



Large Language Models, the ‘Doctrine of Discovery’ and ‘Terra Nullius’ Declared Again?

Reflecting on the phenomenon of Large Language Models (LLMs) and the recent hype around generative AI and ChatGPT, this column interrogates the profound questions around data, rights, and ownership claims, and how any such claims might be viewed critically by computing educators and their students.

Some while ago, I wrote that Google™ was at the vanguard of a “a third wave of colonization for countries such as New Zealand, first by empire and the gun, then by the dollar and economic might, now by the shaping of discourse through distorted delivery of information.” [4] So, do large language models and the possibilities provided by generative AI merely represent another extension of this “third wave” or are they radically different?

Shoshana Zuboff [11,12] has written about the strategies adopted by ‘big tech’ companies in pre-emptively appropriating rights to newly conceived forms of data, expanding to the new world of ‘big data’ and the economic sea change known as “surveillance capitalism.” She argues that “‘Big data’ are constituted by capturing small data from individuals’ computer-mediated actions and utterances in their pursuit of effective life.” [11] She talks of big tech companies accumulating “not only surveillance assets and capital, but also **rights** ... accomplished through a form of unilateral declaration that most closely resembles the social relations of a pre-modern absolutist authority.” [11] Zuboff refers to a calculated secrecy, “concealing a new political equation in which Google’s concentrations of computational power **brush aside users’ decision rights**

This phenomenon of generative AI and how pre-emptive control might operate takes us to questioning how we typically lay claims to property and land?

as easily as King Kong might shoo away an ant, all accomplished offstage where no one can see?” [12]

Then returning to my initial theme of ‘colonization’ in this column, Zuboff observed that “These euphemisms operate exactly in the same way as those found on the earliest maps of the North American continent, in which whole regions were labelled with terms such as “heathens,” “infidels,” “idolaters,” “primitives,” “vassals,” and “rebels.” On the strength of those euphemisms, native peoples—their places and claims—were deleted from the invaders’ moral and legal equations, legitimating the acts of taking and breaking that paved the way for church and monarchy” [11]. While New Zealand may frame itself as a “small, advanced economy,” let’s be quite clear—we, along with the citizens of many other countries, are the newly colonized!

So how does this process of what we might term ‘neo-digi-colonization’ work? And how might AI and the controllers of LLMs go about usurping our ‘places and claims’? One well known strategy is that of “Move fast and Break things” [10], where speed and greed rather than sense predominate, as Liesenfeld and colleagues have noted, “In high-stakes AI research,

data work is often seen as low-level grunt work ... and incentive structures generally encourage a ‘move fast and break things’ mentality over careful scientific work.” [6]

The work of Liesenfeld and colleagues [6] in response to the rise of ChatGPT and LLMs, argue for greater “openness in the fast-moving field of instruction-tuned large language models. We have found projects at varying stages of implementation, documentation, and useability. Most of them offer access to source code and some aspects of pre-training data, **sometimes in legally ambiguous ways.**” [6] They observed in relation to ChatGPT and openness: “There are many shades of openness...yet all of the projects surveyed here are significantly more open than ChatGPT. ChatGPT was announced in a company blog post and rolled out to the public with an interface designed to capture as much free human labour as possible, but without any technical documentation.” [6]

‘Free labour’ has of course been a long-standing marker of colonization!

