear under the supervision and direction of law enforcement capabilities. Equally, were ADF assets to deploy following a call out order or exercise of executive prerogative, their subordination to law enforcement and incorporation under the warrant regime would clothe them in an appropriate degree of legal protection. The construction of the SDA also seems to prefer this approach, as it requires a data disruption warrant to 'authorise the doing of specified things … in relation to the *relevant target computer*'[148] (emphasis added). The Act thus remains entirely agnostic about *who* does those specified things, whether it is a law enforcement officer, ASD or ASIO agent, or military officer or enlisted.

D *Conclusion to Section 2*

The strongest support under Australian law for the use of "hacking back" capabilities appears to be under the SDA as a primarily law enforcement focused exercise. The use of such capabilities under either the intelligence agencies or military forces of Australia, acting alone or in concert, are riddled with statutory imperfections and linguistic difficulties.

Far from being an academic concern, this lack of precision not only compromises the legitimacy of such operations but potentially renders both intelligence and military personnel liable to criminal charge or civil suit. The use of such a controversial capability cannot be supported by reference to imprecise legal frameworks and ambiguous prerogatives.

By focusing on law enforcement as the primary counter-cybercrime agency in Australia, we achieve a degree of 'coherence and consistency between the essential elements of the regime and correlative authorisations elsewhere in legislation'.[149] But could Australia do it better? The next section will examine whether Australia should alter its laws to better reflect circumstances under which it may use its "hacking back" capabilities in the future.

III When Can Australia Hack Back?

Having examined some of the various legal frameworks under which "hacking back" would be legitimised, there remains a policy question at what point recourse to those frameworks should be had. Obviously not all cyber incidents and not all cyber-crimes would qualify for an immediate and overwhelming response by national security agencies and might breach Australia's obligations under international law.[150] On the other hand, failing to respond adequately to a cybersecurity incident may result in further damage and cost which could

---

[145] Ibid, s 27KE(7).

[146] IS Act, s 7(e) and 13(1)(a).

[147] ASIO Act, ss 17(1)(f) and 19A(1)(d).

[148] SDA, s 27KE(1).

[149] Letts and McLaughlin, n 86, 130.

[150] For example, the Budapest Convention on Cybercrime: Roderic Broadhurst, Lennon Y. C. Chang, 'Cybercrime in Asia: Trends and Challenges', in Jianhong Liu, Bill Hebenton, Susyan Jou (Eds.), *Handbook of Asian Criminology* (Springer, New York, 2013) 49-63.

have been avoided. What is therefore required is a policy that has 'the aim of normalizing uncertainty…[and] the aspiration that the uncertainty, the exceptional, be tamed'.[151]

There are international examples of how such a policy could be formulated. Kesan and Majuca describe that hack backs should only be allowed where three conditions can be met: one, that obtaining a court order or similar restraint is unlikely to be successful; two, there is a serious prospect that the hack back will not impact innocent third parties; and three, the damage to the victimised system cannot be adequately mitigated.[152] The trigger for engaging in such activities appears to largely follow a cost-benefit analysis dependent on the likelihood of success by the defender versus the likelihood of collateral damage, mistakes in attribution, lack of legitimacy and the possibility of normalising destructive hacking.[153]

Another benchmark available in international law is the doctrine of countermeasures, being the engagement in 'measures that would otherwise be contrary to the international obligations of an injured State vis-à-vis the responsible State, if they were not taken by the former in response to an internationally wrongful act by the latter in order to procure cessation and reparation'.[154] Under this doctrine, "hacking back" would only be permitted under two conditions: firstly, the State engaging in a prior wrongful act must be identifiable and attributable for that act; and secondly, the hack response must produce a temporary result that is 'instrumentally directed to induce the responsible state to cease its violation'.[155] The obvious limitations upon such a trigger are the cessation of the originating cyberattack, as well as any response that would threaten human rights or amount to an unacceptable use of force.[156]

A third form of policy construction was envisaged by Lahmann, where he examined the lawfulness of Germany's 2005 Aviation Security Act – which permitted the executive to shoot down planes hijacked by terrorists in a 9/11 scenario – and the potential for the Act to 'directly affect innocent individuals and their human dignity'.[157] In Lahmann's view, such regimes needed to abide by four conditions:[158]

- An operative cyber emergency regime needs to comprise precise and workable definitions of key concepts;
- The regime should only require consideration of stakeholder interests if they will foreseeably and directly be negatively affected by the "hacking back" operation;

---

[151] William Vazquez Irizarry, 'Exception and Necessity: The Possibility of a General Theory of Emergency' (2013) <https://law.yale.edu/sites/default/files/documents/pdf/sela/VazquezIrizarry_Eng_CV.pdf> 2.

[152] Kesan and Majuca, n 20, 1.

[153] Mils Hills, 'The Deregulation and Swarming of Cyberwarfare: The Need for and Limitations of Law in Enabling Aggressive Hacking-back and Pre-Emption' (2014) 3(1) *Journal of Law and Cyberwarfare* 43; Lennon Y.C. Chang, Lena Y. Zhong, Peter N. Grabosky, 'Citizen co-production of cyber security: Self-help, vigilantes, and cybercrime' (2018) 12(1) *Regulation and Governance* 101.

[154] UN International Law Commission, *Draft Articles on Responsibility of States for Internationally Wrongful Acts*, UN GAOR, 53d Sess, Supp No 10, UN Doc A/56/10 (2001) art 128.

[155] Oona Hathaway, 'The Drawbacks and Dangers of Active Defense' (Paper presented to the 2014 6th International Conference on Cyber Conflict, Tallinn, 2014) 45.

[156] Eliza Fitzgerald, 'Helping states help themselves: Rethinking the doctrine of countermeasures' (2016) 16(1) *Macquarie Law Journal* 67; Gary Corn, Eric Jensen, 'The use of force and cyber countermeasures' (2018) 32(1) *Temple International and Comparative Law Journal* 127.

[157] Henning Lahmann, 'Hacking Back by States and the Uneasy Place of Necessity within the Rule of Law' (2020) 80(1) *Heidelberg Journal of International Law* 453, 471.

[158] Ibid, 473.

- The regime should eschew *ex ante* attribution (which may not be technically possible at the time) in favour of *ex post* assessment of the legitimacy of the "hacking back" action; and
- As a State with a cyber offensive capability, it has a duty to put some legal process in place that follows principles of the rule of law and ensures consistency.

A fourth, more consequentialist approach by Carr and Schmitt described six criteria: severity, immediacy, directness, invasiveness, measurability, and presumptive legitimacy.[159] However, it is worth observing that any "hack back" which meets all six criteria is suggested to amount to an armed attack which would warrant a military response. As we are focusing the use of State capabilities against non-State actors within a domestic threat scenario, and in many (but not all) circumstances these use of an "armed attack" cannot be justified.

All of these benchmarks stipulate that "hacking back" must not be conducted in an unfettered manner and requires some form of soft limitations, even if the law permits that conduct to be engaged in. For example, Germany's 2005 Aviation Security Act may *legally* allow the shooting down of a hijacked plane – a concept also authorised under Australian law[160] – the question remains whether the Executive could *legitimately* do so in every circumstance. Another obvious limitation for each of these proposed policy frameworks listed above is that they take the approach of "hacking back" as an activity by one State against another, as opposed to our present analysis which involves the activity of a State (Australia) against non-State actors located within its own territory.

In fact, the literature is all but silent on the legitimacy of hacking back as a tool of State power. Arguments broadly seem to centre on three themes: is a hack back effective, proportionate, and preferable compared to other less intrusive and less damaging options.[161] The methodologies of hack back are also highly relevant: actions which install "beacons" – forms of spyware which identify or flag the attacker's computer or keystrokes – are generally considered a more appropriate and rational response than erasing or modifying data in the target system, or ultimately the State's use of ransomware or cyberweapons.[162]

Given the lack of academic consideration in this area, we will propose four matters that national security agencies in Australia (or other States) should have regard to prior to or during the conduct of hacking back activities. These matters draw the important link between what is *legally* authorised under Australian law from the preceding sections and what is *legitimately* warranted. In so doing, we synthesise Lahmann's analysis of the international doctrines of emergency and necessity with

A *Size of Proposed Intervention by Agency*

Under both an ASIO Act and SDA warrant, different levels of interference in a target computer might be warranted. Whilst the warrant must describe what methodologies the

---

[159] Jeffrey Carr, *Inside Cyber Warfare: Mapping the Cyber Underworld* (O'Reilly, New York, 2010); citing Michael N. Schmitt, 'Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework' (1999) 37(1) *Columbia Journal of Transnational Law* 885, 913-915.

[160] See Defence Act, ss 46(5)(d) and 51N(3)(a)(iii), which permits a member of the Defence Force to use lethal force to destroy an aircraft or vessel in limited circumstances.

[161] Ronald L. D. Pool, Bart Custers, 'The police hack back: Legitimacy, necessity and privacy implications of the next step in fighting cybercrime' (2017) 25(2) *European Journal of Crime, Criminal Law and Criminal Justice* 123; Dennis Broeders, 'Private active cyber defense and (international) cyber security—pushing the line?' (2021) 7(1) *Journal of Cybersecurity,* DOI:10.1093/cybsec/tyab010.

[162] Lauren Fiotakis, 'Beacons: A Viable Solution to the Ever-Evolving Problem of Corporate Data Breaches' (2021) 19(3) *Northwestern Journal of Technology and Intellectual Property* 289.

warrant authorises can be performed,[163] exactly whether the effect those methodologies result in is appropriate will vary dependent on the circumstances. We have provided a non-exhaustive spectrum of these effects in Figure 1.

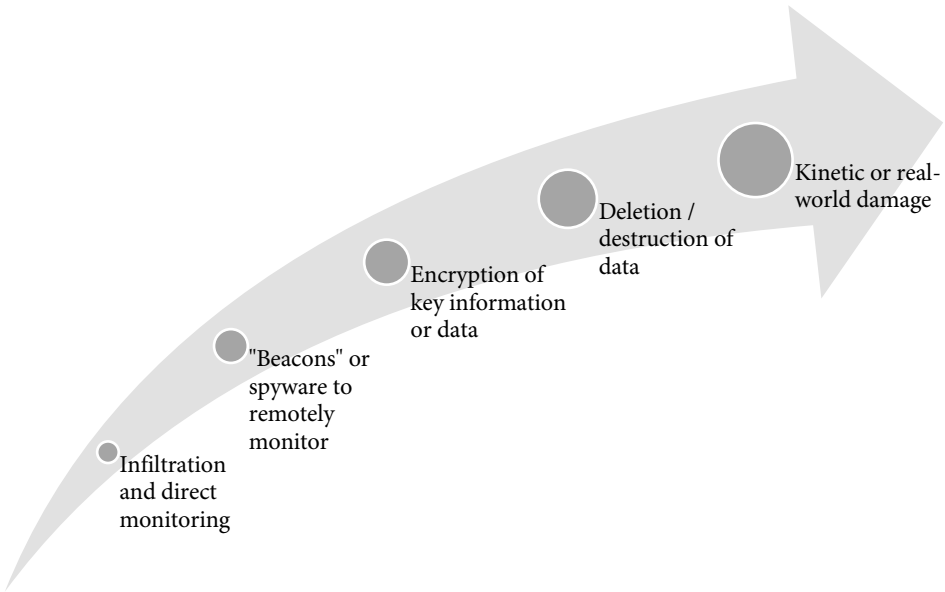


Figure 1. Spectrum of possible effects of "hacking back"

Consider the following by reference to Figure 1. It may be sufficient for investigators to simply gain access to the data on a target computer under an ASIO Act or SDA warrant, and from there directly monitor what activity occurs (including communications between the various parties of interest). This has the benefit of being conducted covertly and may not necessarily involve any physical changes to the data in the target computer. A warrant authorising this type of effect might be more reasonable in circumstances involving preparatory or anticipatory offences where the "hacking back" might be supplemented traditional law enforcement techniques or procedures.

An escalation could involve the installation of "beacons" as described earlier (which flag or identify the physical location of the target computer or persons of interest) or spyware (which might allow for remote monitoring, logging of keystrokes, etc.). Beyond those measures, the spectrum in Figure 1 displays increasingly more obvious forms of harm and damage, such as encrypting the target's data (and thus rendering it unusable[164]), deleting or destroying it, or manipulating a system or some of its connected components to generate a kinetic or real-world damage.[165] Deploying a "hack back" which causes real-world harms would need to be reserved for the most serious of incidents and really as a last resort for fear of normalizing cyber-attacks.[166]

---

[163] ASIO Act, s 25A(3A)(b); SDA, s 27KE(1) and (2).

[164] With respect to law enforcement using "ransomware" style programs in a law enforcement context, see Paul Ohm, 'The investigative dynamics of the use of malware by law enforcement' (2017) 26 *William & Mary Bill of Rights Journal* 303; Vania Mia Chaker, 'Chimaera Unleashed: The Specter of Warrantless Governmental Intrusion Is a Phantom That Has Achieved Greater Life in the Ether of Internet Communications' (2018) 22(2) *Journal of Technology Law & Policy* 1.

[165] Such as how the Stuxnet worm operated to unbalance nuclear centrifuges: Farwell and Rohozinski, n 8.

[166] Lahmann, n 155, 469; see also Jason Healey, 'The implications of persistent (and permanent) engagement in cyberspace' (2019) 5(1) *Journal of Cybersecurity*, tyz008; Homer A. La Rue, 'Outsourcing the Cyber Kill

B *Scale, Scope and Sophistication of the Target Offender or Person of Interest*

A technical method of determining the scale, scope and sophistication of the target offender is to examine which "layer" is being attacked. Under the Open System Interconnection (OSI) model, each system contains seven layers (moving from highest to lowest levels of abstraction):

- Application layer
- Presentation layer
- Session layer
- Transport layer
- Network layer
- Data link layer
- Physical layer

Attacks which target the lower layers are generally more favoured in attacks against infrastructure involving national security,[167] whereas those attacking the higher layers are traditionally associated with wider network or social engineering incidents.[168]

Agencies should then have regard to the nature of the offenders that can be identified and assessed. Is the suspected person a foreign State actor, or an agent acting on behalf of a foreign principal? Or are they merely a motivated opportunist or "lone wolf"? What resources are they suspected or known to have, either in respect of 'relevant offences'[169] or 'security matter'[170] (which could ground an application for a warrant)? If they have or are already launching a cyberattack of some kind, how much time and space do they have to perform follow-up or secondary attacks? Have they made specific threats against known vulnerabilities or targets, or are the threats vague and lacking in substance?

Each of these questions shapes the form of the cyber response that is warranted under either of the ASIO Act or SDA. Obviously a well-resourced, highly motivated actor working on behalf of a foreign State will require a stronger and more decisive "hacking back" response than that used against an individual. In the same manner, offenders with a deep resource pool and the capabilities to conduct follow-up offences should be considered for more serious interventions to disable future incidents and discourage repetition. The nature of the target also helps shape the degree to which different national security agencies might need to coordinate their responses to "hacking back" – the more sophisticated, the more coordination is needed.

C *Quantum and Impact of Actual and Anticipated Harms*

The third matter that national security agencies should have regard to involves the severity of actual or anticipated harms to the affected systems. In the early days of a cybersecurity incident or attack, this may be difficult to achieve; alternately, in the case of recent

---

Chain: Reinforcing the Cyber Mission Force and Allowing Increased Contractor Support of Cyber Operations' (2021) 12(1) *Journal of National Security Law & Policy* 583.

[167] Bernard Everett, 'Optically transparent: The rise of industrial espionage and state-sponsored hacking' (2013) 10 *Computer Fraud & Security* 13.

[168] Lynne Yarbro Williams, 'Catch Me if You Can: A Taxonomically Structured Approach to Cybercrime' (2008) *Forum on Public Policy* 28.

[169] SDA, s 27KA(1)(a).

[170] ASIO Act, s 25A(2).

ransomware attacks against critical infrastructure like a hospital or gas pipeline the actual harms can be more obvious.[171]

A mere risk-benefit analysis is insufficient.[172] A "hack back" by national security agencies to protect a piece of critical infrastructure[173] (even in anticipatory or *ex ante* circumstances) might well invoke a higher order response, even if the scope of the proposed harms to that infrastructure was of a lesser nature. Situations where harms have already occurred – such as the conduct of a terrorist attack, or the use of a weapon of mass destruction for example – which might automatically warrant higher-order national security responses than anticipatory or preparatory actions, or where investigations into conduct are still covert.

A better rule of thumb might be to look to the emergency doctrine in international law referred to by Lahmann and consider whether the actual or anticipated harms involve "grave and imminent peril" to an "essential interest" of Australia.[174] This is especially the case where warrants may be obtained pre-emptively.[175] In cases where national security agencies are contemplating a higher order "hacking back" response, they ought to give consideration to whether their target poses a grave and imminent threat to an essential interest of Australia, and be capable of articulating exactly what those forms of threat and interests are.

Articulating how threats meet those definitions would also be useful for issuing authorities, to assist them in determining whether the issue of warrants is reasonable, necessary and proportionate. Independent arbiters in each warrant regime, being the Commonwealth Ombudsman for the SDA[176] and the Inspector-General of Intelligence and Security in respect of ASIO,[177] would also be better placed to assess the legitimacy of each warrant granted if the application for them was couched with reference to those terms.

D *Likelihood of Collateral Damage*

Finally, agencies should have regard to the likelihood of possible harms to innocent third parties, whether they are other users of a system or network, or in the case of mistaken attribution, i.e., an actor used an interposing computer system to disguise their digital footprint, resulting in agencies "hacking back" into the wrong system. Lahmann counselled that "hacking back" regimes should only consider and protect those private actors whose interests would 'foreseeably and directly be negatively affected'.[178] One of the ways in which he envisaged the regime could take account of these interests would be the issue of warnings to owners or custodians of affected systems of an imminent "hack back", permitting those

---

[171] Charlie Osborne, 'Colonial Pipeline ransomware attack: Everything you need to know', *ZDNET* (online, 13 May 2021) <https://www.zdnet.com/article/colonial-pipeline-ransomware-attack-everything-you-need-to-know/>; Dan Milmo, 'NHS ransomware attack: what happened and how bad is it?', *The Guardian* (online, 12 August 2022) <https://www.theguardian.com/technology/2022/aug/11/nhs-ransomware-attack-what-happened-and-how-bad-is-it>.

[172] Eva Ignatuschtschenko, 'Assessing Harm from Cyber Crime', in Paul Cornish (Ed.), *The Oxford Handbook of Cyber Security* (Oxford University Press, Oxford, 2021) 127-141, 134.

[173] Such as anything defined in the *Security of Critical Infrastructure Act 2018* (Cth), Div 2 of Pt 1.

[174] See for example the treatment of those terms in Michael Schmitt (Ed.), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Tallinn, 2017) 135.

[175] Involving 'one or more relevant offences of a particular kind… are about to be, or are likely to be, committed': SDA, ss 27KA(1)(a) and 35B(1)(a). See also Sayako Quinlan, Andi Wilson, *A Brief History of Law Enforcement Hacking in the United States* (Report, New America, September 2016) <https://d1y8sb8igg2f8e.cloudfront.net/documents/History_Hacking.pdf>.

[176] SDA, ss 49C and 57.

[177] ASIO Act, 34HB.

[178] Lahmann, n 155, 474.

owners or custodians time to implement less intrusive or damaging interventions that might address the original incident.

For obvious reasons, the notion of law enforcement (AFP) or intelligence agencies (ASIO) issuing warnings ahead of a hack back activity – where those agencies could be responding to serious criminal offences or threats to national security – is hardly a practical one. However, both the ASIO Act and SDA do import requirements involving the consideration of stakeholders who might be affected. Both ASIO Act and SDA warrants incorporate provisions which require the applicant to give, and the issuing authority to consider, reasonably foreseeable impacts on innocent third parties.[179] Both warrants also do not extend to authorising the damage of systems which are not contained within the scope of the warrants, and which might be foreseeably affected by the activity.[180]

If something were to go wrong, there are mechanisms in Australian law to hold national security agencies accountable if they have erroneously targeted an entirely uninvolved system or acted disproportionately.[181] The SDA has a specific head of statutory power under which an aggrieved person may obtain compensation from the Commonwealth.[182] In any event, the oft-cited case of *A v Hayden*[183] – which involved allegations against ASIS operatives who, during a training exercise brandished firearms in a hotel, took a manager hostage, and caused criminal damage to the premises – clearly shows that illegal acts undertaken by organs of Australia's national security apparatus will not be excused unless explicitly authorised by statute or order of the court. *A v Hayden* supports the proposition that any act undertaken by an agency of the State involving "hacking back" that is not authorised by law will attach civil liability to the Crown.

IV Conclusion & Lessons Learnt

By examining Australia's statutory frameworks for deployment of "hacking back" capabilities amongst its intelligence, military and law enforcement arms, there are several observations that we can make in conclusion.

Firstly, Australian law largely prohibits the granting of Ministerial authorisations and directions for "onshore" or domestic targets,[184] and the recent Richardson Review of the legislation of NIC agencies recommended against any such laws being changed.[185] In all other cases, the ASIO Act and SDA require the issue of a warrant (either by the Attorney-General or by a Judge or authorised AAT member respectively) on their satisfaction of certain matters. However, the same agencies would largely be authorised to conduct hack backs in the case of protecting critical infrastructure from cyber offensive activities – especially in response to emergency or emergent situations.

In the context of the ASIO Act, the modification or deletion of data must be relevant to the security matter for which the warrant is being sought, subject to the limitations imposed regarding interfering with lawful use of computers or causing loss or damage to other persons.[186] For the SDA, an officer of the AFP or ACIC must set out their reasonable

---

[179] ASIO Act, s 25A(2) and (3A); SDA, s 27KA(3) and 27KC(1)(b).

[180] ASIO Act, s 25A(5); SDA, ss 27KC(2)(cb), (cc) and (cd).

[181] Lahmann, n 155, 475.

[182] SDA, s 64.

[183] [1984] HCA 67; (1984) 156 CLR 532.

[184] ASIO Act, s 27A(9); *Telecommunications (Interception and Access) Act 1979* (Cth), s 11D(5).

[185] The Richardson Review, n 29, [3.84] and [9.121]-[9.125].

[186] ASIO Act, s 25A(4)(a) and (b); cf. s 25A(5).

suspicion that 'disruption of data held in the target computer is likely to substantially assist in frustrating the commission' of offences either involving the data in the target computer, or relevant to the application.[187] The relative satisfaction of those issuing authorities is not reached lightly, and promotes a degree of consistency across the situations in which "hacking back" will be conducted by our national security agencies.

Secondly, these statutes place ambiguous and imprecise boundaries on the lawful mechanism for counter-cybercrime capability to be used in a domestic threat scenario. The closest legitimate analogue is the issue of data disruption warrants under the SDA to law enforcement officers; however, the exact pathway to satisfying the issuing authority, the nature of what is authorised under a data disruption warrant, and the overlap of such warrants with cooperation provisions in the ASIO Act and IS Act in respect of domestic targets remain unnecessarily ambiguous. The cybercrime being targeted would also need to be relatively large in scale and/or organised. It is highly unlikely that Australia would deploy significant cyber offensive capabilities against foreign (or domestic) scammers or individual instances of cybercrime.

