Advance online publication https://doi.org/10.5204/lthj.4001

Addressing Inequitable Openness in Licences for Sharing African Data and Datasets Through the Nwulite Obodo Open Data Licence

Chijioke I Okorie
University of Pretoria, South Africa
Melissa Omino
Strathmore University, Kenya

Abstract

This article examines the relationship between Standard Public Open Licences (SPOLs) and inequity in the artificial intelligence (AI) innovation ecosystem, focusing on how these licences affect access to and use of African datasets. While SPOLs are widely promoted as tools for democratising data access, they often apply uniform conditions to all users, disregarding disparities in infrastructure, capacity and socioeconomic context. As a result, SPOLs may unintentionally reinforce exclusion and enable extractive data practices that disadvantage communities contributing valuable datasets that they have preserved and curated through historically challenging conditions. The study employs a desktop literature review of primary and secondary sources, complemented by analysis of specific case studies from the Masakhane Research Collective in Natural Language Processing and qualitative vignettes based on real-world experiences to identify inherent and systemic limitations of current SPOLs. The research shows how existing SPOLs, particularly those founded on copyright law, fail to accommodate the positionality of African and similarly situated users in the global data economy. In response, the article introduces the Nwulite Obodo Open Data Licence (NOODL Licence), a novel, tiered SPOL designed to foster equitable openness. NOODL differentiates conditions of use based on users' geography and development context, incorporating benefit-sharing obligations and context-sensitive terms. It maintains the simplicity and legal clarity of existing SPOLs while addressing their inequities. By critically analysing the overlooked relationship between SPOLs and inequity, this article contributes a practical, context-aware licensing alternative that centres communities. While grounded in the African experience, the NOODL framework offers a replicable model for promoting fairness and inclusivity in global data governance and AI innovation.

***Keywords**:* Standard public open licences; NOODL Licence; datasets; inequities; equitable licensing; African datasets licensing.

1. Introduction

The rapid advancements in Artificial Intelligence (AI) are reshaping industries, economies and societies globally. This transformative potential is deeply reliant on the availability of vast, diverse and high-quality datasets. A dataset refers to a collection, arrangement or compilation of various forms of data presented in a specific manner for an identifiable purpose.[1] The related content and purpose make datasets highly relevant and indispensable for AI innovation. Standard public open licences (SPOLs) – which provide predetermined terms for access, reuse and/or redistribution of data with minimal restrictions – have emerged as key tools for distribution and sharing of datasets.[2] Widely used SPOLs for sharing datasets include Creative Commons Zero (CC0), CC BY (Attribution), CC BY-NC (Attribution-NonCommercial), Apache 2.0, Business Source Licence

[1] Renear, Definitions of Dataset, 1.
[2] An open licence is one that grants permission (without prior request) to access, reuse and redistribute a work with few or no restrictions. The reference to *standard* and *terms already set* differentiates this kind of open licence from other licences that grant permissions for access and use of protected materials freely to specific and known private persons or entities.

2.0 (BSL 2.0), and the Community Data License Agreement 2.0. These licences promote AI innovation[3] and simplify data dissemination by offering legal clarity and enabling global interoperability.[4]

However, as the demand for open data from low- and middle-income countries (LMICs) grows, critical questions arise regarding ownership, control and the equitable distribution of benefits.[5] While SPOLs promote openness,[6] they often fail to address imbalances in power and access – particularly when datasets originate from under-resourced communities but are primarily exploited by better-resourced global actors.

One example is the Creative Commons Zero (CC0) licence, which places data in the public domain. While seemingly empowering, CC0 often enables extraction without reciprocal value or recognition. For instance, when linguistic or cultural data from marginalised communities are reused by commercial entities, the lack of protections may reinforce existing inequities – a phenomenon sometimes termed 'digital colonialism'.[7] Such dynamics hinder inclusive innovation and contribute to the global inequality in AI development.

SPOLs also present systemic, technical and legal shortcomings that complicate their application across diverse socioeconomic and cultural contexts. While many studies highlight the benefits of SPOLs – particularly their alignment with the principles of responsible and ethical AI – their potential to produce or reinforce inequities within the data ecosystem is less frequently addressed. These inequities, understood here as injustice occasioned by equality or inequality, are especially pronounced in the context of African datasets.[8] Despite longstanding debates over the existence of property or licensable rights in such datasets,[9] and the broader questions about the appropriateness of SPOLs within African legal and ethical contexts,[10] recent experience shows that SPOLs remain a popular choice for sharing African datasets.