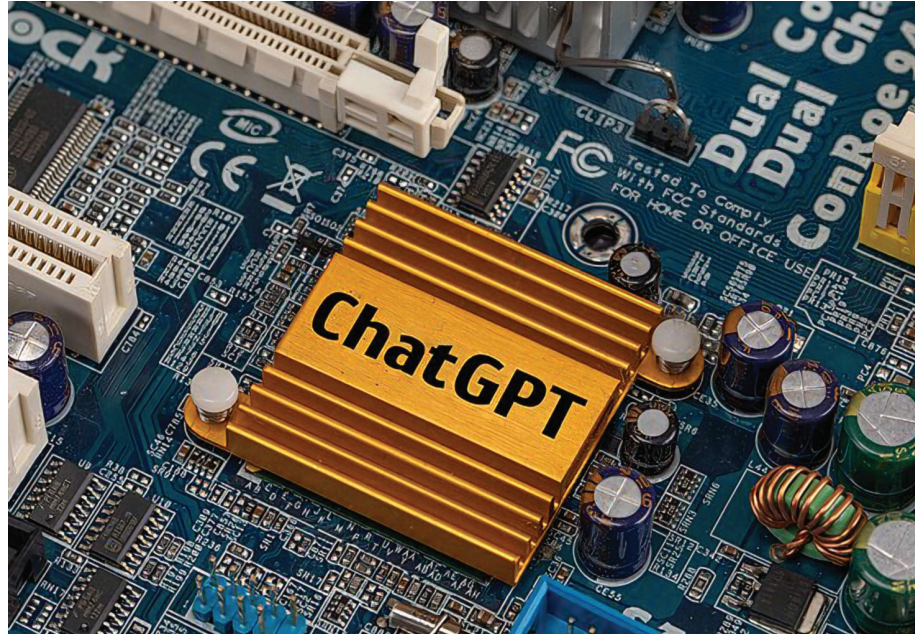
So, this phenomenon of generative AI and how this form of pre-emptive control might operate, takes us to questioning how do we typically lay claims to property and land? The argument developed here takes us away for a period from the

questions of AI but does build the basis for how property ownership claims are built. Simpson, from an Australian legal context has observed that “many of the methods of acquisition derive from Roman law concepts of property ownership. The five classic modes of acquisition of territory at international law were thought to be occupation (or settlement), conquest, cession, accretion, and prescription.” [9]

Further defining the first three modes (as most relevant to this argument), Simpson [9] classifies them in Figure 1.

Historically, European colonization had sanction from the Catholic Church, through what was known as the ‘doctrine of discovery,’ which provided legal justification for the occupation of ‘newly discovered’ territories by the colonizing authorities who laid claim to them. NPR have reported in March 2023 [3] that:

“The doctrine was laid out in a series of papal “bulls,” or decrees; the first one was issued in 1452. They authorized colonial powers such as Spain and Portugal to seize lands and subjugate people in Africa and the “New World,” as long as people on the lands were not Christians.



Nearly 500 years after papal decrees were used to rationalize Europe’s colonial conquests, the Vatican repudiated those decrees on Thursday, saying the “Doctrine of Discovery” that was used to justify snuffing out Indigenous people’s culture and livelihoods is not part of the Catholic faith. [3]

So, while the legitimacy of the doctrine has now belatedly been repudiated and “the mindset of cultural or racial superiority which allowed for that objectification or subjection of people has been renounced,” [3] much of the legal basis for land ownership in settler societies still relies upon it. As one example, Bess has reported that:

The United States Supreme Court in 2005, relying on a series of Indian law cases going back to 1823, specifically cited the Doctrine in its decision denying the right of the Oneida Indian Nation of New York to regain its territory. Justice Ruth Bader Ginsburg wrote in the 2005 decision. “Under the Doctrine of Discovery ...fee title to the land occupied by Indians when the colonists arrived became vested in the sovereign – first the discovering European nation and later the original States and the United States.” [1]

In Simpson’s critique of the use of the doctrine and the tweaking of the definition of the term “terra nullius” in Australia, he commented that “precedent is a deity greater than universally accepted history in some cases... the judiciary ignored international law and history and called its decisions ‘precedent’; in Mabo, it rewrote international law and the common law, and called the decision ‘justice’. What must the

“**Occupation** derives from the natural mode of acquisition in Roman law known as *occupatio*. *Occupatio* could only confer title over objects which were *res nullius* - i.e. belonging to no-one”.

The doctrine, of course, became known as *terra nullius* when it was applied exclusively to land rather than objects generally. If land was *terra nullius* it could be acquired through occupation. The corollary to this was that title **could only be acquired through occupation if the land was *terra nullius***. *Terra nullius* was land that was either deserted or uninhabited ...or inhabited by uncivilized or disorganized groups (this was the general international law view).