Thirdly, though Australia's legislation is broadly compliant with normative customs regarding "hacking back", the policy parameters of such activities require significant academic and governmental examination. By subjecting the capability to a specific warrant regime, the Australian government has clearly established a need to place legislative protections and oversight around this most contentious of powers. Yet much more research needs to be conducted into determining appropriate policy controls and thresholds at which certain methodologies might be employed under SDA data disruption and ASIO computer access warrants.

Fourthly, Australia's data disruption warrant regime needs to be deconflicted from the perspective of ASD. Given its unique position as an IS Act agency, as well as an agency that supports military operations and law enforcement, the precise jurisdiction of ASD operations conducted in a domestic environment (i.e., inside Australian territory or against Australian residents or citizens) needs to be narrowed down with far greater precision. A Ministerial direction is perhaps the most appropriate action in that case.

In concluding, we also observe that our examination of "hacking back" focuses solely on the Australian experience, and neglects that of the other Five Eyes countries with whom they may share methodology (New Zealand, Canada, US and UK). We also have not considered the actions or examined the legitimacy of others active both in cyberspace and the Indo-Pacific region we share, such as Pakistan, India, China, North Korea or Singapore. Further, the voice of our intelligence agencies – represented in the form of empirical studies of their methods and techniques – is sorely lacking in the debate. All three areas are fruitful avenues for future research.

Yet in our jurisdiction there exists patchy legal frameworks in which security translates poorly from being a physical concept. Australia appears to have reconciled itself to the idea that its national security agencies – either ASD or the AFP – can hack into the computers of cybercriminals no matter where they are in the world, and no matter how challenging an intrusion into foreign sovereignty that may be. If those capabilities suffer from a lack of both transparent legality and legitimacy when theoretically applied to domestic threats, that is a situation that cannot be allowed to continue.

Alternately, though we seem more than capable of deploying "hacking back" capabilities, Parliament may have chosen the very deliberate means of excising them from being used *inter*

[187] SDA, s 27KA(1)(c).

alia to combat domestic threats. Given Australia's uncertain future and contested geostrategic position in the Indo-Pacific, the importance of ensuring its national security agencies operate with the powers and functions they need within an appropriate accountability and regulatory framework, cannot be understated.

TOWARDS SOCIETY OF QUANTUM TOMORROW

KATRI NOUSIAINEN,[*] JOONAS KESKI-RAHKONEN,[†] TIM MCDONALD[‡] AND SASCHA FELDMANN[§]

ABSTRACT

Quantum technologies encompass a wide and ever-growing field of applications which leverage unique quantum-mechanical properties for performing tasks that existing, classical technologies could not. To benefit all of society optimally, it is a linchpin that we are aware of not just the potential of these emerging technologies, but also of the risks involved, to analyze them thoroughly, and to take political action accordingly – an A-cubed approach. Here, we follow this approach by first depicting a picture of the future quantum society. We then present a five-point roadmap to examine social, ethical and economical dimensions of quantum technologies, with a call for further discussions on the prospective legal and policy framework. Finally, we look over possible steps we can take on the path towards a bright society of quantum tomorrow.

CONTENTS

I INTRODUCTION

The brave new world of quantum technologies is upon us. As we are entering into the new era of quantum technologies, the sovereign states, institutional and organizational operators, as well as businesses should reflect upon the social and legal relevance of this technological progress. Towards this goal, we propose the A-cubed approach - awareness, analysis, and action - being employed for socio-economic framework construction (visualized in Fig. 1).

[*] Teaching Faculty at Harvard University and Postdoctoral Research Fellow, Program on Negotiation, Harvard Law School.
[†] Postdoctoral Researcher and Teaching Fellow, Department of Physics, Harvard University.
[‡] Assistant Policy Researcher, RAND Corporation.
[§] Research Group Leader and Rowland Fellow, The Rowland Institute, Harvard University.

Figure 1. A-cubed approach to integrate new emerging quantum technologies into a part of future society: Raise awareness, analyze the environment, and take appropriate action.

First, it is of crucial importance that we are aware of these new technologies; secondly, their implications and ramifications must be carefully analyzed within their respective environments, and lastly appropriate actions should be taken to work for their development but also to safeguard for individual and social rights. In the international arena, the path dependency process plays a crucial role in finding balance with various rights and obligations. In the quantum-technological environment, we see that the legal design approach, which aims to empower within improving, supporting, and demonstrating, can pave the way towards a legal framework that is transparent, human-centric, efficient, and comprehensible as well as foster equality and nondiscrimination. We recognize that it is better to be proactive than reactive.

II AWARENESS: FUTURE IS QUANTUM

To take full social benefit of new emerging quantum technologies, we should first be aware of what they encompass. Based on the economic theories on path dependency,[1] knowledge is prevalent in society and learning is considered as gradual. We encounter the same dilemma with quantum technologies: we have the golden opportunity to embrace learnings from the past, to enhance, and build on the understanding that we have today. As a matter of exemplary incident, we may study the policy frameworks of the 1990's for the internet.[2] A more modern reference point for the embodiment of quantum technologies is an envisaged social-legal-ethnic framework for nanotechnology.[3] Though, just reflecting the past is not enough, and a new point of view on quantum matters is a linchpin, even when the learning process comes with a cost. At the end of the day, comprehensive understanding and

[1] See for instance FA Hayek, *New Studies in Philosophy, Politics, Economics, and the History of Ideas* (University of Chicago Press, 2018).

[2] Andrew Chadwick and Christopher May, 'Interaction between States and Citizens in the Age of the Internet: "E-Government" in the United States, Britain, and the European Union' (2003) 16(2) *Governance* 271; Jan Van Dijk, *The Deepening Divide: Inequality in the Information Society* (Sage Publications, 2005).

[3] See for instance, Barry L Shumpert et al, 'Specificity and Engagement: Increasing ELSI's Relevance to Nano–Scientists' (2014) 8(2) *Nanoethics* 193; Antonio G Spagnolo and Viviana Daloiso, 'Outlining Ethical Issues in Nanotechnologies' (2009) 23(7) *Bioethics* 394; Tsjalling Swierstra et al, 'Converging Technologies, Shifting Boundaries' (2009) 3(3) *NanoEthics* 213.

knowledge are pivotal for successful innovation, development, and competition. In this respect, we should raise quantum awareness – there is a new game in town.

The topic of "quantum" has attracted a lot of attention in recent years, starting from science and media until it now has reached every part of society through the first applications coined "quantum" in the consumer market. Yet, for the majority of people it is not actually clear what "quantum" really means, leave alone what implications quantum technologies might have for society in the future. Right now, we find ourselves in the midst of a quantum revolution – however, not in the first, but the second one in the history of quantum physics.

The first quantum evolution began with the discovery of quantum mechanics and its laws in the beginning of the 20th century, pioneered by physicists like Planck, Einstein, Bohr, Heisenberg, to acknowledge only a few of the most notable ones. Following the initial observations on the quantized nature of our underlying reality and the dualism between waves and particles, the understanding of these principles has by now enabled inventions that have had a lasting impact on the development of civilization. Examples include the transistor or laser - essential building blocks in modern computers and telecommunication, thus forming the backbone of our digitalized society and the motor of globalization.

The second quantum evolution is currently underway, and most scientific efforts in the present focus on building, fully controlling and taking advantage of quantum systems. One widely anticipated embodiment of emerging quantum technologies is a functional quantum computer. In a nutshell, a quantum computer is a device that harnesses the properties of quantum mechanics to store data and to perform computational tasks.[4] Although conventional computers have been present in some form since the 20th century, the possibility of a computer operating exclusively with quantum mechanical principles was put forward in the 1980s[5], which has led to the current rise of the quantum computing and information field[6], onside stimulating the evolution of other quantum technologies.

Even though any computational challenge with a classical computer can also, in principle, be targeted by a quantum computer, there are mathematical problems where a quantum computer outshines its classical counterparts spectacularly.[7] Whereas classical computers, such as modern laptops and smartphones, encode information in binary bits, *i.e.*, as either zeros or ones, the fundamental element of a quantum computer is a qubit, which in a simplified picture can be a zero and one at the same time. These quantum bits are nowadays

---

[4] There are a myriad of text books on quantum computing and information, see for example Masahito Hayashi, *Quantum Information Theory* (Springer, 2016); David McMahon, *Quantum Computing Explained* (John Wiley & Sons, 2007); Michael A Nielsen and Isaac L Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2019); Venkateswaran Kasirajan, *Fundamentals of Quantum Computing* (Springer Nature, 2021); John Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018); Mark M Wilde, *Quantum Information Theory* (Cambridge University Press, 2017); Noson S Yanofsky and Mirco A Mannucci, *Quantum Computing for Computer Scientists* (Cambridge University Press, 2008).

[5] Richard P Feynman, 'Simulating Physics with Computers' (1982) 21(6-7) *International Journal of Theoretical Physics* 467; Paul Benioff, 'The Computer as a Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines' (1980) 22(5) *Journal of Statistical Physics* 563; D Deutsch, 'Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer' (1985) 400(1818) *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 97.

[6] TD Ladd et al, 'Quantum Computers' (2010) 464(7285) *Nature* 45; Antonio Acín et al, 'The Quantum Technologies Roadmap: A European Community View' (2018) 20(8) *New Journal of Physics* 080201.

[7] See, for instance, Takashi Yamakawa and Mark Zhandry, 'Verifiable Quantum Advantage without Structure' (2022) *arXiv* 2204.02063.

realized in various physical set-ups.[8] Over the last decade, significant progress has been made, and real-life quantum computers exist. For the time being, classical computers seem to handle most daily computational tasks with which a quantum computer is challenged, with a similar efficiency. On the other hand, based on some companies, such as IBM and Google, we are on the verge of achieving a genuine quantum advantage[9]. However, current devices are early examples of noisy intermediate-scale quantum computers, and these have just begun to find important applications in quantum simulation and chemistry.[10] The biggest adversary for the triumph of quantum computing has been the fragile nature of its building blocks, qubits, that will be rendered into an old-fashioned classical computer employing zeros and ones, i.e., bits, by unwanted disturbance, colloquially known as decoherence. A future challenge is to design devices with an ability to shield their quantum nature from decoherence, while remaining easy to operate. Whether it will be more robust, error-tolerant quantum processors[11] or better ways to correct and mitigate errors[12], or both - only time will show us.

It is the peculiar nature of qubits that gives a quantum computer an edge to solve certain complex problems better than the best conventional supercomputers.[13] Figure 2 gives an overview of potential future applications which could harness this "quantum advantage". For completeness, we want to emphasize that there are also many situations where a quantum computer will always be inferior to a classical one, or where the quantum-boost will be minor.[14] Most likely, the supercomputers of the future are a hybrid[15]; a quantum computer

---

[8] TD Ladd et al, 'Quantum Computers' (2010) 464(7285) *Nature* 45; S Lloyd, 'A Potentially Realizable Quantum Computer' (1993) 261(5128) *Science* 1569; Colin D Bruzewicz et al, 'Trapped-Ion Quantum Computing: Progress and Challenges' (2019) 6(2) *Applied Physics Reviews* 021314; Sergei Slussarenko and Geoff J Pryde, 'Photonic Quantum Information Processing: A Concise Review' (2019) 6(4) *Applied Physics Reviews*; Morten Kjaergaard et al, 'Superconducting Qubits: Current State of Play' (2020) 11(1) *Annual Review of Condensed Matter Physics* 369.

[9] Therefore, the field is accumulating more and more attention and resources from industrial players: not only broad-interests corporations such as Google and Microsoft, but also from companies linked to the area of hardware development, like Intel and IBM. Besides commercial actors, the field has also piqued the wide interest of various universities, international organizations and non-governmental bodies.

[10] John Preskill, 'Quantum Computing in the NISQ Era and Beyond' (2018) 2(2) *Quantum* 79; Kishor Bharti et al, 'Noisy Intermediate-Scale Quantum (NISQ) Algorithms' (2022) 94(1) *Reviews of Modern Physics* 015004; Matteo Ippoliti et al, 'Many-Body Physics in the NISQ Era: Quantum Programming a Discrete Time Crystal' (2021) 2(3) *PRX Quantum* 030346.

[11] Earl T Campbell, Barbara M Terhal and Christophe Vuillot, 'Roads towards Fault-Tolerant Universal Quantum Computation' (2017) 549(7671) *Nature* 172.

[12] Iulia Georgescu, '25 Years of Quantum Error Correction' (2020) 2(10) *Nature Reviews Physics* 519; Simon J Devitt, William J Munro and Kae Nemoto, 'Quantum Error Correction for Beginners' (2013) 76(7) *Reports on Progress in Physics* 076001.

[13] See, for instance, Ethan Bernstein and Umesh Vazirani, 'Quantum Complexity Theory' (1997) 26(5) *SIAM Journal on Computing* 1411; Daniel R Simon, 'On the Power of Quantum Computation' (1997) 26(5) *SIAM Journal on Computing* 1474; Peter W Shor, 'Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer' (1997) 26(5) *SIAM Journal on Computing* 1484.

[14] See, for example, Dorit Aharonov et al, 'A Polynomial-Time Classical Algorithm for Noisy Random Circuit Sampling' (2023) *arXiv* 2211.03999.

[15] Alberto Peruzzo et al, 'A Variational Eigenvalue Solver on a Photonic Quantum Processor' (2014) 5(1) *Nature Communications* 4213; Abhinav Kandala et al, 'Hardware-Efficient Variational Quantum Eigensolver for Small Molecules and Quantum Magnets' (2017) 549(7671) *Nature* 242; Nikolaj Moll et al, 'Quantum Optimization Using Variational Algorithms on Near-Term Quantum Devices' (2018) 3(3) *Quantum Science and Technology* 030503; Jarrod R McClean et al, 'The Theory of Variational Hybrid Quantum-Classical Algorithms' (2016) 18(2) *New Journal of Physics* 023023; Ying Li and Simon C Benjamin, 'Efficient Variational Quantum Simulator Incorporating Active Error Minimization' (2017) 7(2) *Physical Review X* 021050; D Zhu et al, 'Training of Quantum Circuits on a Hybrid Quantum

and a classical computer working in symbiosis to tackle hard computational problems more efficiently. In other words, there will be no future extinction of classical computers, at least not due to quantum computing!

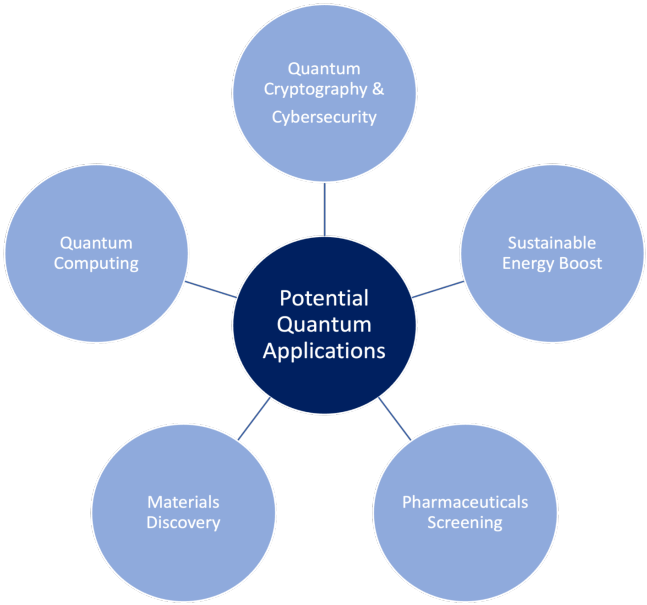


Figure 2. The future is quantum - Potential applications related to quantum technologies that could enter the consumer market within the next decade. Examples include drug and materials discovery, quantum computing and cryptography as well as boosts towards a more sustainable energy consumption and production.

The sought-after quantum advantage[16], sometimes called quantum supremacy[17], stems from the ability of a qubit array to represent and analyze a very large set of information[18]. In fact, a few hundred entangled qubits are sufficient to describe all atoms in the whole Universe, whereas no classical computer would have enough available memory for this task. More precisely, a quantum computer excels in computational tasks which require going through a myriad of possible combinations to find the solution. For instance, the quantum advantage over classical computers can be achieved in solving mathematical optimization issues such as

Computer' (2019) 5(10) *Science Advances* eaaw9918; Ryan R Ferguson et al, 'A Measurement-Based Variational Quantum Eigensolver' (2021) 126(22) *Physical Review Letters* 220501; Bela Bauer et al, 'Hybrid Quantum-Classical Approach to Correlated Materials' (2016) 6(3) *Physical Review X* 031045.

[16] Yulin Wu et al, 'Strong Quantum Computational Advantage Using a Superconducting Quantum Processor' (2021) 127(18) *Physical Review Letters* 180501; Han-Sen Zhong et al, 'Quantum Computational Advantage Using Photons' (2020) 370(6523) *Science* 1460.

[17] Frank Arute et al, 'Quantum Supremacy Using a Programmable Superconducting Processor' (2019) 574(7779) *Nature* 505; Aram W Harrow and Ashley Montanaro, 'Quantum Computational Supremacy' (2017) 549(7671) *Nature* 203; Adam Bouland et al, 'On the Complexity and Verification of Quantum Random Circuit Sampling' (2018) 15(2) *Nature Physics* 159.

[18] See, for example, Lov K Grover, 'Quantum Mechanics Helps in Searching for a Needle in a Haystack' (1997) 79(2) *Physical Review Letters* 325.

the prime number factorization problem[19], which is closely linked to modern encryption methods, and the traveling salesman problem[20], which is in turn, for example, to the optimization of parcel delivery routes. Even though this casts a shadow on the widely employed encryption protocols, quantum can also be an answer[21]: quantum information processing and quantum cryptography are thus promising applications leveraging the laws of quantum mechanics in the future. Furthermore, the quantum enchantment is shown to pay dividends in both machine learning and artificial intelligence that are valuable tools to utilize the ever-increasing mountain of available data.[22] In addition to the computational speed-up, quantum computers are a natural platform to test the fundamental principles of nature[23], as well as to simulate the behavior of (complex) physical systems[24], which could act as a new catalyst to material science and chemistry.

Quantum computers could also assist to map out and benchmark future pharmaceuticals, and thus dramatically accelerate the discovery of new therapeutics or other useful drugs, while screening the highly multidimensional chemical space with classical methods is almost impossible.[25] Indeed, most total synthesis of target drug molecules to date still relies on the chemical intuition and experience of chemists in envisioning potential synthesis routes, which ultimately need to be tested in cost-extensive trial and error

---

[19] Peter W Shor, 'Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer' (1997) 26(5) *SIAM Journal on Computing* 1484.

[20] Charles H Bennett et al, 'Strengths and Weaknesses of Quantum Computing' (1997) 26(5) *SIAM Journal on Computing* 1510; Lov K Grover, 'Quantum Mechanics Helps in Searching for a Needle in a Haystack' (1997) 79(2) *Physical Review Letters* 325.

[21] See, for instance, Dustin Moody et al, *Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process* (US Department of Commerce, NIST, 2022).

[22] Vedran Dunjko and Hans J Briegel, 'Machine Learning & Artificial Intelligence in the Quantum Domain: A Review of Recent Progress' (2018) 81(7) *Reports on Progress in Physics* 074001; Vojtech Havlícek et al, 'Supervised Learning with Quantum-Enhanced Feature Spaces' (2019) 567 *Nature* 209; Yunchao Liu, Srinivasan Arunachalam and Kristan Temme, 'A Rigorous and Robust Quantum Speed-up in Supervised Machine Learning' (2021) 17(9) *Nature Physics* 1013; Vicente Moret-Bonillo, 'Can Artificial Intelligence Benefit from Quantum Computing?' (2014) 3(2) *Progress in Artificial Intelligence* 89; V Saggio et al, 'Experimental Quantum Speed-up in Reinforcement Learning Agents' (2021) 591(7849) *Nature* 229; Andreas Trabesinger, 'Quantum Computing: Towards Reality' (2017) 543(7646) *Nature* S1.

[23] Shruti Dogra, Artem A Melnikov and Gheorghe Sorin Paraoanu, 'Quantum Simulation of Parity–Time Symmetry Breaking with a Superconducting Quantum Processor' (2021) 4(1) *Communications Physics* 26; Simanraj Sadana, Lorenzo Maccone and Urbasi Sinha, 'Testing Quantum Foundations with Quantum Computers' (2022) 4(2) *Physical Review Research* L022001; Scott E Smart, David I Schuster and David A Mazziotti, 'Experimental Data from a Quantum Computer Verifies the Generalized Pauli Exclusion Principle' (2019) 2(1) *Communications Physics* 11.

[24] Ehud Altman et al, 'Quantum Simulators: Architectures and Opportunities' (2019) 2(1) *PRX Quantum* 017003; Bruce M Boghosian and Washington Taylor, 'Simulating Quantum Mechanics on a Quantum Computer' (1998) 120(1-2) *Physica D: Nonlinear Phenomena* 30; Richard P Feynman, 'Simulating Physics with Computers' (1982) 21(6-7) *International Journal of Theoretical Physics* 467; Francesco Tacchino et al, 'Quantum Computers as Universal Quantum Simulators: State-Of-The-Art and Perspectives' (2020) 3(3) *Advanced Quantum Technologies* 1900052; Nathan Wiebe et al, 'Simulating Quantum Dynamics on a Quantum Computer' (2011) 44(44) *Journal of Physics A* 445308; Christof Zalka, 'Simulating Quantum Systems on a Quantum Computer' (1998) 454(1969) *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 313.

[25] See, for instance, Carlos Outeiral et al, 'The Prospects of Quantum Computing in Computational Molecular Biology' (2021) 11(1) *Wiley Interdisciplinary Reviews: Computational Molecular Science* e1481.

processes.[26] In addition, quantum sensing can be utilized, for example, to make structures and functions of individual biomolecules visible under physiological conditions.[27]

Likewise, materials discovery could be dramatically improved using a combination of machine learning approaches based on existing data bases and the highly parallelized nature of quantum simulations to find new material compositions. Examples and pressing challenges to tackle here include building material that could substitute concrete with its extreme carbon and energy footprint, while being more light-weight and versatile to manipulate, high-performance polymers, or highly efficient semiconductors that could substitute silicon in solar cells or electronics components with enhanced energy efficiency.

Aside from directly using quantum systems in quantum simulation and quantum information, there exists also a plethora of applications already in use today, which could see a dramatic "quantum boost" in efficiency in the near future. Examples here include the use of charges or light with a precise control over multiple quantum properties, such as the spin, momentum and polarization[28]. Controlling the spin of electrons in optoelectronic devices could result in a reduction of scattering losses for enhanced solar cell performance, or significantly improve the rate of catalyzed chemical reactions, while in a spinLED (LED = light-emitting diode) the emitted circularly polarized light could double light-outcoupling efficiencies and thus the energy efficiency in most existing LED displays currently featuring anti-glare filters[29]. This would dramatically enhance the battery life or conversely reduce the energy consumption in most digital consumer products. As such, quantum-boosted technology improvements could enable a more sustainable energy future.[30]

Other potential quantum advantages could lie in the energy-consuming nature of computing itself: This occurs currently almost exclusively employing transistors based on electronics, meaning information can be transmitted through electrical current either flowing or not flowing. As such, a bit of information (a "1" or a "0") relies on permanent consumption of electrical energy. In contrast, employing a quantum-mechanical property of an electron called its "spin" would enable spintronics applications which would be by far more energy-efficient than electronics, as preserving a specific state of information would not consume

---

[26] Tameem Albash and Daniel A Lidar, 'Adiabatic Quantum Computation' (2018) 90(1) *Reviews of Modern Physics* 015002; Y Cao, J Romero and A Aspuru-Guzik, 'Potential of Quantum Computing for Drug Discovery' (2018) 62(6) *IBM Journal of Research and Development* 6.

[27] See, for example, Bruno Miranda et al, 'Recent Advances in the Fabrication and Functionalization of Flexible Optical Biosensors: Toward Smart Life-Sciences Applications' (2021) 11(4) *Biosensors* 107; Antoine Reserbat-Plantey et al, 'Quantum Nanophotonics in Two-Dimensional Materials' (2021) 8(1) *ACS Photonics* 85; Tongtong Zhang et al, 'Toward Quantitative Bio-Sensing with Nitrogen–Vacancy Center in Diamond' (2021) 6(6) *ACS Sensors* 2077.

[28] Crassous, Jeanne et al, 'Materials for Chiral Light Control' (2023) 8(6) *Nature Reviews Materials* 365.

[29] Young Hoon Kim et al, 'Chiral-Induced Spin Selectivity Enables a Room-Temperature Spin Light-Emitting Diode' (2021) 371(6534) *Science* 1129.

[30] See, for instance, Jonathan Ruane, Andrew McAfee and William D Oliver, 'Quantum Computing for Business Leaders', *Harvard Business Review* (1 January 2022) <https://hbr.org/2022/01/quantum-computing-for-business-leaders>; Matthias Möller and Cornelis Vuik, 'On the Impact of Quantum Computing Technology on Future Developments in High-Performance Scientific Computing' (2017) 19 *Ethics and Information Technology* 269; Akshay Ajagekar and Fengqi You, 'Quantum Computing and Quantum Artificial Intelligence for Renewable and Sustainable Energy: A Emerging Prospect towards Climate Neutrality' (2022) 165 *Renewable and Sustainable Energy Reviews* 112493; Annarita Giani and Zachary Eldredge, 'Quantum Computing Opportunities in Renewable Energy' (2021) 2(5) *SN Computer Science* 393.

energy and switching it would consume much less energy than switching in electronics[31]. As such, quantum applications would highly enable energy efficient information processing in the future. In particular, quantum technologies could help reaching multiple Sustainable Development Goals set out by the United Nations, including Affordable and Clean Energy (#7), Industry, Innovation and Infrastructure (#9), Sustainable Cities and Communities (#11), and Responsible Consumption and Production (#12)[32].

For the most part, new quantum technologies at the moment are still in the early stage of pioneering and commercialization, but the race for quantum resources has nevertheless clearly begun. There are high expectations and hopes that new emerging quantum technologies, like quantum computers, may assist us to tackle some of today's acute issues, thereby providing a means towards a greener and brighter society. A step into this direction is to remember that there should still ideally be no great knowledge and information asymmetry between all the relevant operators. Furthermore, knowledge on quantum technologies should also be disseminated in understandable terms to the general audience to achieve a wide social comprehension. In particular, the academy and education system has a central role to play in raising public quantum awareness, and to generate "quantum-skilled workforce" which is a necessary prerequisite to sustain and to drive a quantum ecosystem.

In order to learn from best interdisciplinary practices and to create awareness and the most value, it is crucial to bridge more between business, academia and society. We see it is worthy of increasing knowledge and information between different operators, such as the academy, policy makers and industry, regarding interests, incentives, and objectives supporting the creation of standardization and best practices within quantum technologies. Thus, the goal is to thrive the technological development and its social embodiment, and to create a flourishing quantum ecosystem in the future. A pivotal aspect in quantum awareness is to acknowledge that the employment and possession of new technologies create possibilities but also bring responsibilities. As we are now ushering into the next era of quantum, time is ripe to dissect the social-legal-ethical situation of today, and then to act in preparation for the quantum leap ahead.

III Analysis: Quantum Roadmap

As briefly discussed above, the quantum way of thinking has reshaped our worldview about the Universe but has also led to significant practical applications our modern society relies on in the past century. A current trend is to innovate more efficient and greener materials or components for future nanoelectronics by taking a better advantage of quantum resources[33], as it is getting harder to push the boundaries of Moore's law[34], i.e., doubling the transistor density on a microchip every second year. In addition to this evolutionary development, we are now experiencing a new wave of novel quantum technologies that are promising in terms

---

[31] Igor Žutić, Jaroslav Fabian and S Das Sarma, 'Spintronics: Fundamentals and Applications' (2004) 76(2) *Reviews of Modern Physics* 323.

[32] *Transforming Our World: The 2023 Agenda for Sustainable Development,* UN GAOR, 70th sess, Agenda items 15 and 116 UN Doc A/RES/70/1 (21 October 2015).

[33] See, for instance, Philip Ball, 'Materials Innovation from Quantum to Global' (2022) 21(9) *Nature Materials* 962; F Pelayo García de Arquer et al, 'Semiconductor Quantum Dots: Technological Progress and Future Challenges' (2021) 373(6555) *Science* eaaz8541.

[34] G E Moore, 'Cramming More Components onto Integrated Circuits' (1998) 86(1) *Proceedings of the IEEE* 82.

of social impact and commercial applications[35]: quantum sensing[36], imaging[37], metrology[38], communication[39] and computing[40]. Furthermore, different technologies affect each other, especially when combined. For instance, quantum technologies may influence the survival and evolution of other technologies, such as artificial intelligence and big data applications, but this kind of interactions may vice versa increase the power and employment of the quantum technologies.

Although the emerging quantum technologies, such quantum computing, are still at the initial stage of utilization – transiting from the pilot phase to the commercial sphere, they have already begun to influence the structures and functions of society in a spectrum of ways. Sovereign states, institutions, organizations, and corporations should be prepared for the emergence of these new technologies, with the constant goal of improving the current legislative framework and initiating new ones. The upcoming technological shift will take place gradually, thus continuing efforts to stay updated on this development is crucial in order to provide for meaningful legislation initiatives.

The application and possession of new technologies involve harmonizing different rights but also taking account of rising obligations. In anticipation of the social embodiment of quantum technologies, we address this regulatory dilemma within a legal design framework which comes together in our guideline – Quantum Roadmap.[41] It analyzes the emerging legal and ethical responsibilities into five basic principles that are ethics, inclusion, balancing regulatory activities, safeguarding individual rights, and innovating by design (see fig. 3).

---

[35] See, for instance, Antonio Acín et al, 'The Quantum Technologies Roadmap: A European Community View' (2018) 20(8) *New Journal of Physics* 080201.

[36] See, e.g., C L Degen, F Reinhard and P Cappellaro, 'Quantum Sensing' (2017) 89(3) *Reviews of Modern Physics* 035002.

[37] See, e.g., Omar S Magaña-Loaiza and Robert W Boyd, 'Quantum Imaging and Information' (2019) 82(12) *Reports on Progress in Physics* 124401; Paul-Antoine Moreau et al, 'Imaging with Quantum States of Light' (2019) 1(6) *Nature Review Physics* 367.

[38] See, e.g., Vittorio Giovannetti, Seth Lloyd and Lorenzo Maccone, 'Advances in Quantum Metrology' (2011) 5(4) *Nature Photonics* 222; Luca Pezzè et al, 'Quantum Metrology with Nonclassical States of Atomic Ensembles' (2018) 90(3) *Reviews of Modern Physics* 035005; Michael A Taylor and Warwick P Bowen, 'Quantum Metrology and Its Application in Biology' (2016) 615 *Physics Reports* 1; Géza Tóth and Iagoba Apellaniz, 'Quantum Metrology from a Quantum Information Science Perspective' (2014) 47(42) *Journal of Physics A: Mathematical and Theoretical* 424006.

[39] Nicolas Gisin et al, 'Quantum Cryptography' (2002) 74(1) *Reviews of Modern Physics* 145; Nicolas Gisin and Rob Thew, 'Quantum Communication' (2007) 1(3) *Nature Photonics* 165; H J Kimble, 'The Quantum Internet' (2008) 453(7198) *Nature* 1023; Stephanie Wehner, David Elkouss and Ronald Hanson, 'Quantum Internet: A Vision for the Road Ahead' (2018) 362(6412) *Science* eaam9288.

[40] See, e.g., Michael A Nielsen and Isaac L Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2019); T D Ladd et al, 'Quantum Computers' (2010) 464(7285) *Nature* 45.

[41] See also, Katri Nousiainen and Joonas Keski-Rahkonen, *Quantum Computing Era: New Legal Order, Berkeley Global Society: The Tech Book* (Europa Institute at the University of Zurich, forthcoming 2023); Katri Nousiainen and Joonas Keski-Rahkonen, *Navigating in a Post-Quantum Legal Design Landscape, in Legal Design Book* (Cambridge University Press, forthcoming 2023); 'Katri Nousiainen and Joonas Keski-Rahkonen on "Quantum Computing"', *Berkeley Technology Law Journal Podcast* (Seth Bertolucci and Isabel Jones, 23 August 2022) <https://btlj.org/2022/08/berkeley-technology-law-journal-podcast-quantum-computing>.

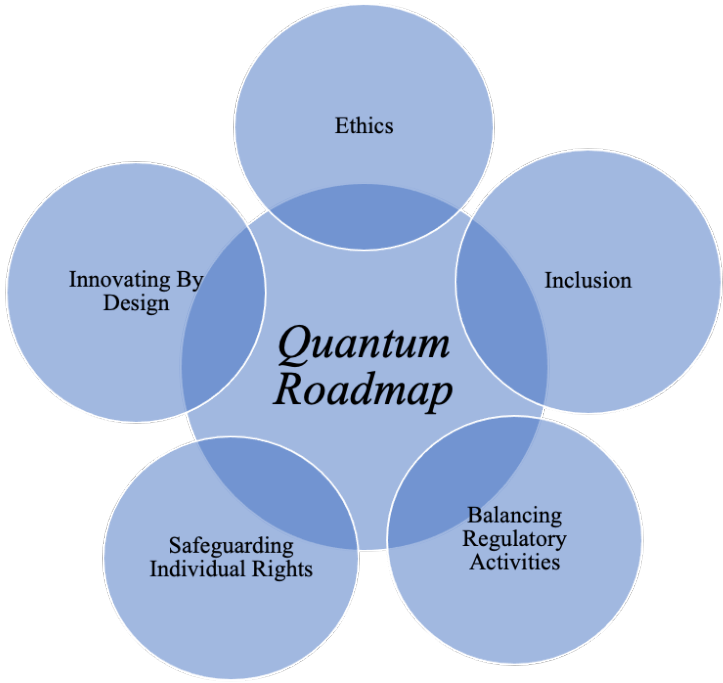


Figure 3. Quantum roadmap suggested by the authors. It has been motivated by the desire to raise quantum awareness and to encourage further debate on the topic. The roadmap charts the social landscape into five interconnecting areas which each is supplemented with a basic guiding principle for the integration of quantum technologies into our daily life.

A *Ethics*

Equal access, public good, and transparency
are the guiding ethical principles.

What ethical risks does quantum technology create, and how can we mitigate those risks? To what extent has quantum technology become a military asset, and what kind of role should international organizations play in governing quantum-based weapon technology?

Law and ethics frequently interrelate, but ethical standards are never a supplementation or replacement to legal measure. In particular, ethics alone is not adequate when regulating high-risk technologies. Nevertheless, ethical aspects *do* provide a valuable direction to construct a legal framework for society. For example, the development or employment of new technologies should not create or aggravate inequalities. It should neither create a different level of standing through its design nor should it leave room for hidden discriminatory practices. Primary calls for the benefit of humankind should be recognized together with commercial incentives.

When it comes to the ethical issues regarding the uprise of new technologies in a society, we do not start ex nihilo. There has been a lot of discussion on ethical rules for different technologies, which has paved the way to modern generic ethical guidelines. Nevertheless, every field has its own special traits; surprisingly the society has relatively recently woken-up to consider the ethical aspects of quantum circumstances.[42] However, for the authors'

[42] See for instance, Mauritz Kop, 'Establishing a Legal-Ethical Framework for Quantum Technology', *Yale Journal of Law & Technology: The Record* (Web Page, 30 March 2021) <https://yjolt.org/blog/establishing-legal-ethical-framework-quantum-technology>.

knowledge, there is no well-established quantum-specific legal-ethics at the current state of affairs. Therefore, we herald for opening a discussion and constructing a legal-ethical framework to cover the emerging quantum technologies. Furthermore, since society is in a constant flux, the ethical norms are thus expected to be dynamic and contextual: the exact quantum-tech regulations will always be a product of their time following the current trend of the applications and implications of the given technology. Consequently, the legal-ethical framework has to be agile and updated with regular intervals.

An exemplary, near-future ethical issue is the dual-usage of quantum technologies.[43] The "dual-usage" term refers to the aspect that technology can be employed in both military and in the commercial sphere. In fact, at the end of 2018, the Commerce Department's Bureau of Industry and Security announced that certain quantum technologies, such as quantum computing, sensing as well as quantum encryption, should be added to a list of blocking U.S exports due to their dual-usage character. [44] Subsequently, the United States has included some quantum technologies on the list of goods whose export is being restricted.[45] There is surely a call for a more extensive and legally binding regulatory framework to address quantum technologies and their export restrictions based on the ethical point of views and the common-good practice.

On the other hand, we must ensure that regulations and export restrictions will not hinder the development of new technologies[46] or cause excessive barriers for their financing, other investments, or slow down scientific dialogues. In general, we see that equal access and transparency lie at the heart of ethics. For this purpose, we suggest following the dogma of legal design approach when approaching the legal-ethical conundrum of quantum technology.

By definition, the legal design is *to apply a user-centered approach to judicial information, services, products, and processes to design them to be more comprehensible*.[47] This approach aims to generate a systemic impact via empowering lay people with law: by supporting equality, creating and building value, increasing tools and products, reducing knowledge and information asymmetry in society, and enabling better access to law and legal information. Within this approach, all actors in the legal field as well as people outside of it are empowered by employing means, such as, design methods, interdisciplinary best practices, and technology.

The legal design operates at least in four prominent ways: empowering, improving, supporting, and demonstrating. It builds on the design thinking process in reaching these

---

[43] See, for instance, Michal Krelina, 'Quantum Technology for Military Applications' (2021) 8 *EPJ Quantum Technology* 24.

[44] See further on *Review of Controls for Certain Emerging Technologies,* 83 FR 58201 (2018). Note: Commerce and Defense both are part of the U.S. Department of the Treasury /CFIUS which makes decisions related to export-control. *See* further, 'CFIUS Overview', US Department of the Treasury (Web Page, 25 August 2023) <https://home.treasury.gov/policy-issues/international/the-committee-on-foreign-investment-in-the-united-states-cfius/cfius-overview>. *See* further on *Export Controls for Quantum Computers*, 15 CFR 774 (2021).

[45] *Implementation of Certain New Controls on Emerging Technologies Agreed at Wassenaar Arrangement 2019 Plenary,* 85 FR 62583 (2020).

[46] On current technological development, *see* for instance, Edward Parker et al, 'An Assessment of the U.S. And Chinese Industrial Bases in Quantum Technology', *Rand Corporation* (Web Page, 2 February 2022) <https://www.rand.org/pubs/research_reports/RRA869-1.html>. Note: Although China is leading on quantum communications, the USA and EU are ahead on quantum sensing.

[47] Katri Nousiainen, 'Legal Design in Commercial Contracting and Business Sustainability New Legal Quality Metrics Standards' (2022) 6(2) *Journal of Strategic Contracting and Negotiation* 137.

functions.[48] The design thinking process determines the challenges and then executes solutions that take end-users' needs into account. These needs are at the essence of solutions and concept development. The design thinking process centralizes understanding, thinking, need-finding, creating, and doing.[49] Here the legal design approach may transform law and legal practice related to quantum technologies becoming more transparent, human-centric, efficient, and comprehensible as well as to foster better quality and the values of non-discrimination. Nonetheless, further research is required to demonstrate and to support the expectations on benefits derived from the legal design approach in the quantum technology field.[50]

B *Inclusiveness*

Democratic involvement and the sharing of knowledge and resources are the guiding inclusive principles.

How will quantum technology affect international trade, trade relations, and trade organizations, and what kind of regulatory challenges does it raise? What kind of positive and negative effects will export and import restrictions have on the quantum-technology industry from a social, economic and innovation perspective? What role should the public and the private sector play, and at which stages? How can or should we fund quantum infrastructure?

We see that the development or employment of new technologies should be inclusive and provide benefits to be utilized for the good of the whole of humankind. The ambition of the inclusiveness is to prevent various risks of increased inequality, e.g., stemming from the monopolization through immaterial property patents, and a quantum division during the commercialization phase, which holds both companies as well as countries. Furthermore, it aims to integrate our democratic values into the social-ontogenesis of new quantum technologies, which, for example, requires educating the general public on quantum-related technologies. For commercial players, a further motivation to this direction is that technology which has gained the trust of the people has a significant market advantage.

From the standpoint of our principle, equal access and openness could, to some extent, help to impede, or at least restrict, the first operators from dominating the field. One concrete future software-level solution could be a cloud-based service enabling researchers and companies to fully tap into the benefits of quantum computers on an equal footing. This cloud-based quantum computing could be either provided by a commercial actor or organized by government authorities. Indeed, the current key actors of the quantum computing market, such as IBM, Google, Microsoft, and Amazon, are on pace to establish

---

[48] Katri Nousiainen, 'What Have I Signed? Do I Really Understand the Contract?', *Contracting Excellence Journal* (Web Page, 12 September 2020) <https://www.worldcc.com/Resources/Blogs-and-Journals/Contracting-Excellence-Journal/View/ArticleID/10812>.

[49] 'A Design Thinking Process', *Stanford Education ME 113* (Web Page, 2 February 2022) <https://web.stanford.edu/class/me113/d_thinking.html>. Note: there exist also various paths as regards the number and content of the design thinking process stages.

[50] See for instance, Katri Nousiainen and Joonas Keski-Rahkonen, *Quantum Computing Era: New Legal Order, Berkeley Global Society: The Tech Book* (Europa Institute at the University of Zurich, forthcoming 2023); Katri Nousiainen and Joonas Keski-Rahkonen, *Navigating in a Post-Quantum Legal Design Landscape, in Legal Design Book* (Cambridge University Press, forthcoming 2023); 'Katri Nousiainen and Joonas Keski-Rahkonen on "Quantum Computing"', *Berkeley Technology Law Journal Podcast* (Seth Bertolucci and Isabel Jones, 23 August 2022) <https://btlj.org/2022/08/berkeley-technology-law-journal-podcast-quantum-computing>.

their quantum clouds, thus allowing the quantum computing experience to a broader audience.[51]

In contrast to quantum-software design, the situation is more challenging on the hardware level, i.e., designing and manufacturing quantum-computer architectures. Whereas we are on the software level, the hardware progress lacks behind, but it is catching up at an accelerating speed. However, unlike quantum-algorithms such as the quantum Fourier-transformation that commonly belong to the public domain, the hardware development is currently strongly led by the private sector so that the corresponding technological breakthroughs fall under immaterial protection rights and corporational trade secrets. At some level, this aspect is problematic for the evolution of quantum computing, e.g., in respect to the openness and further development of the technology. Due to a major innovation or to the early-bird advantage, there may occur a winner-takes-all scenario in the quantum game where one actor begins to dominate the market and the technological development in an unhealthy fashion. In general, it raises a question what role the public and the private sector should play, and at which stages. For instance, there might be a reason for governmental institutes to prohibit or to restrict access to some part of the technology, e.g., because of dual-usage and mitigating security risks.

Some countries have already taken export restriction actions at their national level regarding quantum technologies. These actions have been realized in the form of export controls. It is even expected that we will witness the US and China technology war in the future.[52] However, similarly to the usage of technology, export policy also has a dual character: it can be channeled for good and evil. States should be prudent to goals they aim to achieve through export restrictions. We recognize that the exportation also offers great opportunities to collaborate,[53] learn from each other, and it also provides transparency in the development of new technologies. As the quantum technologies get commercialized, it is a linchpin to find the right balance to safeguard security, and peace, as well as to improve

---

[51] See, for instance, Davide Castelvecchi, 'IBM's Quantum Cloud Computer Goes Commercial' (2017) 543 *Nature* 159; Evan R MacQuarrie et al, 'The Emerging Commercial Landscape of Quantum Computing' (2020) 2(11) *Nature Reviews Physics* 596, 598; Frederic T Chong, Diana Franklin and Margaret Martonosi, 'Programming Languages and Compiler Design for Realistic Quantum Hardware' (2017) 549 *Nature* 180, 187.

[52] Noah Barkin, 'Export Controls and the US-China Tech War Policy Challenges for Europe', *Tendenz Blick* (Web Page, 18 March 2020) <https://tendenzblick.net/wp-content/uploads/2020/08/merics_ChinaMonitor_US-CH-EU-Export-Controls_en_final.pdf>.

[53] See for instance some examples of intergovernmental friendly agreements and research partnerships on quantum technologies; (FIN-USA) 'Joint Statement of the United States and Finland on Cooperation in Quantum Information Science and Technology', *United States Department of State* (Web Page, 6 April 2022) <https://www.state.gov/joint-statement-of-the-united-states-and-finland-on-cooperation-in-quantum-information-science-and-technology>; (AU-USA) 'Cooperation in Quantum Science and Technology', *United States Department of State* (Web Page, 17 November 2021) <https://www.state.gov/cooperation-in-quantum-science-and-technology-aus>; (SWE-USA) Thomas Wong, 'The United States and Sweden Sign Quantum Cooperation Statement', *National Quantum Initiative* (Web Page, 11 April 2022) <https://www.quantum.gov/the-united-states-and-sweden-sign-quantum-cooperation-statement>; (UK -USA) 'The United States and United Kingdom Issue Joint Statement to Enhance Cooperation on Quantum Information Science and Technology', *The White House* (Web Page, 4 November 2021) <https://www.whitehouse.gov/ostp/news-updates/2021/11/04/the-united-states-and-united-kingdom-issue-joint-statement-to-enhance-cooperation-on-quantum-information-science-and-technology>; (UK-SWE) George Freeman, 'New Joint Statement between UK and US to Strengthen Quantum Collaboration', *Department for Business, Energy & Industry Strategy* (Web Page, 4 October 2021) <https://www.gov.uk/government/news/new-joint-statement-between-uk-and-us-to-strengthen-quantum-collaboration>.

technological development. In particular, we call for international influencers such as the United Nations (UN) or World Trade Organization (WTO) to take a bigger role in addressing these issues.

C *Balancing Regulatory Activities*

Innovativeness, common good, effectiveness and being technology-friendly are the guiding regulatory principles.

What type of institutions and governance structures does the emerging quantum technology require? To what extent can we rely on current and emerging regulatory frameworks? What can we learn from the history of technological governance and regulatory restrictions?

The development or employment of new technologies should not be hindered by regulatory measures. In other words, the goal of the regulatory route is to maximize benefits to the whole society and mitigate risks of applied quantum technology. At the same time, the legislative actions should be cohesive and respectful towards the principles of proportionality and subsidiarity while providing a stable and predictive regulatory environment, which is a key element for commercial players. The regulatory actions should be guided by the Aristotelian-like philosophy on the excess and deficiency: balancing legal development, legal rights and obligations, public good, and incentives to innovate as well as to safeguard a fertile soil to develop technologies further.