In cases where the land was occupied by peoples having a system of social organization, land could only be acquired or colonized through either **conquest** or **cession (treaty)**. This was typically the practice in Asia, Latin America and North America... In Asia, *terra nullius* was thought to have little relevance to the well-organized tribal societies in existence at the time of European colonization, and most territorial acquisitions occurred by **cession or treaty**... The Spanish, on the other hand, acquired sovereignty over Latin America by **conquest**...while in North America a whole variety of methods were used ranging from treaties to conquest, but generally not mere occupation.”

In New Zealand the Maori people were thought to fall into category of **cession**, and therefore treaties were concluded between the indigenous inhabitants and the European settlers.... Notoriously, of course, Australia was regarded as falling into the category of ***terra nullius***.

Figure 1: Methods of acquisition for Land and Property [Ex. 9]

Large Language Models, the ‘Doctrine of Discovery’ and ‘Terra Nullius’ Declared Again?

original inhabitants of this land make of such mysticism?” [9]

The argument now links again to our questions about AI and property rights. How will these legal games play out in the face of today’s new “terra nullius” being the data libraries, datasets, and scrapings from territories of the internet, newly ‘discovered’ by the creators of LLMs such as ChatGPT and their design and use of generative AI systems based on these implicit territorial claims?

In the face of this new attempted form of colonization, major risks exist for users and owners of systems, and major battles lie ahead over intellectual property rights.

In the New Zealand context, the users of chatbots such as ChatGPT should exercise caution over ownership issues with copyright experts questioning “who ‘owns’ parts of the essay, song lyrics, poems, speeches, blogs or other features that the chatbot spurts out? [arguing that] the chatbot technology arrived so quickly that users have not had time to think through the implications.” [8]

New Zealand copyright expert Moon has made the point that “somewhere along the way, those words, sentence sequences, images and sounds are likely to have been input by a human. And under New Zealand law, that content is automatically protected by copyright for the life of the author or creator, and 50 years beyond that person’s death.” [8]

While Moon has observed that “in the US, only works produced by a human can be registered. But in New Zealand, all computer-generated work, including ‘new’ content created by a chatbot like ChatGPT, is protected under copyright for 50 years. But apart from ChatGPT outputs, there is a copyright risk for AI users who “educate” or “train” their systems to generate material, Moon says. That includes AI systems like GitHub Copilot, used to help write computer programs, which are protected under literary works.” [8]

Further “Moon predicts there could well be a debate over whether New Zealand should drop its copyright protection for computer-generated works altogether. He thinks it’s a debate worth having, given that few other countries offer that protection.” [8]



Again, in other countries, such as Australia and the US, we are seeing debates and lawsuits over AI and intellectual property rights. In Australia concerns have been raised about the Books3 dataset, where Josephine Johnston, chief executive of Australia’s Copyright Agency, described the Books3 development as ‘a free kick to big tech’ at the expense of Australia’s creative and cultural life. ‘We’re going to need greater transparency – how these tools have been developed, trained, how they operate – before people can truly understand what their legal rights might be,’ she said... Australian copyright law protects creators of original content from data scraping. Litigation in the US against ChatGPT creator OpenAI over use of allegedly pirated book datasets, Books1 and Books2 (which do not appear to be affiliated with Books3) has already commenced. [2]

So, the risks and legal exposure for educational institutions and students using generative AI through LLMs and systems such as ChatGPT are unclear. The users of ChatGPT agree to the terms of use, which among other things restrict users from such misuse as “represent that output from

the Services was human-generated when it is not.,” [7] so students who plagiarise using ChatGPT are basically in violation of its terms of use. In response then should institutions report offenders to Chat GPT? Users indemnify ChatGPT against any claims, losses, and expenses, with legal jurisdiction based in the US. What exposure this involves for students, academics and their institutions is unclear? User contributions (unless they explicitly opt out) are made available to ChatGPT “When you use our non-API consumer services ChatGPT or DALL-E, we may use the data you provide us to improve our models.” [7] So what implications does that have for our students and the confidentiality and intellectual property rights over the work assigned to them by academics?

But overall, the echoes of well-rehearsed strategies for colonization shine through. Not so far off are the mindsets of cultural and racial superiority that justified the legal subterfuge of “terra nullius,” when ‘big tech’ seeks to arrive and take the digitized traces of our lives and use them without permission or recompense to generate copies and derivative analogues for their own commercial purposes.