Although some initial steps have been taken,[54] contemporary legal frameworks are inadequate to cover quantum technologies. We want to emphasize that there is particularly a pressing demand for an international regulatory framework for the employment of quantum technologies in the society-wide global context. For example, WTO could be a prominent and connecting entity between different nations for addressing the commercial usage of future quantum technologies. Overall, it is of the utmost importance to find the regulatory balance.

With quantum technologies maturing rapidly, it will be seen whether the current incentives are enough for different operators to come together to take precautionary actions. Despite being better to be proactive than reactive, a cynical prediction is – by reflecting on the past – that a major incident often needs to occur before appropriate measures are set in place.

D *Safeguarding Individual Rights and Liberties*

Prioritizing individual autonomy, fundamental rights and liberties, equality and fairness are the guiding principles.

How will quantum technology affect digital surveillance, privacy, fairness, trust, access to information, and human rights? What are recommendations for the private sector to collaborate with the government?

The development or employment of new technologies should not interfere with recognized individual rights and liberties, exclude individual access without good cause unreasonably, create or aggravate inequalities between individuals, interfere within individual autonomy, create barriers to access to justice or other recognized democratic fundamental principles. Safeguarding equal standing, non-interference on individual rights,

---

[54] Ibid; See for instance, *National Quantum Initiative Act*, Pub L No 115-368, 132 Stat 5902-5103; *National Quantum Initiative Act*, HR 6227, 115th Congress (2017-2018) <https://www.congress.gov/bill/115th-congress/house-bill/6227/text>.

and safeguarding for taking larger frameworks into account when making justified decisions affecting individual's rights and obligations.

For instance, the future quantum applications and innovations can be expected to comply with the legislation on data protection, governance and privacy. However, it is currently unknown to what extent can we rely on current and emerging regulatory frameworks such as General Data Protection Regulation (GDPR)[55],The California Consumer Privacy Act (CCPA)/ California Privacy Rights Act (CPRA),[56]Proposal for a regulation of the European parliament and of the council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (AI Act), [57]Digital Markets Act (DMA)[58], The Digital Services Act, (DSA),[59] Data Act,[60] Wassenaar Arrangement[61] to mention a few. This privacy issue culminates in the matter of cybersecurity where one must switch eventually to new quantum-proof encryption standards as quantum computers scale up. Like with balancing the regulatory actions, it remains to be seen whether the current incentives are enough for the field itself to take the precautionary step, or if a governmental nudge is required to motivate its reformation.

Currently, we are living in the age of information. In fact, we are almost drowning in the sea of data, but a quantum way of thinking may give us an ability to efficiently process enormous data sets. In the coming decades, synergies of quantum technology and AI may open a new chapter in data science. As regards, for instance, quantum-boosted artificial intelligence can be employed to categorize data, to track patterns, to benefit process development, and to make more accurate forecasts. Moreover, it is speculated that quantum-enhanced AI will play a major role in the rise of autonomous decision making. Nevertheless, quantum data utilization should not violate human rights, including human dignity, agency and oversight with the right to an explanation, and the rights of humans with respect to artificial intelligence. This core principle should be methodically embedded in existing and future regulatory structures. To ensure human-centricity, one can employ the design

---

[55] See *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1.

[56] *Cal Civ Code* § 1798.100-1798.199.100 ('California Consumer Privacy Act of 2018').

[57] European Commission, Directorate-General for Communications Networks, Content and Technology, 'Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts', *Eur-Lex* (Web Page, 21 April 2021) <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206>.

[58] Digital Markets Act (DMA), expected to be adopted in September or October 2022. 'Digital Markets Act (DMA)', *European Commission* (Web Page, 8 July 2022), archived at <https://web.archive.org/web/20220708202056/https://ec.europa.eu/competition-policy/sectors/ict/dma_en>; 'Europe Fit for the Digital Age: Commission Proposes New Rules for Digital Platforms, *European Commission* (Web Page, 15 December 2020) <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347>.

[59] 'The Digital Services Act Package', *European Commission* (Web Page, 25 September 2023) <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>; 'Europe Fit for the Digital Age: Commission Proposes New Rules for Digital Platforms, *European Commission* (Web Page, 15 December 2020) <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347>.

[60] 'Data Act: Commission Proposes Measures for a Fair and Innovative Data Economy', *European Commission* (Web Page, 23 February 2022) <https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113>.

[61] Bureau of Nonproliferation, US Department of State, 'Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies' (Web Page, 22 March 2000), archived at <https://1997-2001.state.gov/global/arms/np/mtcr/000322_wassenaar.html>.

methods mentioned previously above as one possible solution, e.g., for designing and generating data sets according to the principle.

E *Innovating by Design*

A steered innovation design towards human centricity, transparency, openness and sustainability is the guiding principle.

What type of innovation should we want? How can and should governments and public entities shape innovation in quantum technology, and what path dependencies might the corresponding actions and inactions create?

A fundamental question is what type of innovations we want in the future. The governments, public and private entities, like different institutions, and other players, such as the developers and investors, have a possibility – and responsibility – to shape innovation in quantum technology, but at the same time keeping track of the path dependencies that the corresponding actions and inactions create in respect to the other four guidelines above. The development or employment of new technologies should be designed in accordance with equality, transparency, ethics, and human centricity. The design of the technology should help provide an equal access to technology, designing technologies to foster non-discriminatory practices, transparency, and sustainability.

First of all, when we investigate, develop and design quantum technology, academia plays a central role and is a good medium to initiate the quantum debate. Researchers do have the duty to steer research and innovation; various risks, legal gaps, ethical questions, societal implications and other unknown ramifications associated with quantum technologies should be factored in. Prospective practices should be designed and tested. Subsequently, insights should be shared and disseminated openly within and outside of the academic community.

Side by side with academia, a public sector needs to step in. For example, governments and governmental institutions can bring the quantum community together, which instead can forecast future trends of quantum technology evolution for the service of the public. With this information, the public sector can become more aware of risks and engage in potential benefits related to quantum technologies. Moreover, it enables the public sector to set up quantum-targeted strategies and policies to steer the progress into the right direction, to maximize the social benefit of the technology. This also enables governments to found new specialized public institutes to offer legal-ethical guidance on the current possibilities associated with the development and usage of quantum technologies from the public point of view. The public sector should also have healthy dialogue with the private sector to establish a pathway for commercial innovations.

Basic research conducted by academia through public funding is usually a precursor for commercial incentives. For example, quantum computing is rapidly evolving field whose the basic research funding still mostly comes from public resources, where the European Union[62]

---

[62] For further reading, European Commission's *Shaping Europe's Digital Future - Quantum Technologies Flagship*, that intends to place the European Union at front of the second quantum revolution in fostering long term research and innovation, 'Quantum Technologies Flagship', *European Commission* (Web Page, 4 October 2022) <https://digital-strategy.ec.europa.eu/en/policies/quantum-technologies-flagship>; *Council Regulation on Establishing the European High Performance Computing Joint Undertaking and Repealing Regulation EU2018/1488* [2021] OJ L 256/3 ('EuroHPC'). The regulation aims to foster making the EU the leading actor in super computing. 'The European High Performance Computing Joint Undertaking,' *European Commission* (Web Page, 30 June 2023) <https://digital-strategy.ec.europa.eu/en/policies/high-performance-computing-joint-undertaking>.

with 7,2 billion and China with 15 billion have a clear leadership on funding. The USA[63] has announced a planned governmental funding for 1.3 billion dollars. However, the private funding has also significantly increased during the recent years, in 2021 quantum computing start-ups raised 1,7 billion (Fig. 4). It is expected that the private funding will just further increase as the commercial applications gain attraction. Operators such as IBM, Amazon, Alibaba, Microsoft, and Google have already launched their quantum computing services in the commercial sphere.[64]

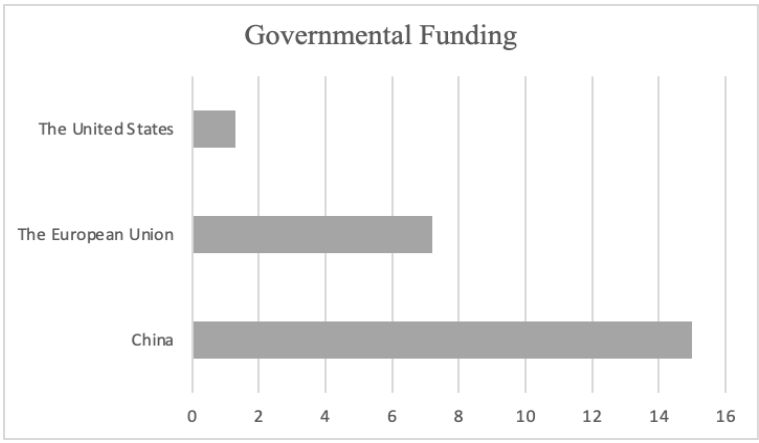


Figure 4. Governmental funding in billion US-dollars in China, EU, and the US.

In consequence, we can ask what type of governmental and public ties the emerging technology requires to achieve a bright quantum future. For example, if the development of

---

[63] See also *National Quantum Initiative Act*, Pub L No 115-368, 132 Stat 5902-5103; *National Quantum Initiative Act*, HR 6227, 115th Congress (2017-2018) <https://www.congress.gov/bill/115th-congress/house-bill/6227/text>; 'National Quantum Initiative', *National Quantum Coordination Office* (Web Page) <https://www.quantum.gov>; 'National Quantum Initiative Supplement to the President's FY 2023 Budget', *National Quantum Coordination Office* (Report, January 2023) <https://www.quantum.gov/wp-content/uploads/2023/01/NQI-Annual-Report-FY2023.pdf>; 'National Quantum Initiative Supplement to the President's FY 2022 Budget', *National Quantum Coordination Office* (Report, December 2021) <https://www.quantum.gov/wp-content/uploads/2021/12/NQI-Annual-Report-FY2022.pdf>. See further discussion around quantum computing and cybersecurity: Christopher Monroe, Michael G Raymer and Jacob Taylor, 'The U.S. National Quantum Initiative: From Act to Action' (2019) 364(6439) *Science* 440; *National Quantum Initiative Act*, Pub L No 115-368, 132 Stat 5902-5103; Arthur Herman, 'At Last America Is Moving on Quantum', *Forbes* (Web Page, 20 August 2018) <https://www.forbes.com/sites/arthurherman/2018/08/20/at-last-america-is-moving-on-quantum/#71eaa5d55327>; Office of Science and Technology, The White House, 'White House Office of Science & Technology Policy and U.S. National Science Foundation Host "Quantum Workforce: Q-12 Actions for Community Growth" Event, Release Quantum Workforce Development Plan' (Web Page, 1 February 2022) <https://www.whitehouse.gov/ostp/news-updates/2022/02/01/white-house-office-of-science-technology-policy-and-u-s-national-science-foundation-host-quantum-workforce-q-12-actions-for-community-growth-event-release-quantum-workforce>.

[64] See, for instance, Edwin Cartlidge, 'Europe's Billion-Euro Quantum Flagship Hands out First Grants', *Science* (Web Page, 29 October 2018) <https://www.science.org/content/article/europe-s-billion-euro-quantum-flagship-hands-out-first-grants>; Garrelt J N Alberts et al, 'Accelerating Quantum Computer Developments' (2021) 8(1) *EPJ Quantum Technology* 18; Jonathan Ruane, Andrew McAfee and William D Oliver, 'Quantum Computing for Business Leaders', *Harvard Business Review* (1 January 2022) <https://hbr.org/2022/01/quantum-computing-for-business-leaders>.

such basic research is carried out or supported by public funding, the fruits could be then shared accordingly. This could mean that the fundamental research results should be announced as open access to be utilized, and the commercialization could take place via licensing to prevent the centralization of the crucial quantum innovations on one player. This kind of proactive involvement of the public sector could be also a precursor to establish industry-wide hardware standards which further stimulate technological evolution on a broader front, e.g., to lure smaller, new players into the quantum play. Therefore, along with the legislational route, an effective regulatory tool is to control the flows of (public) funding in order to design socially and ethically equitable quantum infrastructure without sacrificing the evolution and integration of the technology.

IV ACTION: STEPS TOWARDS A FUTURE QUANTUM SOCIETY

By knowing the "bigger picture", we can take steps to ascertain the functionality and gain of society, while not smothering the evolution and integration of quantum technologies. In practice, success, most likely, requires a strategy plan with concrete steps for how to incorporate these technologies in order to fully capture the commercial opportunities, and to deliver maximum benefit to society at the same time. For example, one of these measures can be to launch a fleet of mission-driven flagships to solve industrial and societal challenges related to the embodiment of quantum technologies. Naturally, our legislational environment needs to be ready for the upcoming quantum change. For instance, the immaterial property right framework could encourage commercialisation as well as accessibility. In the long run, there may be a necessity for an international framework to ensure the coherence and the optimal functionality of the global quantum community in respect to the values presented in Quantum Roadmap. To support the advancement of quantum technologies while mitigating risks for destructive conflict, there need to be frameworks and new institutions that address legal, economic, political, and security issues. This will require institutional innovation, as quantum technologies exist in terms of the "tri-sector" of government, industry, and non-governmental organizations.[65]

Quantum technologies present challenges in terms of both shared development and governance. Companies and nations are cooperating and in competition, or what has been described as "co-opetition," referring to when stakeholders can gain through working together, but are also in fierce competition and must balance the risks of over-exposure, protecting security or trade secrets.[66] In contrast to the Cold War, when Western and Soviet-aligned nations had entirely different economic, political, and security institutions – such as the European Economic Community and NATO, or the Soviet-aligned Council for Mutual Economic Assistance (Comecon) and Warsaw Pact – nations today are closely linked, even when they have extensive divisions. The new competition between the United States, Europe, and China for example is fundamentally different, with integration. Economies have much more integration, there are more exchanges of citizens, and many shared interests.

Co-opetition is possible when both parties can gain without putting critical factors at risk, or the two parties together can gain an advantage over others. The key is in how partnerships are structured. The task is to manage these tensions and be proactive, to ensure benefits and manage risks. and we may need new institutions and processes. Naturally, our legislational environment needs to be ready for the upcoming quantum change. For instance,

---

[65] Nick Lovegrove and Matthew Thomas, 'Why the World Needs Tri-Sector Leaders', *Harvard Business Review* (Web Page, 13 February 2013) <http://hbr.org/2013/02/why-the-world-needs-tri-sector>.

[66] Adam Brandenburger and Barry Nalebuff, 'The Rules of Co-Opetition', *Harvard Business Review* (January-February 2021) <http://hbr.org/2021/01/the-rules-of-co-opetition>.

the immaterial property right framework could encourage commercialisation as well as accessibility. An international framework will be required to ensure the coherence and the optimal functionality of the global quantum community in respect to the values presented in Quantum Roadmap.

This can be done by creating an architecture of the system.[67] Currently legal, economic, political, and security issues are negotiated through international bodies like the United Nations, World Trade Organization, regional bodies like the European Union, academic societies, and Non-Governmental Organizations such as ICANN. These are voluntary, and their formation is led often by a smaller group of powerful actors.

Regulations need to cover data privacy, and access for government authorities such as law enforcement, shared governance of quantum internet, managing norms around cyberattacks, developing solutions to shared challenges such as climate change, and establishing a common language and terms among all three sectors. These have been significant challenges for the United States and China, and quantum computing provides an opportunity to form new institutional arrangements for a fundamentally new technology. These could take the form of new voluntary bodies modeled after the World Trade Organization which have governed challenging economic issues, or the Internet Corporation for Assigned Names and Numbers which has facilitated cross-national governance of the internet.

Quantum-safe data transfers and storage is closely entangled with the security and defense field. Most likely, there will be a demand for a new intergovernmental legal rule framework and surveillance in certain research areas of quantum computing to ensure worldwide security. The possession and employment of new technologies creates opportunities but also responsibilities. A great part of decision power is often vested and employed by public authorities and governments. Furthermore, there will be questions of a deeper functional collaboration and legal rule framework between sovereign states to prevent abusing quantum technologies, such as employing it for purposes to produce war materials.

A solution could be to establish a legal collaborative framework for "Mandatory Reporting and Supervision" to ensure international peace and security. For instance, this could be realized as a form of a Security Council or of a Union of Sovereign States - committing to the same goals on security and sustainability. The operations and accomplishment of the goals should be overseen equally by all the coalition members, and the power should not be centered upon a few selected parties. Ideally, these member states should represent comprehensively sovereign states - not just a few powerful ones - but more equally the sovereign states of the world. The more equal standing of the sovereign states in this possible "Mandatory Reporting and Supervision Body" would allow actions to be taken with less political and historical impact, that is quite the opposite, for instance, to the unfortunate situation with the United Nations Security Council (that is mostly comprised of the War winning countries). The historical burden and political impact have frequently caused the United Nations Security Council to be toothless in taking appropriate measures and actions in reply to threats on international peace and security. Often, a mere "condemn" is insufficient to resolve the incidents occurring at the international arena. The former challenges with international organizations and supervisory bodies should be converted into knowledge for anticipatory and precautionary practices. Therefore, we could learn from the past to ensure the future peace and security. According to economic theory, it should be in

[67] Michael J Mazarr and Tim McDonald, 'Competing for the System: The Essence of Emerging Strategic Rivalries', *Rand Corporation* (Web Page, 10 November 2022) <http://www.rand.org/pubs/perspectives/PEA1404-2.html>.

the interest of operators to collaborate as without collaboration the stakes are extremely high and can lead in the worst scenario, a full-scale mutual destruction. Thus, the game theory advises that it is best to collaborate.

In general, our vision about a bright society of tomorrow is established upon broad scientific capabilities in a coordinated symbiosis with the tech industry to push the cutting-edge quantum technologies forward. At the same, we see an importance to deepen the dialogue within the "Tri-Sectors" of industry, academia and government so that social-level actions are taken in the "right" direction. In the process, a virtuous circle may be set up. Public and private funding stimulates basic research yielding blooming quantum hubs and eventually connecting into a thriving quantum ecosystem. On the other hand, some money will flow back into research to generate more knowledge we can transform into further advantageous innovation, and into more benefits to society.

V CONCLUSION

A new quantum revolution is underway, with innovation enabling the building and controlling of quantum systems in areas of computing, cryptography and cybersecurity, sustainable energy, pharmaceuticals, and materials.

To conceptualize the changes, we propose an A-cubic approach of awareness, analysis, and action to organize legal design. Awareness provides for knowledge of quantum to the general public and to regulators and specialists, bridging between business, academia, and society. To assist in the analysis of quantum capabilities we propose creation of standardization and best practices that cross national and sectoral boundaries. This can be supported through a Quantum Roadmap regulatory framework organizing emerging legal and ethical issues relating to quantum technologies into five categories of ethics, inclusion, regulatory activities, safeguarding individual rights, and innovating by design.

The ethical principles to guide a regulatory framework include equal access, public good, and transparency. The inclusiveness principles include democratic involvement and sharing of knowledge and resources. The principles for balancing regulatory activities are supporting innovativeness and the common good. The principles behind safeguarding individual rights are prioritizing individual autonomy, and fundamental rights such as equality and fairness. Innovating by design means steering innovation design toward human centricity, openness, and sustainability.

Quantum technologies are "tri-sector" and impact industry, academia, and government, mirroring the tensions of both cooperation and competition. New institutions will be needed for this regulation. We propose the establishment of a legal collaborative framework for mandatory reporting and supervision, reflecting a type of Security Council or a Union of Sovereign States, to coordinate across these boundaries and ensure that the development of quantum technologies advances, rather than inhibits or destructs, the betterment of society.

WHOSE DATA IS IT ANYWAY? COPYRIGHT PROTECTION OF DATABASES AND BIG DATA THROUGH THE LOOKING GLASS

TANA PISTORIUS* AND JUAN-JACQUES JORDAAN†

ABSTRACT

This paper addresses a number of copyright issues that arise in relation to the protection of data and databases in the data economy. The paper questions the copyrightability and the ownership of aspects of Big Data. A related issue is the nature and scope of the copyright protection of electronic databases from a common-law perspective. Is the recognition of computer-generated works in South Africa and New Zealand helpful in navigating copyright protection of collections of data in the data economy? The lawful processing of personal information also gives raise to several new copyright issues. The paper addresses the nature of data subject participation rights and consumer data rights and their impact on the copyright protection of databases. For example, where data subject participation rights allow data subjects, under certain circumstances, to reach over the proverbial database-ownership copyright wall and cause the database owner to remove or amend personal information, it may have the effect of amending the original work. It is questioned whether new legislation in the EU, which seeks to protect the public interest while promoting private enterprise, should be adopted in Australasia.

CONTENTS

I INTRODUCTION

Technological advancement has for ever changed the protection and value of data, datasets and databases. What was previously known as the 'database problem' has been compounded

* Professor and Head of Department of Commercial Law, The University of Auckland.

† Legal Practitioner, Durban, Kwa-Zulu Natal. This article is partially based on Juan-Jacques Jordaan *Legal Implications for the Processing of Big Data – a South African Perspective* (LLM dissertation, University of South Africa, 2020). Funding through the SARChI Chair (South African Department of Science and Technology) is acknowledged.

by the EU's *Data Act Proposal* and the *Data Governance Act*. What used to be known as the database problem has now become a Gordian knot.

II WHO OWNS BIG DATA?

A *Big Data as an Asset*

Businesses have come to realise the value of their data and are beginning to treat it as an asset in itself and a means of creating value.[1] As businesses merge and split, the data these businesses possess, is considered to have economic value that form part of the businesses' asset base.[2] Mergers and business takeovers are essential aspects of business and the economy and they affect many stakeholders, including those falling within the scope of various regulators.[3]

Businesses' needs, which are essential to driving and developing the modern data economy, encompass access to and the use of databases and datasets.[4] The protection of the intellectual property that vests in these databases and datasets is of increasing importance. The ownership of data, datasets and databases, however, must be balanced against the data subjects' right to privacy which includes not having their personal information exploited. Ultimately, data-protection principles are conceptually and practically a fine balancing act between individual rights, societal values, national security and economic efficiencies, among other considerations.[5]

The ownership of Big Data[6] as an asset and its valuation is a topic of considerable discussion and bargaining.[7] It has also been argued that the universal principle of property ownership should extend to virtual property, which exists only electronically: so-called virtual property rights.[8] By extension, this would mean that any creation of Big Data made by computers and existing only in electronic format should be recognised as a form of property. The World Economic Forum has even suggested that personal information in itself is a new asset class that will require new interpretations regarding the individuals to whom the information relates.[9] It therefore remains to be seen whether Big Data will be classed as an asset or whether the data subjects' rights will trump such ownership.

This exploitation of data brings up the debate about who owns Big Data, or rather segments of Big Data, and therefore who is permitted to exploit the data contained in them.

---

1 Ira Rubinstein 'Big Data: The End of Privacy or a New Beginning?' [2013] *International Data Privacy Law* 74, 76.

2 Jacques B Stander 'The Modern Asset: Big Data and Information Valuation' (MSc thesis, Stellenbosch University, 2015) 132.

3 MAL Phakeng 'Deal Protection Measures in Takeovers and Mergers: Break Fees' (2018) 39(2) *Obiter* 430, 438.

4 A 'dataset' is defined as a particular collection of data, gathered for a purpose. See 'Introduction to data', *Data.govt.nz* (Web Page, 23 April 2021) <https://www.data.govt.nz/toolkit/intro-to-data/>.

5 O Tene and J Polonetsky J 'Judged by the Tin Man: Individual Rights in the Age of Big Data' (2013) 11 *Journal of Telecommunications and High Technology Law* 351, 363.

6 Big Data is the information asset characterised by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value. ''See Andrea De Mauro, Marco Greco and Michele Grimaldi 'A formal definition of Big Data based on its essential features' (2016) 65 *Library Review* 122, 131.

7 Stander (n 2) 2.

8 Wien Erlank 'Don't touch My Virtual Property: Justifications for the Recognition of Virtual Property' (2016) 133(3) *South African Law Journal* 664, 687.

9 World Economic Forum *Personal Data: The Emergence of a New Asset Class* (Initiative, January 2011).

The debate about copyright ownership of Big Data is far from over, especially considering the rapid speed of technological development and economic progress which is dependent upon this data.

B *Datasets as Training Data*

The recent proliferation of bots that scrape datasets for the development of artificial neural networks (more commonly known as artificial intelligence (AI)) applications highlights the copyright issues related to the use of copyright works as training data.[10] Copyright owners assert the need for proper licensing where AI training data includes copyright works whereas the developers are in favour of a license-less approach.[11]

Besides the question of the copyright ownership of databases, one may question whether it is possible to have ownership and control over both information and data.[12] The recent recognition of consumers' rights in data generated by connected devices, discussed below, has compounded the issues.

The current position regarding raw data is to view a database as a storage space which is capable of ownership; but the data contained in the database is not owned or capable of being possessed without a legal foundation,[13] such as copyright or data-protection laws that explicitly provide for such ownership or possession. It must be borne in mind that a vast part of the Big Data can comprise or at least include personal information[14] and that this may have an effect on the legalities applying to a database.

From a copyright point of view, Big Data ownership is hampered if the database contains data that can identify an individual, as data-protection legislation compounds the meaningful ownership of that data[15] and therefore reduces the commodity value of the Big Data. In reality, the very purpose of using Big Data to identify trends and persons through the analysis of data may create a situation where raw data becomes personally identifiable.[16] This element of identifiability in Big Data would lead to such data falling within the scope and under the

---

[10] See, eg, Jan Bernd Nordemann and Jonathan Pukas 'Copyright exceptions for AI training data—will there be an international level playing field?' (2022) 17(12) *Journal of Intellectual Property Law & Practice* 973, 973 where the authors note: 'Music AI relies on audio recordings protected by the copyrights of the composers and the neighbouring rights of the performing artists and record producers'. Recently two copyright infringement cases were filed in the United States District Courts for the Northern District of California and the District Court for the District of Delaware respectively related to the use of copyrighted images as training data for AI applications: See *Andersen v Stability AI Ltd* (D Cal, Case no 3:23-cv-00201, 13 January 2023) and *Getty Images (US) Inc v Stability AI Inc* (D Del, Case no 1:23-cv-00135, 3 February 2023).

[11] Stuart Dredge 'UK government rethinks plans for AI-training copyright exception', *Musically* (News, 2 February 2023) <https://musically.com/2023/02/02/uk-government-rethinks-plans-for-ai-training-copyright-exception/>.

[12] MN Njotini 'Evaluating the Position of Information or Data in the Law of Property' (2015) 26(1) *Stellenbosch Law Review* 220, 239.

[13] Ibid 224.

[14] See *Protection of Personal Information Act 2013* (South Africa) (*POPIA*) section 8 sv 'Personal Information'.

[15] Xavier Seuba, Christophe Geiger and Julien Pénin (eds) *Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data* (CEIPI/ICTSD Publications Series Issue 5, June 2018) 71.

[16] Paul Ohm 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 *UCLA Law Review* 1701, 1701; Moira Paterson and Maeve McDonagh 'Data protection in an era of big government: the challenges posed by big personal data' (2018) 44 *Monash Law Review* 1, 1; K Krasnow Waterman and Paula J Bruening 'Big Data analytics: risks and responsibilities' (2014) 4 *International Data Privacy Law* 89, 90.

scrutiny of data-protection laws,[17] again making ownership protection difficult or impossible. Organisations that aggregate data generally assume that they hold the rights to the data they possess and, as such, have the right to analyse it and exploit the results or findings of their analyses.[18] Similarly, some organisations that exploit datasets for the purpose of machine learning also operate under those assumptions. Hugh Stephens has pointed to the fact that some AI developers belong to the 'better to ask for forgiveness after rather than permission before' school of thought.[19]

It is not clear-cut whether the exploitation of datasets as data training falls under the data mining exception under EU law.[20] The Digital Single Market Directive defines text and data mining in article 2(2) 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'.

Although one could argue that the use of training data to create AI applications involve the analysis of data in order to recognise patterns which are then applied to parameters of the AI, the fact that a black box approach is used in AI applications obscures the issue.[21] The data mining exceptions and limitations have neither been adopted in the copyright law of Australia nor in New Zealand. This certainly casts a longer shadow on the lawfulness of the unlicensed use of copyright works as training data.

Despite the economic and copyright ownership and infringement debates, the reality is that data already exists as a commodity.[22] Baron notes that in the use of databases, a database owner could exploit the economic potential of the data if the database contents could be effectively structured in a manner in which the owner could claim ownership of both the database[23] and the information contained in it.

III Copyright Protection of Databases

A *Originality*

Databases are a collection of recorded and organised data or information in an electronic or digital format from which data or information may be accessed, reproduced or retracted. Databases are generally protected in terms of copyright law in the same manner as literary works. In South Africa the *Copyright Act* provides the definition of a "literary work", which includes tables and compilations, including tables and compilations of data stored or embodied in a computer or a medium used in conjunction with a computer, but shall not include a computer program.[24]

---

[17] Yvonne Mcdermott 'Conceptualising the right to data protection in an era of Big Data' (2017) 4 *Big Data and Society* 1, 4; See also Seuba, Geiger and Pénin (n 15) 71.

[18] Marcus R Wigan M and Roger Clarke 'Big Data's Big Unintended Consequences' (2013) 46 *Computer* 46, 51.

[19] Hugh Stephens 'Will the Year of the Rabbit be the Year of Contentious Copyright Litigation over AI-Generated Content?' *Hugh Stephens Blog* (Blog, 1 February 2023) <https://hughstephensblog.net/2023/02/01/will-the-year-of-the-rabbit-be-the-year-of-contentious-copyright-litigation-over-ai-generated-content/>.

[20] Nordemann and Pukas, above n 10, 974.

[21] Ibid.

[22] Herbert Zech *Data as a Tradeable Commodity – Implications for Contract Law* (Edward Elgar, 2017) 1; P Baron 'Databases and the commodification of information' (2002) 49(1) *Journal of the Copyright Society of the USA* 132, 144.

[23] Baron, above n 22, 144.

[24] *Copyright Act 1978* (South Africa) section 1(1)(g) of the sv 'literary work'.

The definitions of 'tables' and 'compilations'[25] include Big Data databases. Data and datasets, as would be the case in Big Data databases, should qualify for protection in South Africa as 'literary works', with the requirement of originality being the determining factor for qualification in respect of the works[26] or in this case, the data or the dataset.

Some jurisdictions such as New Zealand and South Africa, still infuse the originality requirement with skill and labour as opposed to creative input – also known as the the 'sweat -of-the-brow' approach to database protection.[27] In New Zealand courts held that skill, judgment or labour[28] or effort, skill and labour[29] were sufficient to impart originality to the works.

In *Haupt t/a Softcopy v Brewers Marketing Intelligence (Pty) Ltd* [30] Streicher JA confirmed that, because the South African *Copyright Act* originated from the English law, creativity is not a requirement for copyright protection. The court then confirmed the test for originality in South African copyright law is: 'Save where specifically provided otherwise, a work is considered to be original if it has not been copied from an existing source and if its production required a substantial (or not trivial) degree of skill, judgement or labour'.[31]

This low threshold for originality suffices and neither a higher standard nor any level of creativity is required. For example, a directory of telefax users,[32] a catalogue and a price list[33] and an electronic database[34] qualified for copyright protection. Dean[35] submits that the skill and labour which go into the compilation must be such that the compilation cannot simply be regarded as a copy of existing subject-matter, but rather as a work that contains features and qualities absent in the material form from which it was initially composed.

The arrangement and selection of data are critical components pertinent to the originality requirement for the protection of databases. However, the selection aspect may be removed where a database is too comprehensive, with the result that very complex databases will enjoy less protection.[36] The digital embodiment of electronic databases meet the other intrinsic requirement for copyright protection, namely the material embodiment requirement. It must be noted, however, that the copyright protection of a database does not extend to the raw data contained in the database.

---

[25] David Rüther 'Government Data and Copyright Protection in South Africa' [2015] *South African Intellectual Property Law Journal* 55, 63.

[26] Ibid 73.

[27] See *Waterlow Publishers Ltd v Rose* The Times 8 Dec 1989; *Waterlow Publishers Ltd v Reed Information Services Ltd* The Times 11 Oct 1990 as quoted by Morton 'Draft EC Directive on the Protection of Electronic Databases: Comfort After Feist' 8 (1992) *Computer Law & Practice* 38, 39; Cornish '1996 European Community Directive on Database Protection' (1996-1997) 21 *Columbia VLA Journal of Law & the Arts* 1, 2.

[28] Labroke (Football) Ltd v William Hill (Football) Ltd [1964] 1 All ER 465 (HL).

[29] Bleiman v News Media (Auckland) Ltd [1994] 2 NZLR 673 (CA).

[30] Haupt t/a Softcopy v Brewers Marketing Intelligence (Pty) Ltd 2006 (4) SA 458 (Supreme Court of Appeal).

[31] Haupt t/a Softcopy v Brewers Marketing Intelligence (Pty) Ltd, above n 30, 473A–B.

[32] Fax Directories (Pty) Ltd v SA Fax Listings CC 1990 (2) SA 164 (Local Division).

[33] Payen Components SA Ltd v Bovic CC and Others 1995 (4) SA 441 (Appellate Division).

[34] See *Haupt t/a Softcopy v Brewers Marketing Intelligence (Pty) Ltd*, above n 30.

[35] OH Dean Handbook of South African Copyright Law (Juta, 1987), 1-8A.

[36] A Roos 'Data Privacy Law' in DP Van der Merwe (ed) *Information and Communications Technology Law* (LexisNexis, 3rd ed, 2022) 363, 361.

B *Authorship*

The question of human authorship is important as Big Data often exist of machine-generated data. Guadamuz notes that creative works qualify for copyright protection if they are original, with most definitions of originality requiring a human author. Most jurisdictions, including Spain and Germany, state that only works created by a human author can be protected by copyright.[37]

There are two ways in which copyright law can deal with works where human interaction is minimal or non-existent. It can either deny copyright protection for works that have been generated by a computer or it can attribute authorship of such works to the creator of the computer-generated work. In the United States, for example, the Copyright Office has declared that it will 'register an original work of authorship, provided that the work was created by a human being.' This stance flows from case law[38] which specifies that copyright law only protects 'the fruits of intellectual labor' that 'are founded in the creative powers of the mind.'

In Europe the Court of Justice of the European Union has also declared on various occasions, particularly in its landmark *Infopaq* decision[39] that copyright only applies to original works, and that originality must reflect the 'author's own intellectual creation.' This is usually understood as meaning that an original work must reflect the author's personality, which clearly means that a human author is necessary for a copyright work to exist.

Similarly, in an Australian case,[40] a court declared that source codes were not original because they were generated by a computer, not written by a human author or by joint authors. Shortly, a work generated with the intervention of a computer could not be protected by copyright because it was not produced by a human.

The second option, that of giving authorship to the programmer, is evident in a few countries such as India, Ireland, New Zealand and the UK. This approach is best encapsulated in UK copyright law, section 9(3) of the Copyright, Designs and Patents Act (CDPA), which states: 'In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.'

Furthermore, section 178 of the *Copyright Designs and Patents Act* defines a computer-generated work as one that 'is generated by computer in circumstances such that there is no human author of the work'. The idea behind such a provision is to create an exception to all human authorship requirements by recognizing the work that goes into creating a program capable of generating works, even if the creative spark is undertaken by the machine.

C *Sui Generis Database Right*

In Europe, the *EU Database Directive*[41] deals with the matter of databases under copyright law and also provides database creators with a *sui generis* right to databases. The European Union adopted a novel approach in the *Database Directive*[42] after nearly eight years of

---

[37] Andres Guadamuz 'Artificial intelligence and copyright' *WIPO Magazine* (Article, October 2017) <https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html>.

[38] *Feist Publications v Rural Telephone Service Company Inc,* 499 US 340 (1991).

[39] *Infopaq International A/S v Danske Dagbaldes Forening* (C-5/08) [2009] ECR I-06569.

[40] *Acohs Pty Ltd v Ucorp Pty Ltd* [2012] FCAFC 16.

[41] *EU Directive 96/9*: Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ 1996 L77/20 ('*Database Directive*')).

[42] See the *Database Directive*.

deliberation. The Directive provides a two-tier form of protection. It strives to create a harmonised level of copyright protection for 'original' databases.[43] A novel 'sui generis' right to protect investments in databases was also introduced.[44] Both rights differ in terms of requirements for protection, duration of rights, scope of protection, the exceptions or limitations that apply and the determination of the right holders (both natural and legal).[45]

The *Database Directive* extends copyright protection to databases that constitute 'the author's own intellectual creation' -- databases which evidence some measure of 'originality' or 'creativity' on the part of the author.[46] Article 5 states that compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such.

When, from a qualitative or quantitative perspective, a substantial investment has been shown to be made in a database, its makers may claim a right to it.[47] European legislation therefore provides some protection for the contents of databases,[48] which would also apply to the protection of Big Data databases.

IV THE DATABASE PROBLEM

Big Data is premised on the collection of as much raw data as possible.[49] In order to create a commodity, private ownership of data must be possible. Technological advances have facilitated the creation of big databases which are becoming more efficient and valuable. These databases often consist of vast collections of personal information. In such cases, their creation has raised questions about data subjects' rights to participate and to oversee or control the manner in which their data is being used in these databases – a phenomenon which has become known as the 'database problem'.[50]

As society and businesses have developed, novel and previously unimaginable threats to privacy have emerged. These include data matching, where different sets of unrelated data are compared using a common denominator to match records; profiling, where historical information or records are used to create a profile about a data subject; data mining, which is the processing of databases for the discovery of knowledge,[51] and web harvesting or the use of bots to scrape data and datasets from websites and other online repositories of data. Given these developments, it may be argued that data subject participation rights is a mechanism to guard against these modern harms.

Data subject participation rights[52] provided for in data protection legislation are akin to giving data subjects a right of control over their personal information. A data subject has a

---

43 See Articles 3-5 of the *Database Directive*.

44 See Articles 7, 10 and 11 of the *Database Directive*.

45 See Articles 6, 8, 9 and 15 of the *Database Directive*.

46 See recital 15 and art 3(1) of the *Database Directive*.

47 Article 7(7) of the *Database Directive*.

48 Julia Johnson 'Database Protection a Reality? How the Professional and Fantasy Sporting World Could Benefit from a sui generis Intellectual Property Right' (2015) 27(2*) Intellectual Property Journal* 237.

49 Joseph Jerome 'Buying and Selling Privacy: Big Data's Different Burdens and Benefits' (2013) 66 *Stanford Law Review Online* 47, 49.

50 Neil M Richards 'Reconciling Data Privacy and the First Amendment' (2005) 52 *UCLA Law Review* 1149, 1150.

51 Anneliese Roos The law of data (privacy) protection: a comparative and theoretical study (LLD Thesis, University of South Africa, 2003), 5.

52 Sections 23–25 of *POPIA*; see also articles 12-23 of the *EU Regulation* 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC ('*GDPR*').

prima facie right to exercise such rights over a database controlled and owned by a responsible party, simply because the data subject's personal information is contained in that database. In effect, they are exercising a right over someone else's property.

By way of example, where the data subject requests a database owner (i.e. a responsible party) to remove or modify their personal information contained in a database,[53] this will have the effect of forcing the database owner to amend its intellectual property at the instruction of the data subject. This would then lead to the question if one may argue that the data subject granted an implied licence to the database owner to include the data subject's personal information in the database.

Consequently, a database may soon only be considered a collection of licensed information that can be protected through contractual agreements with data subjects. These data subject participation rights[54] allow the data subject, in certain circumstances, to reach over the proverbial database-ownership copyright wall and cause the database owner to remove or amend their personal information.[55]

In the case of a large data processor, for example one of the tech giants, the negative possibilities are considerable, especially if all users simultaneously request that their personal information be removed. Before data protection legislation was enacted, it would have been challenging, if not near-impossible, for data subjects to unilaterally change the content of a third-party database owner's database.

A dramatic shift has taken place in the EU regarding the ownership and use of Big Data. In February 2022, the *Data Act*[56] was proposed to promote the sharing of data, particularly data generated by the use of connected objects and the Internet of Things, between companies (B2B) and consumers (B2C).[57] The *Data Act Proposal* defines *who can use what data*, and under *what conditions*. The *Data Act Proposal* also enhances consumer protection by allowing users to control their data generated by digital technologies and to transfer it to third parties. This has a positive impact on competition in the digital markets as it curbs the data power of entrenched tech giant companies.[58] The *Data Act Proposal* creates legal certainty, for both consumers and businesses, around access to data generated by products and services.[59]

The *Data Governance Act*[60] sets out a framework for data intermediation service providers (DISPs). Data intermediation services are defined in the Act as services which aim to establish commercial relationships between data subjects and data holders on the one hand

---

[53] See articles 16-17, article 21 of the *GDPR*; see also section 24 of *POPIA*.

[54] Relevant data subject participation rights contained in the *GDPR* include the right to rectification (section 16), the right to erasure (section 17), the right to restrict processing (section 18) and the right to data portability (section 20); See also sections 23–25 of *POPIA*.

[55] For example through the right to erasure ('right to be forgotten') in terms of article 17 of the *GDPR*.

[56] *Proposal for a Regulation on harmonised rules on fair access to and use of data* (*Data Act*) [2022] COM 68 final.

[57] 'European strategy for data: the CNIL and its counterparts comment on the Data Governance Act and the Data Act' *National Commission on Informatics and* Liberty (Web Page, 13 July 2022) <https://www.cnil.fr/en/european-strategy-data-cnil-and-its-counterparts-comment-data-governance-act-and-data-act>.

[58] Pascal D König 'Fortress Europe 4.0? An analysis of EU data governance through the lens of the resource regime concept' (2022) 8(4) *European Policy Analysis* 484.

[59] Ibid 487.

[60] *Regulation (EU) 2022/868 on European data governance and amending Regulation 2018/1724 (Data Governance Act)* [2022] OJ L 152/1.

and data users on the other, for the purposes of data sharing through technical, legal, or other means, such as infrastructure, platforms, or databases.[61]

The *Data Act Proposal* and the *Data Governance Act* are crucial pillars of the European Strategy for Data.[62] The *Data Governance Act* entered into force on 23 June 2022 and, following a 15-month grace period, will be applicable from September 2023. The *Data Act Proposal* was adopted by the European Parliament on 14 March 2023, and it must still be approved by the European Council. These new instruments aim to create market fairness in the allocation of the value created by data. DISPs manage commercial relationships between data subjects and data holders on the one hand and data users on the other, for the purposes of data sharing.

As noted above, the *Data Governance Act* and the *Data Act Proposal* form part of a European strategy for data. The *Data Act Proposal* covers both personal and non-personal data (as opposed to the GDPR which relates to personal data only) so both apply to mixed data. The *Data Act Proposal* strengthens the *GDPR* and existing rights and obligations under the *GDPR* remains unaffected. It has been noted that the *Data Act Proposal* also strengthens data portability and gives users the right to access and port both personal and non-personal data.[63] The *Data Governance Act* and the *Data Act Proposal* complement each other as the former sets out rules for data intermediaries and data altruism and the *Data Act Proposal* clarifies the rules related to the creation of value from data.[64]

The commodification of information has been firmly entrenched in the European Union. Provision is made for consumers' 'personal data spaces'. Furthermore, provision has been made in the *Data Governance Act* for the creation of 'e-wallets' to secure consumers' personal data. This will also have a ripple effect and will compound the database problem.

V CONCLUDING REMARKS

In summary, the question of data ownership has become more pertinent than ever before. This is due, firstly, to the fact that much of the data being stored in huge databases are created and owned by a few dominant tech giants. Secondly, the use of training data and the proliferation of data scraping techniques have become prominent due to the rise in AI applications. The regulation of personal information, consumer data and machine generated data have become important building blocks of the European data economy. The regulatory environment in Europe will enable organisations to move beyond surveillance capitalism to data capitalism. In short, to innovate through data. To the exploit data for profit.

The copyright protection of databases continues to play an important role. The same underlying policy objectives that support the protection of literary works also underlie the protection of databases, a species of literary works, especially in jurisdictions that have a low threshold of originality. However, these assumptions should be questioned as far as the copyright protection of electronic databases is concerned.

Big Data is not a unique form of data, but simply vast amounts of data which is difficult to process using traditional data-processing methods. Big data is immensely valuable to the

---

[61] Article 10.

[62] Francesco Vogelezang 'The Data Act: five implications for the Datasphere' *Datasphere Initiative* (Article, 22 August 2022) <https://www.thedatasphere.org/news/the-data-act-five-implications-for-the-datasphere/>.

[63] Blanca Escribano and Sofía Fontanals 'The Data Act: new EU rules for data sharing' EY Spain (Article, 8 November 2022) <https://www.ey.com/en_es/law/the-data-act-new-eu-rules-for-data-sharing>.

[64] Ibid.

data economy. With data processors using novel and unique technologies to process this Big Data and performing a vast amount of work on the data, the question of the ownership, from an intellectual property perspective, arises, but it is one that remains unclear and unsettled. This is something that the regulators would have to consider in dealing with updated proprietary forms of protection for data and databases in the future.

As the law currently stands, Big Data as such is not regulated in most jurisdictions, but certain aspects – such as personal information contained in Big Data – are regulated to ensure the personal information of identifiable individuals are processed lawfully. The regulation of data is therefore based on the content of a database and not on a database as a system. This position has been described as a flawed basis for the protection of personally identifiable information.[65] This is especially true if we consider the vast economic significance of consumer data.

The *EU Database Directive* deals with databases both as a copyright work and as a *sui generis* right, whereas New Zealand and South Africa view databases as literary works to be protected under traditional copyright law. The *EU Database Directive* excludes, with good reason, single source databases. A series of South African cases[66] illustrate that the copyright protection of single source databases may be used in a defensive and anti-competitive manner to lock out competitors. It submitted that the underlying policy objective of copyright law is not being served where the extensive protection afforded to the owners of electronic databases are used in a manner that deter competition. This is especially true where a database is the sole source of information. Copyright protection may act as a barrier to competitors especially where such a database has become an industry norm or where it is the single source of information; and it functions akin to an essential facility. In these circumstances it can be argued copyright abuse rears its ugly head.

It may be argued that a *sui generis* right of ownership exists in respect of any personal information belonging to the data subject. The data subject may license such personal information to the responsible party, which licence is subject to revocation at any time if no other legitimate reason for retention or use exists. The definition of property is changing, but the question remains if the definition of property could be expanded to include a data subject's digital identity and thus provide for the concomitant protection of such property. With the increase in cybercrime, specifically that focused on identity theft, it is these authors' view that if identity is something that can be stolen, and therefore something that must be capable of ownership, it should be granted a *sui generis* right to its protection.