Yet is there an inevitability to the forces of capitalism, driving monopolistic big tech companies seeking to exploit

these new opportunities? In the face of this inexorable private 'land grab' an ongoing 'regulatory race' is one necessary response as a strategy for asserting human rights through the public sphere. Capitalist societies have long shown the ability to regulate natural monopolies for the common good, for instance the European Union's definition of 'big tech' companies as 'utility providers' of "very large commercial online platforms." [5] Ironically the European Union, the original source of colonisation, has shown insight through the ability to act with respect to AI through its recent "Regulation on Artificial Intelligence (the EU AI Act)" [5]. The Act classifies AI into four levels of risk based on the intended use of a system: 1) unacceptable; 2) high; 3) limited, and 4) minimal risk, where the Act is most concerned with 'high-risk AI.'

I can envisage the dominant large language models in due course being similarly defined as 'utility platforms' with accompanying and evolving regulations,

as a likely development to rebalance the private and public spheres. With our students we need to become aware that resistance and vigilance on the part of citizens in these data wars will be a necessary part of this ongoing struggle against neo-digi-colonization. ❖

References

1. Bess, R. New Zealand's Treaty of Waitangi and the doctrine of discovery: Implications for the foreshore and seabed. *Marine Policy*, 35, 1 (2011), 85-94.
2. Burke, K. 'Biggest act of copyright theft in history': thousands of Australian books allegedly used to train AI model. *The Guardian*, September 28, 2023; <https://www.theguardian.com/australia-news/2023/sep/28/australian-books-training-ai-books3-stolen-pirated>. Accessed 2023 Sep 30.
3. Chappell, B. The Vatican repudiates 'Doctrine of Discovery,' which was used to justify colonialism. *NPR*, <https://www.npr.org/2023/03/30/1167056438/vatican-doctrine-of-discovery-colonialism-indigenous>. Accessed 2023 Sep 30.
4. Clear, T. Google™ - "Do No Evil" - Yeah Right! *SI/GCSE Bulletin*, 38, 4 (2006), 8-10.
5. Edwards, L. The EU AI Act: a summary of its significance and scope *Artificial Intelligence (the EU AI Act)*. (London, UK: Ada Lovelace Institute, 2021). <https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/>. Accessed 2023 Dec 15.
6. Liesenfeld, A., Lopez, A. and Dingemans, M. Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned

text generators. *arXiv preprint arXiv:2307.05532*.

7. OpenAI. "Terms of Use", March 14, 2023. <https://openai.com/policies/terms-of-use>. Accessed 2023 Sep 30..
8. Phare, J. Warning: Using AI chatbots like ChatGPT could get you sued for copyright breaches. *NZ Herald*, September 11, 2023; <https://www.nzherald.co.nz/nz/warning-using-ai-chatbots-like-chat-gpt-could-get-you-sued-for-copyright-breaches/BWZGMWDN6JHG6VFAQI6U6DLYSSM/>. Accessed 2023 Sep 30.
9. Simpson, G. Mabo, international law, terra nullius and the stories of settlement: an unresolved jurisprudence. *Melbourne University Law Review*, 19 (1993), 195.
10. Vardi, M.Y. Move fast and break things. *Communications of the ACM*, 61, 9 (2018), 7-7.
11. Zuboff, S. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30,1 (2015), 75-89.
12. Zuboff, S. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. (London: Profile Books, 2019).



Tony Clear
 School of Computing and
 Mathematical Sciences
 Auckland University of Technology
 Private Bag 92006
 Auckland, 1142 New Zealand
 Tony.Clear@aut.ac.nz

DOI: 10.1145/3638564

Copyright held by author.

Digital Dreams Have Become Nightmares What We Must Do

Second Edition

Ronald M. Baecker
 with Jonathan Grudin

ISBN: 979-8-4007-1768-0
 DOI: 10.1145/3640479

<http://books.acm.org>



 **ACM BOOKS**
 Collection III

 ASSOCIATION FOR COMPUTING MACHINERY