In respect of licensing considerations, many of the global social media networks license the use of their databases to third parties for a fee, but without consideration being given to the rights of the data subjects that provided the data. We question whether data subjects' rights to their own data are being overridden by unilateral terms of use, and what that means for the ownership debate. Almost two decades ago writers have suggested that new legislation which protect the public interest while promoting private enterprise should be adopted; secondly that databases should be removed from the ambit of copyright law.[67] It was argued

---

[65] Paul Ohm 'The Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation' (2010) 57 *UCLA Law Review* 1701, 1777.

[66] See *Board of Healthcare Funders v Discovery Life; Discovery Health (Pty) Ltd and the Council for Medical Scheme* (Unreported decision case number 35769/2011 Local Division decision dated 2 May 2012); *Transunion Auto Information Solutions (Pty) Limited V Autobid (Pty) Limited*, (Unreported decision case number 6494/2011 Provincial Division decision dated 14 March 2012).

[67] Jacqueline Lipton 'Balancing Private Rights and Public Policies: Reconceptualizing Property in Databases' (2003) 18(3) *Berkeley Technology Law Journal* 773, 830.

that such legislative approach would allow for the commercialisation of rights in information property and simultaneously require government oversight due to the dangers inherent in allowing vast collections of data to be commercially exploited.[68]

As society races into the future creating and collecting ever more data, the need for data-protection laws has become abundantly clear. However, while adequately protecting the interests of data subjects, these laws must not stand in the way of technological progress.[69] Whether this is done in the form of providing a *sui generis* right of ownership to databases of personal information or identity or through some other manner of protecting databases remains to be seen. What is certain is that existing laws do not apply to Big Data with ease.

Specific laws could be created to deal with the ownership of Big Data. A case in point is the holistic approach in the EU through the *Database Directive*, the *Data Act Proposal* and the *Data Governance Act*. Collectively, these instruments provide innovative solutions that form a legislative framework for the promotion of the growth of the EU data economy. We are of the opinion that the debate over the ownership of data has not been adequately explained or justified beyond Europe. The issues concerning the ownership of data requires much more exploration and analysis in Australasia.[70]

In essence, the world is a borderless place where the free flow of data between countries is commonplace and the quest for data is the ultimate goal. What is needed appears to be a global shared framework based on principles or conditions for the protection of data. In 2008 one author naively noted the following regarding the protection of information with reference to the database right:[71]

> Policy considerations underlying the regulation of access to information and access to knowledge should be heeded. It can never been seriously proposed that information itself should be protected (except by the law regarding trade secrets)
>
> …
>
> There is a long-standing principle that copyright should not be extended to cover basic information or "raw" data. However, as evidenced by the ECJ's differentiation between the "creation" of data and its *obtaining* demonstrate, the "sui generis" right comes precariously close to protecting basic information.[72]

---

[68] Ibid 831.

[69] Asunción Esteve 'The Business of Personal Data: Google, Facebook, and Privacy Issues in the EU and the USA' (2017) 7(1) *International Data Privacy Law* 36, 47.

[70] Václav Janeček 'Ownership of Personal Data in the Internet of Things' (2018) 34(5) *Computer Law and Security Review* 1039, 1052.

[71] Tana Pistorius 'The IP protection of electronic databases: copyright or copywrong?' in H Venter, M Eloff, J Eloff and L Labuschagne (eds) 'Proceedings of the ISSA 2008 Innovative Minds Conference' (ISSA Pretoria, 2008).

[72] Pistorius, above n 71, 12; See also Stone and Kernick 'Protecting Databases: Copyright? We don't Need No Stinkin' Copyright' (1999) 16 *The Computer* Lawyer 17; Fieldhouse and Bolton 'Copyright? Wrong! – Copyright protection of computer programs as literary works' (2003) *Copyright World* 22, 25; H Sun 'Copyright law under siege: An inquiry into the legitimacy of copyright protection in the context of the global divide' (2005) 36 *International Review of Industrial Property and Copyright Law* 192; Tana Pistorius 'Copyright in the Information Age: The Catch-22 of Digital Technology' (2006) 2 *Critical Arts* 47, 54; Michael J Bastian 'Protection of 'Noncreative' Databases: Harmonization of United States, foreign and international law' (1999) 22 *Boston College Environmental Affairs Law Review* 425, 426; Lionel M Lavenue 'Database rights and technical data rights: the expansion of intellectual property for the protection of databases' (1997) 38 *Santa Clara Law Review* 1.

With the benefit of hindsight it is clear that their assumptions about the value of data were naive. It is clear that technological advancement has for ever changed the landscape of database protection and our preconceived notions about data, datasets, and data ownership.[73]

[73] Christopher Kuner and others 'The (Data Privacy) Law hasn't even checked in when Technology takes off' (2014) 4 *International Data Privacy Law* 175, 176.

THE ACCC'S PROPOSED DIGITAL PLATFORM OMBUDS SCHEME: DOES IT GO FAR ENOUGH?

KAREN LEE* AND DEREK WILDING†

ABSTRACT

A proposal by the Australian Competition and Consumer Commission for the establishment of a new Digital Platform Ombuds Scheme is being considered by the Australian government. Drawing on our research into options for digital platform complaint handling, and a round table consultation we held with industry, government and consumers at the end of 2022, we support the proposal and also suggest that the existing Telecommunications Industry Ombudsman scheme could be adapted for this purpose. Using a typology for digital platform complaints that we developed as part of our research, we observe that the proposed ombuds scheme would cover only 'transactional' type disputes between end-users and platforms, such as unmet contractual obligations. Recognising the likely expansion of complaints between end-users, and the fluidity of complaint types, we argue for a more comprehensive approach that would address a broader range of complaints, coupled with the development of internal dispute resolution standards.

CONTENTS

In November 2022, the Australian Competition and Consumer Commission (ACCC) recommended the creation of a new independent external ombuds scheme to help address the market power imbalance that exists between consumers and digital platforms.[1] Government is now deciding whether a new scheme is warranted or if an existing body such as the Telecommunications Industry Ombudsman (TIO) should undertake 'any or all functions proposed for the new body'.[2]

Drawing on our research, report and round table consultation exploring options for an external dispute resolution scheme for digital platforms,[3] we argue that, while an expanded

* Senior Lecturer, Faculty of Law, University of Technology Sydney.
† Professor, Faculty of Law, University of Technology Sydney and Co-Director, UTS Centre for Media Transition.

1 ACCC, *Digital Platform Services Inquiry: Interim Report No 5 – Regulatory Reform* (September 2022) 16.
2 Treasury, *Digital Platforms: Government Consultation on ACCC's Regulatory Reform Recommendations, Consultation Paper* (December 2022) 9.
3 See Holly Raiche, Derek Wilding, Karen Lee, and Anita Stuhmcke, *Digital Platform Complaint Handling: Options for an External Dispute Resolution Scheme* (UTS Centre for Media Transition, 2022). See also UTS Centre for Media Transition, *Submission to ACCC Discussion Paper for Interim Report No. 5: Updating competition and consumer law for digital platform services*, April 2022 and *Submission to The Treasury,*

TIO is preferable, the adoption of either option would be a significant, positive step forward for consumers. However, with its narrow focus on what can be characterised as 'transactional' complaints that users make against platforms, the proposed scheme leaves consumers and citizens without an external avenue to resolve complaints against platforms that are more 'social' in nature, as well as complaints that users make against each other (rather than against the platform itself). Attention needs be directed to the former in the medium term (if not sooner) and to the latter in the medium to longer term.

I BACKGROUND

In its 2019 Digital Platforms Inquiry (DPI) Final Report, which included a series of recommendations designed to address the market power of some digital platforms, the ACCC made two suggestions for improving the way platforms handle complaints from customers.[4] Recommendation 22 proposed that the Australian Communications and Media Authority (ACMA) develop standards that would apply to internal dispute resolution ('IDR'), while Recommendation 23 proposed the establishment of an ombuds scheme to deal with escalated complaints under external dispute resolution ('EDR'). The ACCC suggested the TIO be considered for the role, or if that were not feasible, then a standalone ombuds be established. The ACCC also said any ombuds scheme should be expected to adjudicate complaints relating to scam content, business users' complaints involving advertising campaigns and suspended business accounts, but added the ACMA should 'consult broadly to identify all areas which could benefit from the recommended ombudsman scheme.'[5]

The then Coalition government did not endorse the ACCC's suggestion concerning the ACMA, but it gave in-principle support to Recommendations 22 and 23, proposing that the ACCC work with the major platforms on a pilot EDR scheme that could inform any decision to establish a Digital Platform Ombudsman.[6] By 2021, the Department of Infrastructure, Transport, Regional Development and Communications commissioned some background research into digital platform complaints.[7] However, a pilot EDR scheme with the major platforms was never developed.

In 2022, the ACCC repeated its call for the adoption of IDR standards and an external ombuds scheme in a Discussion Paper[8] and Interim Report No 5[9] relating to its Digital Platform Services Inquiry. However, in Interim Report No 5, published in November, the ACCC changed its thinking about the body that should perform the role of an ombuds. Whereas previously it had suggested that the TIO could be considered, the ACCC concluded that 'an industry-specific ombuds would be preferable given that an existing body may not have the capability and capacity to undertake this role'.[10] In addition, although it suggested further consideration should be given to the types of disputes the ombuds should handle, the

---

Digital Platforms: Government Consultation on ACCC's Regulatory Reform Recommendations, 22 February 2023. The round table was held at UTS on 7 December 2022.

4 ACCC, *Digital Platforms Inquiry: Final Report* (June 2019) 37-38.

5 Ibid 509.

6 Australian Government, *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry* (Policy Statement, December 2019) 13. On 13 October 2022, ACMA announced it would undertake research into digital platform reports and complaints over the next 12 months. ACMA, 'ACMA Releases 2022-23 Research Program' (Media Release, 13 October 2022).

7 This research was commissioned and completed in 2021 but was not published.

8 ACCC, *Digital Platform Services Inquiry Discussion Paper for Interim Report No 5: Updating Competition and Consumer Law for Digital Platform Services* (February 2022) 51.

9 *Interim Report No 5* (n 1) 74-107.

10 Ibid 103.

ACCC indicated the scheme would primarily be expected to resolve user complaints concerning the conduct of the digital platforms involving customers' unmet contractual expectations (eg decisions to suspend services or terminate their accounts) and/or infringement of an amended Australian Consumer Law (ACL).[11]

Existing industry ombuds and regulators do not currently have jurisdiction to resolve any of the complaints mentioned by the ACCC in Interim Report No 5. Consequently, if digital platform customers are to have recourse to EDR for these complaints, a new independent body will need to be created or the functions of an existing body expanded. A key question is which option is best?

II A New Independent External Ombuds Scheme?

A new body would avoid some of the complexities, identified in our July 2022 report,[12] involving modifications to the constitutions and funding arrangements of existing schemes.[13] In the case of the TIO, modifications would be needed to accommodate an expansion of its memberships to digital platforms – arguably enabling faster scheme set up and quicker redress for consumers. In contrast, a new scheme would involve the creation of a brand new body and require this body to set up the kind of administrative frameworks under which existing schemes operate. It would also require resources to educate consumers about the new scheme to ensure the scheme had some consumer brand recognition – recognition that existing schemes already enjoy. Further, a new ombuds scheme might not be able to fully leverage the deep knowledge and expertise of dispute resolution gained by existing schemes. Existing regulators and schemes would need to be willing to work with the new stand-alone scheme. And given the possibility of some complaints made to the new body raising additional matters that fall within the jurisdiction of multiple regulators and schemes, the development of new memoranda of understanding and other administrative arrangements between all potentially relevant parties would be needed.

III An Expanded TIO?

In our July 2022 report, we also considered whether nine existing bodies and regulators could potentially handle the types of complaints identified by the ACCC in its DPI Final Report.[14] However, our review of them suggested that the TIO was the only existing body that could perform the functions the ACCC is now suggesting should be performed by a new ombuds scheme. This was not because of any perceived failings on the part of the other bodies; rather, the other bodies all have functions that render them ill-equipped to take on user-to-platform transactional complaints, or the addition of those complaints would be likely to impede their existing work. Moreover, the TIO itself acknowledged in its 2019 submission on the government's response to the ACCC's DPI Final Report that complaints like the ones the

---

[11] For information about the ACCC's proposal to amend the ACL, see ibid 64-71.

[12] Raiche, Wilding, Lee and Stuhmcke (n 3) 44-5.

[13] A 'purpose-built' digital platform scheme might also facilitate nimbleness and flexibility (eg, if the government decided in the future to expand the scheme to include user-to-platform social complaints). See further below.

[14] They included the TIO, eSafety, ACMA, the Digital Industry Group Inc (DIGI), the ACCC, the OAIC, the Australian Small Business and Family Enterprise Ombudsman, Ad Standards, the Australian Press Council and other news standards organisations. See Raiche, Wilding, Lee and Stuhmcke (n 3) 24-35, 42-3.

ACCC highlighted in its DPSI Report No 5 are a natural fit for an expanded TIO.[15] This is because the TIO currently administers a resolution scheme based on similar consumer complaints about telecommunications service providers.

An expanded TIO would have the advantages of reducing some of the brand-generation and recognition costs and potential for consumer confusion likely to arise as a result of the creation of yet another external complaints scheme. This is especially the case given the TIO has reported that consumers already contact it seeking resolution of digital platform complaints.[16] Using the TIO would also avoid the need to replicate existing administrative frameworks and procedures and allow it to bring its 30 years of experience and successful track-record with dispute resolution in the telecommunications sector to digital platforms.

However, existing regulators and schemes would need to be willing to continue to work with an expanded TIO, requiring changes to their memoranda of understanding and other administrative arrangements to facilitate the smooth operation of the scheme. In addition, the TIO would need to agree to take on this additional function and be adequately funded by government (at least in the short term) to acquire the necessary capabilities and capacities to perform this role.

IV Expanded TIO is Preferable but Multiple Types of Complaints are Left Without Any Means of External Resolution

We agree with the ACCC that the body entrusted with resolving user complaints involving unmet contractual expectations and potential ACL infringement by the digital platforms must have the capability and capacity to do so. But the ACCC does not make it clear why the TIO may not. The number of complaints involved is likely to be very high, but without further information, and subject to adequate funding of the TIO and swift adoption of any necessary amendments to the *Telecommunications (Consumer Protection and Service Standards) Act 1999* (Cth) and the TIO's constitution, we do not see why the TIO could not perform this additional function should it wish to, and the other regulators and schemes with which it currently works agree. Our preference therefore is for the TIO to assume these new responsibilities. However, as our analysis highlights, adoption of a new digital platform ombuds scheme is an equally acceptable way forward.

A more important concern in our view is the limited remit of the proposed independent external ombuds scheme (whether new or an expanded TIO) – a point we illustrate below by reference to the typology of complaints developed in our July 2022 report, which focused on social media platforms and the leading social media service in Australia – Facebook.[17]

We determined that complaints can be about the conduct of the social media platforms themselves or about the conduct of third-party users of those platforms, including advertisers, sellers and users who post content. Further, they can be distinguished as *social disputes* based on, for example, harmful content one user posts about another, or complaints against a platform for exposure to illegal content, misinformation or other harmful material;

---

[15] Telecommunications Industry Ombudsman, *Submission from the Telecommunications Industry Ombudsman to the Treasury's Consultation on the Final Digital Platforms Inquiry Report* (September 2019) 18. The TIO confirmed its suitability in its submission to Treasury on the ACCC's recommendations. See TIO, *Submission to the Commonwealth Department of the Treasury: Consultation on ACCC's Regulatory Reform Recommendations* (February 2023).

[16] Ibid 6.

[17] This was due to budgetary considerations.

or *transactional disputes* often involving unmet contractual expectations but sometimes involving misuse of user data etc.[18]

Using these distinctions, complaints can be grouped into four categories:

- user-to-platform transactional complaints
- user-to-user transactional complaints
- user-to-platform social complaints
- user-to-user social complaints.

Table 1 below shows how we classified the various complaints involving social media platforms using this typology.

	Social	Transactional
User-to-platform Complaints	• Illegal content including terrorism, CSAM, instruction in criminal acts • Pornography and other offensive content • Disinformation and misinformation • News content eg, accuracy and fairness • Content moderation disputes • Advertising content eg, community standards, offensive material • Sale of prohibited goods or services • Election advertisements • Propagation via fake accounts and other inauthentic behaviour	• Unfair digital platform business practices • Complaints about digital platform service, charges etc. • Privacy / other personal violations by digital platform • Failure to protect user eg, account hacking • Complaints about service disruption eg, account suspension • Dispute over terms of service / account suspension etc. • Complaints involving digital platform failure to comply with dispute resolution obligations
User-to-user Complaints	• Abuse, harassment and discrimination and other personal harms with an online dimension • Damage to reputation • Identity theft, impersonation • Disclosure of confidential or protected information • Other privacy breaches by third parties	• Scams and fraudulent transactions • Misleading advertising and product claims, unfair terms, product defects, other sales disputes • Breach of copyright • Comments in reviews of products and services • Spam and unwelcome notifications or communications

Table 1. Types of complaints made about content and conduct on digital platforms

The complaints listed in the table are not intended to be exhaustive and we were and remain conscious that some topics of complaint (eg, privacy breaches, spam and unwelcome notifications of communications) could be allocated to another category or more than one category. It is also possible that a user-to-user social complaint (eg, one about extremist content) could become a user-to-platform transactional complaint, if, for example, the user's account were suspended or cancelled as a result of the extremist content and the account owner made a complaint. In addition, what may begin as a user-to-user dispute may become a user-to-platform dispute where one or more of the users considers the platform has failed to fulfil its obligations (where they exist) relating to the resolution of user-to-user disputes.

[18] This categorisation draws on the work of Ethan Katsh and Orna Rabinovich-Einy, 'The Challenge of Social and Anti-Social Media' in Ethan Katsh and Orna Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes* (Oxford University Press, 2017) 109-130, 113. Katsh and Rabinovich-Einy limited their category of 'social disputes' to user-to-user disputes; we have expanded this category to include complaints that users might have against the platforms themselves.

Despite these qualifications, the typology allows us to highlight that the examples of complaints cited by the ACCC in DPSI Report No 5 are mostly 'transactional disputes'. They largely encompass user-to-platform complaints, but also include some user-to-user complaints (eg, reporting and removal of scams and fake reviews). Two important consequences flow from this.

1. There will be no external means of resolving many types of user-to-platform social complaints that arise on social media platforms if the ACCC's proposed scheme were adopted: existing regulators and industry schemes do not have jurisdiction over these types of complaints. Examples of such complaints include the failures of social media platforms to discharge their obligations in relation to: disinformation and misinformation (apart from the narrow category of complaints under the Australian Code of Practice on Disinformation and Misinformation[19] that amount to failure to implement systems and processes); news content and breaches of community standards in advertising content (where the complaint is about how the platform itself treats that content); election advertisements (except for the narrow category of actions covered by some electoral laws); censorship; disclosure of confidential or protected information; and damage to reputation (apart from the narrow class of actions against platforms that might succeed, at great expense, via the law of defamation).
2. The proposed ombuds would have no jurisdiction to resolve user-to-user complaints (social and transactional).

To overcome these weaknesses, attention should be directed in the medium term (if not sooner) to how an independent external ombuds (new or an expanded TIO) could be modified to accommodate user-to-platform social complaints disputes. This is especially important given the possibility of user-to-platform transactional complaints becoming user-to-platform social complaints (and vice versa), and user-to-user complaints becoming user-to-platform complaints (and vice versa).

Consideration should also be given to how internal dispute resolution standards could be used to encourage platforms to provide effective means of resolving disputes between users (eg, online dispute resolution) over matters that arise as a result of the use of the platform, apart from the schemes administered by Ad Standards and the Australian Press Council which provide a forum for the resolution of complaints about the content of advertising and news. Social disputes are likely to increase; there is a strong public policy argument for encouraging social media providers to fund easily accessible and no-cost dispute mechanisms; and there is an additional community benefit in helping to address defamation claims in a forum that helps claimants – and courts – avoid the costs of defamation litigation.

[19] Digital Industry Group Inc, *Australian Code of Practice on Disinformation and Misinformation* (22 December 2022). In June 2023 the government released an exposure draft of a Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023 which, if implemented, would give the ACMA powers to register and enforce codes of practice as well as to create its own standards. One of the examples of matters that may be dealt with by codes and standards is 'policies and procedures for receiving and handling reports and complaints from end-users' (see cl 33(3)(i)).

A STUDY ON EXPLAINABLE AI IN HEALTHCARE: A BRIEF REPORT

RITA MATULIONYTE*

ABSTRACT

Despite its exponential growth, artificial intelligence (AI) in healthcare faces various challenges. One of the problems is a lack of transparency and explainability around healthcare AI. This arguably leads to insufficient trust in AI technologies, quality, and accountability and liability issues. In our pilot study we examined whether, why, and to what extent AI explainability is needed with relation to AI-enabled medical devices and their outputs. Relying on a critical analysis of interdisciplinary literature on this topic and a pilot empirical study, we conclude that the role of technical explainability in the medical AI context is a limited one. Technical explainability is capable to addresses only a limited range of challenges associated with AI and is likely to reach fewer goals than sometimes expected. The study shows that, instead of technical explainability of medical AI devices, most stakeholders need more transparency around its development and quality assurance process.

CONTENTS

I BACKGROUND

AI technologies, such as machine learning (ML), are gaining importance in healthcare. AI-enabled medical applications have been developed that promise to: improve diagnosis; assist in the treatment and prediction of diseases; and improve clinical workflow. AI-enabled medical devices are expected to comply with a number of ethical principles and policy recommendations, such as benevolence, privacy and protection of data, safety, fairness, accountability and responsibility, avoidance of bias, governance, and others. A sought-after principle is that of transparency and/or explainability, which is found in most ethical AI guidelines.[1] Generally speaking it mandates that certain information about AI in healthcare should be made available and that outcomes of AI tools should be explainable and interpretable.

In response to this, computer scientists have been working to develop AI explainability techniques, with some of them focusing specifically on explainable AI (XAI) in the healthcare sector. In order to ensure explainability of complex and thus intrinsically inexplainable algorithms (such as those based on deep learning and artificial neural networks) and their

* Senior Lecturer, Macquarie Law School, Macquarie University.

[1] Eg Australia's Artificial Intelligence Ethics Framework (2022), <https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles> ('There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI and can find out when an AI system is engaging with them.).

outcomes, numerous so-called post-hoc XAI approaches and techniques have been developed and discussed in the literature.[2] At the same time, the literature has shown signs of increasing disagreement as to whether explainability should be a required feature of AI devices, including those intended for the healthcare sector. While some commentators argue that the black box nature of AI-enabled medical devices has led to a lack of trust and quality, and, consequently, a slow adoption of these technologies in practice, others are increasingly suggesting that AI explainability is not a necessary or adequate measure in ensuring the quality of AI or, indeed, the trust in AI.[3]

The aim of this study was to examine whether, why, and to what extent AI explainability should be demanded with relation to AI-enabled medical devices and their outputs. To achieve this aim, we posed the following questions: First, what exactly an AI explainability principle means and how it could be delineated from other terms, such as transparency and interpretability; second, what goals AI explainability can be expected to achieve and which stakeholders will likely benefit from AI explainability; and finally, is AI explainability capable of achieving the identified goals or does it merely create a 'false hope', as suggested by some commentators?

In the study, we adopted a dual methodology. First, we have reviewed, synthesised, and critically analysed medical and computer science literature exploring the question of explainability of AI-enabled medical devices. Secondly, we adopted the Focus Group method to supplement our analysis with first-hand empirical data. We organized two pilot focus group discussions (5-6 participants each) to collect views from clinicians, AI developers and policy makers on the need of explainability for AI-enabled medical devices.

This study was conducted by an interdisciplinary team: Dr Rita Matulionyte (Macquarie Law School, Macquarie University), Paul Nolan (Macquarie Law School, Macquarie University), Prof Farah Magrabi (Australian Institute for Health Innovation) and Prof Amin Beheshti (School of Computing, Macquarie University).

II FINDINGS

Since 'AI explainability' does not have an agreed definition and various meanings of it are provided in different contexts, we first developed the definition to be used in this study. We noted that both in literature and in policy documents, AI explainability is sometimes used as a synonym to AI transparency, while in other instances it is delineated from the latter. In our study we distinguish between AI explainability and AI transparency principles. We refer to 'AI explainability' in a narrow sense, as an explanation of *how* an AI system generates outputs, which in most cases will require using specific explainable AI (XAI) approaches or techniques. This is similar to 'technical explainability' as defined by the EU Principles on Trustworthy AI.[4] In contrast, we understand 'transparency' as a requirement to provide information *about* the model. It may require disclosing very general information such as 'when AI is being used (in a prediction, recommendation or decision, or that the user is

---

[2] Eg J Amann et al. 'Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective' (2020) 20 *BMC Med Inform Decision Making* 310.

[3] Eg A J London, 'Artificial Intelligence and Black-Box Medical Decisions: Accuracy Versus Explainability', (2019) 49(1) *Hastings Center Report* 15-21.

[4] Eg European Commission, 'Ethics Guidelines for Trustworthy AI' (2019), <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

interacting directly with an AI-powered agent, such as a chatbot)'[5] or more specific information about the AI use, its technical configuration, limitations, etc.

After clarifying the concept of explainability, we identified the main reasons, as proposed in the literature, why and by whom technical explainability of AI medical devices could be required. We identified 4 main rationales for explainable AI, as discussed in legal, healthcare and computer science literature: trust in technology; patient autonomy and clinician-patient relationship; quality of AI and improved clinical decision making; and accountability and liability.

First, the 'black box' nature of AI arguably fails to elicit trust, both among clinicians and their patients. If clinicians cannot interpret and understand the decision made by AI, such as a diagnosis or a treatment recommendation, or if they cannot understand the criteria taken into account when making the decision, trust and reliance issues will arise.[6] Secondly, a lack of explainability, arguably, is incompatible with patient-centered medicine, as it adversely affects both a patient's ability to make informed decisions and the clinician-patient relationship. [7] Thirdly, the lack of explanation may arguably lead to technical errors or bias in AI that, due to the opaque nature of AI, cannot be readily identified by technical or medical specialists.[8] Such errors or bias, if AI is applied to numerous cases, could lead to harm to multiple patients. Finally, many experts cite explainable AI as the answer to ensuring professional accountability and determining legal liability for wrong decisions generated by AI.[9] Arguably, the opaque nature of AI arguably leads to problems in defining moral accountability and legal liability as it makes it unclear as to who would be held accountable for harm caused by a black box algorithm – the clinician, the AI developer, both, or none of them. Explainable AI would arguably help more clearly and appropriately allocate accountability for incorrect AI decisions.

As a next step, we critically analysed these rationales for explainable AI in healthcare and, through focus group discussions, examined whether stakeholders (clinicians, patients, policy makers) agree with these propositions. We made four main conclusions.

First, AI explainability is not the only (or the best) way to ensure trust in AI among clinicians. We argue that a causal explanation is not always necessary in clinical decision-making as clinicians have traditionally used or relied on technologies that they do not fully understand. For instance, physicians and others rely on laboratory test results in their decision making, even if they do not precisely know how the pathology laboratory testing works. Similarly, clinicians routinely prescribe pharmaceutical interventions without knowing their specific mechanisms of action. Further, XAI techniques are still facing a number of technical challenges and are yet to attain sufficient certainty, and therefore the explanations that they produce cannot be themselves trusted. [10] In additional, we suggest that

---

5 OECD, Recommendation of the Council on AI (2022), para 1.3, <https://oecd.ai/en/dashboards/ai-principles/P7>.

6 See eg K Rasheed et al, 'Explainable, Trustworthy, and Ethical Machine Learning for Healthcare: A Survey', (2021) Comput Biol Med. 2.

7 J C Bjerring, J Busch, 'Artificial Intelligence and Patient-Centered Decision-Making', (2021) 34(2) *Philosophy & Technology* 349-371.

8 H Maslen, 'Responsible Use of Machine Learning Classifiers in Clinical Practice', (2019) 27(1) *Journal of Law and Medicine* 37-49.

9 Eg M Sendak et al, 'The human Body is a Black Box: Supporting Clinical Decision-Making with Deep Learning', (Conference paper, *Fairness, Accountability, and Transparency*, 2020, 99-109, 101).

10 J J Wadden, 'Defining the Undefinable: The Black Box Problem in Artificial Healthcare', (2021) *J Med Ethics* 2.

there are more optimal alternatives to ensure trust in AI systems. Empirical research suggests that trust of the AI system in healthcare could be ensured by, e.g., building relationships with stakeholders from the beginning of the project to the final implementation stage; by respecting professional discretion and elevating the expertise of stakeholders rather than replacing them with technology; and by creating ongoing information feedback loops with stakeholders.[11]

Second, we argue that a lack of explainability does not inhibit patient autonomy, nor their relationship with the clinician or trust in the medical system generally. While patients may need certain information about AI technology, like any other technologies applied in the healthcare sector, the information they would require would fall under the 'transparency' concept defined above, rather than a technical explainability concept on which we focused in this study.

Third, we contend that XAI techniques may be helpful in ensuring the quality of AI during the development process, but the utility of XAI techniques in eliminating AI errors in a clinical setting is questionable. We agree that XAI may be useful, or perhaps even necessary, for AI developers in ensuring the quality, accuracy, and absence of bias when developing AI modules. However, it is questionable whether XAI techniques could help clinicians to improve clinical decision making. In most if not all instances, XAI techniques and their outputs cannot be understood and interpreted by those lacking AI expertise, such as clinicians.[12] Also, empirical evidence suggests that additional explainability features do not necessarily improve clinical decision making. [13] In addition, explanations may lead to an over-trust and over-reliance on the technology, thereby introducing a risk of missing obvious mistakes.

Finally, we also question the need of explainability functions to clearly allocate accountability and liability among different stake holders (clinician, AI developer and healthcare provider institution). We submit that it is yet not clear how explainability functions in an AI-enabled medical device will ultimately affect the determination of liability. Explainability of an AI system is something that, from a legal perspective, potentially cuts both ways: it could both decrease the potential for errors, negative patient outcomes and associated liability for clinicians, or it could increase the standard of care demanded of clinicians, leading to the potential to breach their duty.

III Conclusion

These findings suggest that the role of an AI explainability principle in the medical AI context is a limited one. Technical explainability, as we define it here, can address only a limited range of challenges associated with AI and is likely to reach fewer goals than sometimes expected. This should be considered by policy makers when making demands for AI-enabled medical devices to be explainable, and by companies and data scientists when deciding whether to integrate an explainability function in an AI-enabled medical device.

The full reference to the report is: R Matulionyte, P Nolan, F Magrabi, A Beheshti, 'Should AI Medical Devices be Explainable?', 30(2) *International Journal of Law and*

---

[11] Sendak et al (n 9) 100.

[12] See e.g. interpretability analysis by E Zihni et al, 'Opening the Black Box of Artificial Intelligence for Clinical Decision Support: A study Predicting Stroke Outcome', (2020) *15*(4) *Plos one* e0231166.

[13] Eg H J Weerts et al, 'A Human-Grounded Evaluation of SHAP for Alert Processing', (2019) *arXiv preprint arXiv:1907.03324.*

Information Technology, 151-180 (2022). A pre-print copy could be accessed via: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829858>.

PRESS COUNCILS: ADAPTING AN EXISTING SELF-REGULATORY MODEL FOR THE SOCIAL MEDIA AGE

DIANA NESTOROVSKA*

ABSTRACT

The debate on whether social media can and should be regulated has become polarised in the United States. Some view traditional forms of regulatory intervention as a threat to free speech and a bridge too far towards censorship, while others are sceptical of the efficacy of self-regulation. While Australia and New Zealand do not face the same legislative hurdles in regulating social media platforms, both jurisdictions are grappling with how to regulate social media content. Consideration could be given to adapting the self-regulatory model of a press council to a social media context. It is submitted that adapting this model may be a tentative first step towards greater accountability of digital platforms to the public.

CONTENTS

I INTRODUCTION

Digital platforms have garnered a reputation for being incubators of disinformation and misinformation. In the United States, they are protected from their good faith attempts to regulate content by section 230(c) of the *Communications Decency Act 1996*. In essence, the provision precludes the social media and tech giants from being treated as either "publishers" or "speakers" and grants them immunity from civil liability for hosting third-party content.

The conceptual distinction between disinformation and misinformation is one of intent.[1] Colloquially referred to as "fake news", disinformation is content that has the "look and feel" of traditional news but is designed to deceive its audience.[2] On the other hand, misinformation is not necessarily deliberate in misleading an audience, although attempts to

* Practising Australian Lawyer and Public Member, Australian Press Council. This article was originally submitted as a paper during the author's international exchange program at UCLA Anderson's School of Management, Summer Term 2022. The author thanks Professor Steven E Zipperstein for his comments on the original paper. The views expressed in this article are the author's own personal views and do not in any way represent the views of any organisation affiliated with the author.

1 Andrew M Guess and Benjamin A Lyons, 'Misinformation, Disinformation, and Online Propaganda' in Nathaniel Persily and Joshua A Tucker (eds), *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge University Press, 2020) 10-33.

2 Ibid.

harness misinformation in an orchestrated campaign for political purposes can be characterised as a species of disinformation.[3]

Digital platforms have been historically reluctant to remove content that is characterised as "fake news" or misinformation on the basis that such regulation would impede free speech. Facebook (now Meta) has articulated its position in its "commitment to voice" as follows:[4] "*In some cases, we allow content – which would otherwise go against our standards – if it's newsworthy and in the public interest.*"

While the commitment to free speech is a worthy value, particularly in the context of the First Amendment in the United States, the consequences of taking no or delayed action on disinformation can be serious. Indeed, the storming of the U.S. Capitol on 6 January 2021 exposed the dark underbelly of a laissez-faire approach to content regulation. The attempted insurrection was largely fermented through the "*echo chamber*" effects of social media: Mr Trump's calls that the 2020 Presidential election had been rigged stirred his supporters into a mob and eventually culminated into the storming of the Capitol.[5] Even as police were securing the Capitol, Mr Trump continued to post inflammatory statements on social media.[6] Facebook and Twitter responded by removing his posts and indefinitely blocking Mr Trump from using the platforms on the basis that the content promoted violence, which violated their Terms of Service.[7] While the social media giants did take action to take down Mr Trump's posts, it was seen by some as a case of too little too late.[8]

II Community Standards and Self-Regulation

Under increased public pressure over the proliferation of misinformation and "fake news," Facebook and Twitter have promulgated their Community Standards. Both platforms require users, through their Terms of Service, to adhere to these Community Standards, violation of which can lead to the removal of non-complying content and/or suspension of the user's account. Facebook has a "Community Standard on misinformation" which identifies content that incites interference in electoral processes, and content that undermines public health responses, as objectionable.[9] Twitter goes further and has specific policies relating to health

---

[3] Ibid.

[4] See 'Facebook Community Standards', Meta (Web Page) <https://transparency.fb.com/en-gb/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards>.

[5] Pablo Barberá, 'Social Media, Echo Chambers, and Political Polarization' in Persily and Tucker (n 2) 34-55; Dmitriy Khavin, Haley Willis, Evan Hill, Natalie Reneau, Drew Jordan, Cora Engelbrecht, Christian Triebert, Stella Cooper, Malachy Browne and David Botti, 'Day of Rage: How Trump Supporters Took the U.S. Capitol', *New York Times* (online, 18 July 2023) <https://www.nytimes.com/spotlight/us-capitol-riots-investigations>.

[6] 'Case decision 2021-001-FB-FBR', (Oversight Board) <https://www.oversightboard.com/decision/FB-691QAMHJ/>.

[7] Ibid.

[8] Craig Timberg, Elizabeth Dwoskin and Reed Albergotti, 'Inside Facebook Jan. 6 violence fuelled anger, regret over missed warning signs', *The Washington Post*, (online, 22 October 2021) <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>.

[9] *Terms of Service*, cl 3.2.1, Meta (Web Page) <https://www.facebook.com/terms.php> and 'Community Standard on Misinformation', Meta (Web Page) <https://transparency.fb.com/en-gb/policies/community-standards/misinformation/>; *Terms of Service*, cl 4, Twitter (Web Page) <https://twitter.com/en/tos> and 'Community Standards on Platform Integrity and Authenticity', Twitter (Web Page) <https://help.twitter.com/en/rules-and-policies#twitter-rules>.

and electoral misinformation, namely, the "Covid-19 misleading information policy" and the "Civic integrity misleading information policy".[10]

The question remains whether the mere promulgation of Community Standards is enough, especially given Facebook has explicitly stated in its "commitment to voice" that it will allow content which would otherwise go against its standards to remain on the platform if Facebook considers it in the public interest. First, decisions as to whether content is in the public interest is a traditional editorial function of a publisher. One queries whether section 230 is still appropriate in these circumstances, given that the provision granted immunity to digital platforms on the basis that they were mere facilitators of third-party content. In any event, some commentators have urged social media giants to move quickly and decisively towards self-regulation, largely as a way of keeping unwanted government regulation at bay.[11] Others are more sceptical, dismissing Community Standards as a public relations exercise and questioning whether social media giants are capable of meaningful self-regulation.[12] The premise is that so long as the platforms profit from the exploitation of user content, regardless of whether the content is true or otherwise, then they are maximising shareholder return. On this view, the platforms will not choose to limit their ability to make profit unless there are negative consequences imposed on them from an external body,[13] or the erosion of public trust in them is so great as to translate into a tangible decline in shareholder value. Thus, social media giants are incapable of regulating themselves, and calls for greater government regulation of social media should be no more concerning than regulatory intervention in cases of product liability. In such cases, companies can and have been held to account for faulty products and this has not irrevocably undermined free enterprise in the United States.[14]

Aside from updating its Community Standards, Facebook has taken additional steps towards self-regulation by creating a quasi-regulatory body called the Oversight Board ("*Board*"). The Board is funded by a Trust in which Facebook is the sole contributor, and is overseen by independent Trustees who are appointed by Facebook.[15] According to the Board's Charter, a request for review of a Facebook decision on content can be submitted to the Board by either the original poster of the content, or a person who previously submitted the content to Facebook for review.[16] However, the request can only be submitted in circumstances where they do not agree with Facebook's decision and have exhausted other internal avenues of appeal.[17] Facebook can also submit requests for review to gain the Board's opinion on whether any action it has taken is justified or to request direction on a new emerging area of policy.[18] The right to be heard is not guaranteed: the Board has the discretion to decide whether it will review a request and will be guided by essentially utilitarian

---

[10] Twitter (Web Page) <https://help.twitter.com/en/rules-and-policies/election-integrity-policy> and <https://transparency.twitter.com/en/reports/covid19.html#2021-jul-dec>. Note that the 'Covid-19 misinformation information policy' has not been enforced since 23 November 2022.

[11] See, eg, Michael A Cusumano, Annabelle Gawer and David Yoffie, 'Social Media Companies Should Self-Regulate. Now.', *Harvard Business Review*, (online, 15 January 2021) <https://hbr.org/2021/01/social-media-companies-should-self-regulate-now>.

[12] See, eg, Yolanda Redrup and Andrew Tillett, 'Social Media Platforms Can't Self-Regulate', *Australian Financial Review*, (online, 28 March 2019) <https://www.afr.com/technology/social-media-platforms-can-t-self-regulate-20190327-p517y5>.

[13] Ibid.

[14] Ibid.

[15] *Oversight Board Charter*, art 5.

[16] Ibid, art 2.

[17] Ibid.

[18] Ibid, art 2 and 5.

principles: only cases that have the greatest potential to guide future decisions and policies will be heard, though any case which could result in criminal or regulatory sanctions will be declined.[19]

Facebook's Board was put to the test on the issue of the Capitol insurrection. On 21 January 2021, Facebook announced that it had referred this case to its Board, asking it to consider whether Facebook had correctly decided on 7 January 2021 to prohibit Mr Trump's access to posting content on Facebook and Instagram for an indefinite amount of time.[20] The Board ruled that "*it was not appropriate for Facebook to impose the indeterminate and standardless penalty of indefinite suspension.*"[21] In other words, Facebook went a step too far and needed to take action that was consistent with consequences that are applied to other users of the platform.

It is commendable that Facebook has sought to create a mechanism for users to air their grievances over content decisions. Ultimately, however, the Board does not have the hallmarks of accountability that would give it the legitimacy of a regulatory body. The term "accountability" has been described as "*the process of being called to 'account' to some authority for one's actions.*"[22] Accountability has several features in that it involves giving account to an external third party; one party demanding account and the other responding and accepting sanctions; and it implies that one party has the authority to assert rights over those who are accountable.[23]

Thus, while the Board is technically a separate entity with Trustees that do not answer to Facebook, Facebook finances the Trust and appoints the Trustees. Even in the absence of any actual conflict of interest, this still creates a perceived conflict, which may undermine public trust in the Board. Also problematic is that the Board will only review a handful of decisions, and even then, effectively perform an internal audit function: that is, it will determine whether the decision was consistent with Facebook's content policies and values.[24] In other words, Facebook is accountable to rules that it sets by a body that it effectively funds, and not rules set by a third-party authority. Prior Board decisions have "*precedential value*", and decisions will be published, but this pronouncement is watered down by the qualification that past decisions "*should be viewed as highly persuasive when the facts, applicable policies or other factors are substantially similar.*"[25] Facebook also commits to the independent oversight of the Board in relation to its decisions on content, and states that it will provide reasonable assistance to the Board and implement its recommendations.[26] However, this is only to the extent that the requests are "*technically and operationally feasible*" and not an undue drain on resourcing.[27] Apart from reputational consequences for failing to adhere to Board rulings, there is nevertheless no sanction imposed by the Board.

---

[19] Ibid, art 2.

[20] 'Case decision 2021-001-FB-FBR', (Oversight Board) <https://www.oversightboard.com/decision/FB-691QAMHJ/>.

[21] Ibid.

[22] The author has previously considered the issue of accountability in a regulatory context: see Diana Nestorovska, 'Assessing the Effectiveness of ASIC's Accountability Framework' (2016) 34(3) *Company and Securities Law Journal* 193, 199 citing R Mulgan, 'Accountability: An Ever-Expanding Concept?' (2000) 38 *Public Administration* 555, 555–5.

[23] Ibid.

[24] *Oversight Board Charter*, art 2.

[25] Ibid.

[26] Ibid, art 5.

[27] Ibid.

The Board's purpose is consistent with Facebook's publicly stated policy position on the regulation of content. Indeed, it is difficult to resist a conclusion that the Board was set up for a self-serving purpose. Facebook has articulated a view that procedural regulation – in other words, requiring companies to maintain certain systems and procedures—is the preferred way forward, at least for jurisdictions outside of the United States.[28] Procedural regulation would include requirements that Facebook has already implemented, such as requiring companies to publish their content standards, provide avenues for people to report to the company any content that appears to violate the standards, respond to such user reports with a decision, and provide notice to users when removing their content from the site. Such regulations could require a certain level of performance in these areas to avoid regulatory consequences. However, Facebook does not elaborate on what those regulatory consequences could be.

It is beyond the scope of this paper to critically assess whether self-regulation is appropriate or whether firmer regulatory intervention is required. Self-regulation is not uncommon and in the United States, this accountability mechanism has included companies in the business of video games, an industry that has also seen its fair share of community concern.[29] In the context of section 230 and general wariness of subjecting free speech to potentially overreaching and unconstitutional government control, self-regulation of social media may be a more incremental, palatable, and achievable policy reform in the short to medium term.

III SAFE HARBOUR

By way of comparison, there is no equivalent of "section 230" in Australia: current safe harbour provisions for intermediaries are narrow in scope and do not necessarily extend to social media platforms.[30] In New Zealand safe harbour is available to intermediaries that "facilitate" defamatory content provided they follow the take down procedure set out in the *Harmful Digital Communications Act 2015*. In general, however, legislative interventions targeting social medial content have been reactive and address specific harms (e.g. cyberbullying) rather than targeting disinformation or misinformation. For example, in 2019, the Australian Federal Parliament enacted laws requiring platforms, under pain of criminal sanction, to take down "*abhorrent violent material*" capable of being accessed in Australia.[31] This was in response to the live streaming on Facebook of the Christchurch terrorist act against a mosque. Australia has also moved to give victims of cyberbullying and harassment the right to apply for take down orders of content hosted by online platforms.[32]

---

[28] See 'Charting a way forward on online regulation', Meta (Web Page), <https://about.fb.com/news/2020/02/online-content-regulation/>.

[29] See, eg, Entertainment Software Rating Board (Web Page) <https://www.esrb.org/>.

[30] See, eg, *Copyright Act 1968* (Cth), pt V div 2AA, which provides safe harbour for intermediaries against copyright infringement; Max Mason, 'Google and Facebook excluded from safe harbour copyright reforms', *Australian Financial Review*, (online, 5 December 2017) <https://www.afr.com/companies/media-and-marketing/google-and-facebook-excluded-from-safe-harbour-copyright-reforms-20171205-gzz3fw>.

[31] Monica Biddington, 'Regulation of Australian online content: cybersafety and harm', (Parliamentary Library Briefing Book, July 2019) <https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook46p/Cybersafety>.

[32] See *Online Safety Act 2021* (Cth); eSafety Commissioner (Web Page) <https://www.esafety.gov.au/whats-on/online-safety-act>.

There is some civil jurisprudence in Australia that holds online platforms liable for defamatory content posted by third-party users.[33] Each case, however, turns on its facts, as shown by the recent High Court decision in *Voller*. That case arose out of lower court defamation proceedings launched by Dylan Voller, whose mistreatment at a juvenile detention facility sparked a formal government inquiry.[34] The High Court ruled in a 5-3 decision that the appellant media organisations were strictly liable for comments they invited on their own Facebook posts, but did not rule that Google as the search engine was liable as a publisher.[35] Following the *Voller* case, former senior politician, John Barilaro, won a case for defamation in the Federal Court against Google. Interestingly in that case, Mr Barilaro launched defamation proceedings against Google as the publisher of You Tube videos uploaded by one Mr Shanks, with the Court finding that the tirades against the then politician amounted to online harassment, and Google's failure to remove them a clear violation of its own policies that amounted to publication.[36]

Of the three jurisdictions discussed in this essay, New Zealand has perhaps articulated the most cohesive framework for social media content regulation that would also broadly cover misinformation and fake news. New Zealand's internet watchdog, Netsafe, and industry group NZTech have launched a voluntary *Aotearoa New Zealand Code of Practice for Online Safety and Harms* ("*Code of Practice*") which covers the following thematic areas: child sexual exploitation and abuse; bullying or harassment; hate speech; incitement of violence; violent or graphic content; misinformation and disinformation.[37] Signatories to the Code include Meta, TikTok, Google, Amazon and Twitter, all of which have agreed to comply with the Code by, *inter alia*: publishing annual reports on their systems, policies and procedures for removing harmful content; and adhering to a public complaints process for breaches of the Code for which they may receive sanction.[38] At the time of writing, however, details of the complaints process have yet to be released.

IV PRESS COUNCIL MODEL

It is submitted that policy makers in the United States, Australia and New Zealand do not necessarily have to "reinvent the wheel" when it comes to a model for social media content regulation: the model of a press council is worthy of further consideration. In its traditional form, a press council is a body established by the major actors of the media industry, namely media owners, editors, journalists, and the public.[39] It is responsible for investigating

---

[33] For an interesting discussion on why New Zealand is not likely to follow Australia down this route, see Alex Latu, 'Why NZ is unlikely to follow Australia's lead on social media defamation laws', *The Spinoff*, (online, 17 September 2021) <https://thespinoff.co.nz/media/17-09-2021/why-nz-is-unlikely-to-follow-australias-lead-on-social-media-defamation-law>.

[34] 'Facebook defamation ruling by High Court exposes all page owners to lawsuits, not just the media', *ABC News* (online, 12 September 2021) <https://www.abc.net.au/news/2021-09-12/facebook-defamation-high-court-ruling-exposes-more-than-media/100451198>.

[35] *Fairfax Media Publications v Dylan Voller; Nationwide News Pty Ltd v Dylan Voller; Australian News Channel Pty Ltd v Dylan Voller* [2021] HCA 27, [173] per Kiefel CJ, Keane and Gleeson JJ.

[36] *John Barilaro v Google LLC* [2022] FCA 650, [403].

[37] 'Netsafe, NZTech and global tech companies act to tackle digital harms', NZTech (Web Page) <https://nztech.org.nz/2022/07/25/netsafe-nztech-and-global-tech-companies-act-to-tackle-digital-harms/>.

[38] 'Aotearoa New Zealand Code of Practice for Online Safety and Harms Draft', (Netsafe, 2 December 2021) <https://netsafe.org.nz/aotearoa-new-zealand-code-of-practice-for-online-safety-and-harms-draft/>.

[39] See Lara Fielden, 'Regulating the Press: A Comparative Study of International Press Councils' (Reuters Institute for the Study of Journalism, April 2012) <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-11/Regulating%20the%20Press.pdf>.

potential breaches in the ethical codes of conduct that are adopted by members.[40] Media organisations are not compelled to join a press council, and its efficacy depends on the funding provided by member organisations and the cooperation of all parties. This is important in a democracy on the basis that the media performs a key role in informing the public and holding a government to account. Thus, in principle, a press council should not be funded by government, nor should it have government appointees on the council. In the context of social media, it is submitted that a "social media council" could be established which includes in its constituent membership social media and technology companies, contributors of content, and diverse representatives from the public. This would be an improvement on Facebook's Oversight Board because while the body would be industry-funded, it would also be responsible for creating a set of industry-wide Community Standards that members would need to adhere to, and a process of adjudication for dealing with content complaints that is more at arm's length. Applied to the case of the Capitol insurrection, any adjudication would be based on industry-accepted standards rather than standards set by Facebook itself. This would be a modest step towards greater transparency.

At the time of writing, the United States does not have a press council to which the public can submit complaints about the traditional media.[41] Arguments for and against the idea of a press council have been posited for decades and are not revisited here.[42] The position is different in other countries including Australia and New Zealand, which still have active press councils, namely the Australian Press Council and the New Zealand Media Council. Indeed, in the context of New Zealand's Code of Practice, it is envisaged that a "*multi-stakeholder governance group*" will administer the Code.[43] Arguably, public policy makers could leverage the existing model of New Zealand's Medial Council as a starting point.

Taking the example of the Australian Press Council ("*Press Council*"), its stated purpose is to promote freedom of speech and responsible journalism, set standards, and respond to complaints about material in Australian newspapers, magazines, and online-only publications.[44] Most complaints are resolved without resorting to adjudication and result in a correction or apology from the publication.[45] For those that cannot be resolved at the initial stages of the process, the Press Council may refer the matter to an adjudication panel to hear from the complainant and the publisher. Adjudication panels are drawn from a pool of Press Council members that represent the public and independent journalists (but not member publications), and a group of panel members with community and media backgrounds who are not Council members.[46] While adjudication decisions are published online, it is important to note that adjudications can only be made in relation to members, and as such, there may be media organisations that are not members and therefore beyond the reach of the Press Council. In addition, the Press Council does not have the power of enforcement: generally, it

---

[40] Ibid.

[41] Accountable Journalism (Web Page) <https://accountablejournalism.org/press-councils/USA>.

[42] John A Ritter and Matthew Leibowitz, 'Press Councils: The Answer to Our First Amendment Dilemma' [1974] (5) Duke Law Journal 845; Dr Ralph Lowenstein, 'Press Councils: Idea and Reality' (Freedom of Information Foundation, April 1973) <https://repository.uchastings.edu/cgi/viewcontent.cgi?filename=0&article=1078&context=nnc&type=additional>; Ray Finkelstein and Rodney Tiffen, 'When Does Press Self-Regulation Work?' (2015) 38 *Melbourne University Law Review* 944.

[43] 'Netsafe, NZTech and global tech companies act to tackle digital harms', NZTech (Web Page) https://nztech.org.nz/2022/07/25/netsafe-nztech-and-global-tech-companies-act-to-tackle-digital-harms/>.

[44] Australian Press Council (Web Page) <https://www.presscouncil.org.au/>.

[45] Ibid <https://www.presscouncil.org.au/complaints>.

[46] Ibid <https://www.presscouncil.org.au/about-us/who-we-are/>.

can issue a reprimand or censure, and call for (but not require) the publication to apologise, correct or revise content.[47] Under the New Zealand model, members of the public are required to submit their complaint directly to the member publication for resolution at first instance.[48] If the complainant is not satisfied with the publication's response, then they may complain to the Media Council. Each complaint is assessed against the Statement of Principles and referred to the Chair, a Committee of Council, or the full Media Council depending on the outcome of the assessment. The members, who are drawn from industry and the public, will determine whether the complaint should proceed and whether there has been a breach of ethics.

V Conclusion

No doubt digital platforms may find the concept of an industry body akin to a press council worrying on the basis that such a model stemmed from the traditional print and online media businesses. This may draw an unwanted inference, contrary to section 230, that digital platforms are comparable to publishers and hence liable for third party content they publish. Nonetheless, it is submitted that the model of a social media council is a starting point, noting that the model is not posited to be a panacea for all actual and perceived social media ills. The purpose of an industry-wide body would not be to undermine section 230, but rather, to acknowledge the significant role that social media plays in facilitating free speech, and to progress a constructive debate around social media regulation.

[47] Ibid <https://www.presscouncil.org.au/complaints/handling-of-complaints/>.

[48] New Zealand Media Council (Web Page), <https://www.mediacouncil.org.nz/faqs/>.

DEDICATED CYBER INSURANCE OR BUST – LESSONS FROM INCHCAPE

BENJAMIN DI MARCO* AND ANTHONY KUMAR†

Blended insurance products commonly used to cover emerging specialist risks such as cyber are increasingly likely to leave insured organisations without adequate protection.

The recent decision of *Inchcape Australia Limited v Chubb Insurance Australia Limited* [2022] FCA 883, demonstrates that organisations must purchase specialised cyber insurance policies, to effectively cover the losses and exposures caused by cyberattacks and ransomware threats.

In this case Inchcape sought cover for ransomware losses under an Electronic and Computer Crime (ECC) policy. This blended insurance covers commonly found in crime policies for funds transfer, redirection, and push payment frauds, together with insuring clauses for direct financial loss arising from computer viruses and the modification of electronic data and electronic media.

Unfortunately, the ECC policy did not include any of the core insurance clauses found in market standard cyber liability insurance policies. Because it lacked proper cyber insurance coverage, the court found that Inchcape was unable to use the policy to recover losses sustained after a major ransomware attack.[1]

I UNDERSTANDING THE KEY EXPOSURES

Blended insurance products like Inchcape's ECC policy are common in the market and often attempt to cover both traditional risks as well as emerging specialist risks such as cyber liability.

By their nature, blended products contain narrower insuring clauses, when compared against comprehensive risk specific wordings. Despite this, blended products can be attractive to organisations, and in some cases are more cost-effective than pursuing specialist risk policies such as a cyber insurance policy. They may also be easier to obtain than specialist products, as they often require fewer underwriting questions to be answered.

However, as the Inchcape decision demonstrates, blended products can leave significant uninsured gaps in insurance programs, unless they are carefully analysed against an organisation's key exposures and matched to specific insurance needs.

In this case, Inchcape sought coverage for financial loss sustained following a ransomware attack which included:

1. repairs and/or replacement of hardware, software, and data, including investigation costs
2. hardware and data recovery costs
3. resource and additional staffing costs.[2]

* Cyber and Technology Risk Specialist, WTW.

† Senior Associate, Cyber and Technology, WTW.

1 See *Inchcape Australia Limited v Chubb Insurance Australia Limited* [2022] FCA 883, [11] – [15] (Jagot J) ('*Inchcape*').

2 Ibid 7.

All of these would have been affirmatively covered under a specialist cyber liability insurance policy and are losses commonly sustained after many ransomware attacks. The court found that Inchcape, in relying on an ECC policy, was not covered for the ransomware losses.[3]

Where other organisations have included blended wordings in their insurance programs or have failed to procure a specialist cyber insurance policy, they will be exposed to a similar uninsured fate. Given the Inchcape decision, each organisation should carefully investigate whether they hold sufficient insurance that appropriately addresses their realistic cyber exposures and consider the need for specialist cyber insurance.

II Cyber Risks Impact Insurance Programs as a Whole

The Inchcape decision also highlights the need to examine how an organisation's entire insurance program collectively responds to cyber and technology risks. WTW's recent Global Directors Liability Report identified that cyber-related issues were the top risk concerns for respondents for 2022 with 65% saying the risk of cyberattacks was "very significant" or "extremely significant," and 59% saying they fear a "very significant" or "extremely significant" exposure to cyber extortion attacks.[4]

In handing down the Inchcape decision, Justice Jagot highlighted that cover under the ECC policy was limited to the "direct financial loss" sustained by the company.[5] While this language is common in crime policies, it is immediately problematic for claims caused by a cyber event, because the nature and extent of a cyber loss is determined by:

- the intervening steps taken by the insured after the attack including how they investigate the suspected incident
- any decisions taken to shut down and isolate parts of their IT environment
- the extent of engagement with the malicious actor
- the type of restoration work performed.

These intervening acts reduce the proximate cause, and add an element of indirect or consequential losses, which make wordings like those under the ECC policy significantly less likely to respond.

Some cyber events will however create direct financial losses, particularly where an authentication compromise or fraudulent instructions result in the organisation losing or transferring funds to an incorrect party.

Similar tensions can also arise in situations where an organisation's cyber and technology exposures may also create liabilities under Directors and Officers, Professional Indemnity, and Property Insurance. In some cases, covers may be impacted by specific cyber exclusions which are commonly being added to traditional wordings. This makes obtaining overall coverage for technology risk more difficult, and often requires a higher level of expert advice so that the organisation is properly advised on best-in-class insurance options and potential areas of risk that are not insurable. Without this advice, it is difficult for the organisation to make an informed risk management decision.

---

[3] Ibid 43 – 47.

[4] John Moran and Marc Voses, *Directors' Liability Survey 2022 April 2022* (Report), p 33 <https://www.wtwco.com/-/media/WTW/Insights/2022/04/directors-liability-survey-2022.pdf?modified=20220523170432>.

[5] See *Inchcape* (no 1) [41] (Jagot J).

For large and complex organisations, there is often strong benefit to engage with your broker to examine how all the relevant wording interplays, and the extent to which technology exposures may require multiple insurance policies to address key exposures.

III WHY CYBER RISK AND INSURANCE EXPERTISE IS CRITICAL

The Inchcape judgment demonstrates that cyber risk management and cyber insurance are complex matters. Indeed, in parts of the ruling, it appears that even the Court struggled with the intricacies of cyber incident response management and misunderstood how certain triage tasks were conducted directly following the ransomware event.[6]

Cyber security and risk management is a new industry, and few experts know how to properly bridge these topics. Cyber insurance is particularly complex, as it requires knowledge of both the cyber risk landscape and rapidly evolving insurance products created to meet this risk.

In the current market there is significant variance between the insurance policies offered by different carriers, and the underwriting information required to obtain cyber insurance. Those wordings which seem cheaper or easier to obtain, and often deliberately drafted to reduce the insurer's exposure. These solutions may be suitable in some instances, but if they are not properly scrutinised, can result in the insurance program failing to meet cyber risks.

Had Inchcape obtained support from a dedicated cyber insurance expert it is unlikely an ECC policy would have been recommended as:

1. It contained extremely narrow causation language requiring both that Inchcape suffered direct financial loss, and further that this must directly arise from a small numbered of covered incidents;
2. The wording did not properly call out the incident response costs and steps which Inchcape would need to perform following a major cyber event, or the key ransomware losses that the organisation would suffer;
3. The insuring clauses in the policy did not address the range of malicious acts which are commonly employed by modern ransomware and cyber threat actors;
4. Hurdles in the policy that required damaged or destroyed electronic data, electronic media, or electronic instruction, do not reflect how most modern cyberattacks and cyberextortions are performed. This created further coverage uncertainty; and
5. General conditions in the wording imposed significant limitation of the covers in the policy relevant to cyber events.

[6] Ibid 43.

AUSCL CONTRIBUTIONS TO POLICY ON COMPUTERS AND THE LAW

RAM SUNTHAR*

AUSCL aims to be Australia's leading interdisciplinary think-thank on issues relating to the law, at the intersection of technology and society. It is a registered Australian non-profit charity with a charter to advance education and advocacy at that intersection.

AUSCL provides a lively forum for debate and is committed to providing balanced, informed and transparent advocacy on critical issues and promoting the education of its members and the wider community.

AUSCL members include legal and technology professionals, business leaders, government officials, academics across all disciplines and members of the bar and retired judiciary

AUSCL hosts several training sessions and master classes. Some of the sessions are recorded and hosted on AUSCL Youtube® channel. There are a total of 52 records sessions hosted within AUSCL Youtube Channel.

I SUBMISSIONS

The AUSCL has been actively providing submissions to state and federal governments on various bills and amendments.

A *AI Action Plan*

The AUSCL Policy Lab, together with the Allens Hub, The Law Society of New South Wales Future of Law and Innovation in the Profession (FLIP), and The Disability Innovation Institute at UNSW (DIIU), has made a submission to the Department of Industry, Science, Energy and Resources in response to a call for input into the development of an Artificial Intelligence (AI) Action Plan.

The submission covered the full spectrum of AI capabilities, from defining artificial intelligence to AI maturity. It addressed skills and capabilities that are required for AI enabled future, AI's contribution to legal sector, the expected law reform action for AI action plan and lowering the barriers to entry.

B *Online Safety Bill 2020 Exposure Draft*

This submission proposed establishing guard rails to address online safety, including cyber-bullying, image-based abuse, harassment and abhorrent violent material, through policy development and draft legislation.

It proposed that the guiding principles must include respect for the rule of law and human rights, including the right that individuals not be subjected to arbitrary interference with their privacy. An emphasis on these guiding principles will promote human dignity, natural justice, procedural fairness, transparency and accountability, and predictability and consistency in the application of law. In circumstances where powers exist to limit an

* Architect, CitiPower, Powercor and United Energy.

individuals' rights and freedoms, including freedom of expression, it is necessary to ensure adequate safeguards to ensure those powers, and the limits they impose, are necessary, reasonable, proportionate and justifiable on the grounds of public interest.

C *Digital Advertising Services ("AdTech") Inquiry*

This submission reviewed ACCC's Digital Advertising Services Inquiry: Interim Report ('Interim Report') by focusing on three issues

- The scope and focus of the inquiry;
- The analysis of issues at the intersection of competition and privacy; and
- International (trade and conflicts of laws) considerations that should be included in the analysis.

D *Data Availability and Transparency Bill 2020 and the Data Availability and Transparency (Consequential Amendments) Bill 2020*

This submission acknowledged that the Bill contains some commendable transparency measures such as the public availability of Data Sharing Agreements ('DSA'). The submission covered following propositions.

Objects of Bill. A reference to accountability should be inserted into the Bill's Objects. This would strengthen the functionality of existing safeguards and ensure accountability plays a central interpretive role. In addition, the Objects clause should note that consent remains the primary basis for sharing personal information.

Private Sector Research and Research Ethics. Private sector organisations seeking to use data for research should be required to prove a rigorous ethics process.

New Data Attributes. Interaction with the review of the Privacy Act 1988 definition of "personal information" should be managed.

International Data Sharing. Accreditation of foreign entities should be subject to proof that the relevant foreign country has a comparable privacy law framework.

Transparency. Transparency measures should be put in place with respect to the operation of Clause 15(4). Further, there should be ongoing transparency about flaws in the data protections applied in clause 16(7).

Interaction with Other Legislation. Details of interaction with other legislation should be published, ideally within the Bill. Consistent terminology across legislation should be a long term goal.

Handling of Data After Project Completion. Requirements on termination of a project or suspension of an accredited entity, such as data deletion, should be specified.

Accountability. Transparency and accountability should be enhanced through additional language in privacy policies and a requirement for data scheme entities to raise complaints. Data subjects should also be encouraged to make complaints.

Consent. The threshold for circumstances when it is unreasonable or impracticable to seek consent should be incorporated as part of the ethics function governed by the National Data Advisory Council.

Data Sharing Controls and Environment. There should be minimum standards for security and data protection practices, including training.

Guidelines to Address Data Procurement. The scope of guidelines be amended to cover data procurement and pre-processing as well as the operation of clause 15(4).

E *Inquiry into Draft Critical Infrastructure Asset Definition Rules*

This submission focused on the following topics.

- The scope of what constitutes a critical infrastructure asset should be narrowed to ensure proportionality with respect to the grant of government powers contemplated by the Bill.
- In deciding what is a critical infrastructure asset, it is important to understand network interactions; dependencies are relevant in determining which components are critical.

F *The Privacy and Personal Information Protection Amendment Bill 2021*

AUSCL submission focused on seven issues:

- The scope and focus of the proposed changes
- Definition of 'Eligible Data Breach'
- Resourcing the regulator
- Extensions to assessment periods
- Public Notification
- Exemption from Notification to affected individuals; and
- Reconciliation with other data breach notification obligations.

G *Australian Government Digital Identity Legislation*

AUSCL contributed to digital identity legislation on Position Paper (Phase 2) and Exposure Draft (Phase 3).

- Independent oversight of the system
- Onboarding to participate in the system
- Individual and User expectations
- Privacy and consumer safeguards;
- Security requirements and incident responses;
- Record keeping and data retention; and
- Liability and redress framework.

H *Australian Data Strategy Discussion Paper*

AUSCL contributed to the Data strategy discussion paper covering the following areas:

- Data sharing agreement
- International data sharing
- Data provenance
- Data sovereignty
- Data quality
- Data access
- Data traceability
- Standards for data
- Data integration
- Data lifecycle management

Further submission covered the expected top three outcomes from Australian Data Strategy by 2025.

I *Modernising Document Execution*

AUSCL submission focused on ensuring that regulations that govern the execution documents are light-touch fit for purpose and reflect the way businesses and consumers want to engage and communicate digitally.

J *Privacy Act Review Report*

Privacy review submission addresses a number of proposals.

1. The existence of policy issues remains beyond the scope of the Privacy Act Review;
2. The fact that information cannot be cleanly divided into 'personal' and 'nonpersonal', but that much data that has undergone a de-identification process is reidentifiable with some measure of risk (proposal 2);
3. Concerns about the relationship between the small business exemption and Consumer Data Rights;
4. The importance of removing or narrowing the employee records exemption in order to ensure cyber security requirements apply in this context;
5. A requirement for privacy notices to be accompanied by code that can be automatically processed by computers (so that privacy settings can be used to control what an individual agrees to);
6. The limitations of relying on consent as a solution to consumer issues with privacy (proposal 9);
7. The importance of transparency about research uses (proposal 10.4);
8. How proposal 11 could be strengthened;
9. The ability to use a range of existing security standards to reduce compliance costs in proposal 19.2;
10. Support and further suggestions for strengthening enforcement powers (proposal 24);
11. Suggestions for a direct right of action (proposal 25); and
12. Specific issues around biometric identifiers.

K *Copyright Amendment (Access Reform) Bill 2021*

The AUSCL submission on Copyright Amendment (Access Reform) bill addressed a selection of the issues covered by the proposed legislation in relation to orphan works, namely, Schedule 1 – Limitation on remedies for use of orphan works.

THE AUSCL FUTURE LAW NETWORK

NATALIA CRNOMARKOVIC*

The Future Law Network had a very successful 2022. On 11 March we made a submission to the ALRC in relation to the Legislative Framework for Corporations and Financial Regulation focusing on the need to look beyond XML.

On 15 March 2022, the AUSCL Future Law Network hosted, a Global RaC2.0 Plenary together with the Allens hub for Technology, Law and Innovation. The purpose of the Plenary was to engage with the Global RaC Community to share learning and develop actionable recommendations to assist the community to "take RaC to the next level" RaC 2.0.

The Plenary was attended by well over 100 participants, across 10+ countries with keynotes by globally recognised luminaries - Prof Mireille Hildebrant and Pia Andrews. Together we explored pressing design challenges in theory and practice from running RaC projects (private/public) sectors and technical issues (platforms, standards and interoperability), through to public law considerations, democracy, Rule of Law and the future RaC workforce - distilling key principles and recommendations. Since the Plenary, a RaC Working group continues to provide a valuable forum for cross-disciplinary professionals to share insights on current and future RaC directions.

In May, Natalia Crnomarkovic, Future Law Network leader moderated two sessions at the Legal Innovation and Tech Fest 2022. The first was on Rules as Code with guest speakers Síobháine Slevin, trailblazer in regulation as digital infrastructure and CEO and founder of Realta Logic, Prof Andrew Mowbray and Philip Chung from AustLII.

In second session Natalia unpacked Next Generation Contracting aka Smart Legal Contracts with leaders in the field - Natasha Blycha, Managing Partner with Stirling Rose and Tim Bass, CEO Block8.

The Future Law Network also ran our 6th RaC Masterclass exploring the critical legal coding in the new world of digital legislation featuring Dr Megan Ma from Stanford X and Dr Jason Grant Allen, our AUSCL Tasmanian Chapter Lead.

A huge heartfelt thank you to our Future Law Steering Committee - Mark Staples, Síobháine Slevin, Heidi Richards, Tim De Sousa and Martin Harford for giving so generously of your time, expertise and energy to make 2022 the success it was.

* Future Law Portfolio Leader, Australasian Society for Computers and Law (AUSCL).