On the one hand, SPOLs are widely celebrated for enabling clear relinquishment of rights in licensed materials (including datasets), offering transparent specifications for reuse,[11] and fostering innovation and collaboration through streamlined permissions and global interoperability.[12] On the other hand, they have been critiqued for failing to respond adequately to the power asymmetries and contextual differences that characterise global data ecosystems. For example, ambiguities around terms such as 'non-commercial' in Creative Commons licences have generated uncertainty for dataset users, contributed to potential loss of income for data creators and limited sublicensing options.[13] SPOLs also lack robust enforcement mechanisms and make it difficult for licensors to prove damages in cases of violation or misuse.[14] Specific SPOLs such as the CC0 licence, which aim to waive all rights for maximal openness, may clash with national laws – particularly in Europe, where moral rights cannot be waived – thereby undermining the legal certainty that SPOLs claim to offer.[15]

Moreover, the one-size-fits-all approach of SPOLs fails to account for key contextual factors such as differences in technological infrastructure, institutional capacity and legal systems. As a result, dataset users in high-income countries are typically better equipped to extract value from open datasets, while users in African and other low- and middle-income countries often lack the same capacity. This creates structurally unequal outcomes under formally equal licensing terms, a dynamic that reproduces rather than redresses global data inequities.

In the African context, SPOLs continue to be promoted as valuable tools for enabling broader access to datasets and accelerating AI development.[16] However, their limitations have spurred interest in less-restrictive SPOL variants – such as CC0 – as more suitable options for the dissemination and distribution of datasets.[17] While zero-restriction licences are often seen as optimal

[3] Labastida, "Licensing FAIR Data," 199–202; Schindler, "Data Documentation."
[4] Goss, "Codifying a Commons," 965; Forsythe, "Creative Commons," 346–348; West, "Little Victories," 903.
[5] Craig, "The AI-Copyright Trap," 7–10; Okorie, "Challenges of Copyright." Also, Senftleben, "Generative AI," 1535–1540.
[6] Labastida, "Licensing FAIR Data for Reuse," 199–202; Schindler, "Data Documentation Beyond Provenance."
[7] Coleman, "Digital Colonialism," 417.
[8] An African dataset is a collection, arrangement or compilation of various forms of data about or from African people, African language, culture, resources or environments, presented in a specific manner for an identifiable purpose. Adapted from the definition of 'Māori data' in Te Mana Raraunga, "Māori Data."
[9] Craig, "The AI-Copyright Trap," 7–10.
[10] Okorie, "Challenges of Copyright." See also Senftleben, "Generative AI," 1535–1540.
[11] Goss, "Codifying a Commons," 965; Forsythe, "Creative Commons," 346–348.
[12] West, "Little Victories," 903.
[13] Artusio, "A Sui Generis Challenge"; Geiger, "Freedom of Artistic Creativity," 413–415.
[14] Labastida, "Licensing FAIR Data for Reuse," 199–202; West, "Little Victories," 903.
[15] Labastida, "Licensing FAIR Data for Reuse," 199–202.
[16] Ncube, "Intellectual Property and Fourth Industrial Revolution."
[17] Okorie, "Challenges of Copyright."

for openness, licences such as the Creative Commons Attribution ShareAlike licence, with their reciprocal use obligations, are still considered relevant in contexts where ethical data reuse and redistribution are concerns.[18] The complexity of legal and regulatory frameworks surrounding data, particularly when combining data from multiple sources with different SPOLs, also introduces significant legal uncertainty, exacerbated by the absence of comprehensive compatibility information or guidance.[19] Additional concerns include the potential for misuse of licensed data,[20] whether through non-compliance with licence terms or by using the data in ways that were never intended by the original creators.[21] Nevertheless, SPOLs remain highly regarded in scholarly and policy circles for advancing the FAIR data principles – ensuring that datasets are findable, accessible, interoperable and reusable.[22]

In most studies of the use of SPOLs for sharing datasets, these licences are portrayed as tools that democratise data access and, in doing so, accelerate AI innovation.[23] Their emphasis on openness, standardisation and permissive reuse is widely viewed as central to advancing technological development. However, much of this literature overlooks a critical dimension: the structural inequities that arise from applying SPOLs to African datasets specifically.[24] Because SPOLs are designed as public and uniform instruments, they typically treat all users as part of a homogeneous global community, without acknowledging or accounting for the significant disparities in the positionality, capacity and access to infrastructure needed to make meaningful use of the licensed datasets. This universalist approach assumes a level playing field where one does not exist. In the context of African datasets, this results in inequity – namely, injustice that stems from treating unequally situated actors as if they are the same. For example, users in high-income countries often benefit from the well-established technological ecosystems, robust institutional support and advanced computational infrastructure needed to make speedy and meaningful uses of datasets for AI innovation. In contrast, users in Africa and other developing regions frequently face limited connectivity, resource constraints and restricted access to AI tools – all of which undermine their ability to derive comparable value from the same open datasets.[25] The identical licensing terms, when applied across these unequal contexts, produce asymmetrical outcomes. Far from levelling the playing field, SPOLs can entrench disparities in who benefits from open data and who remains excluded from its promise – injustice brought about by treating unequal persons the same or equally.

Conventional open data licensing models, including SPOLs, fall short of facilitating equitable data-sharing or ensuring fair benefit distribution, particularly for historically marginalised or under-resourced communities. Without new approaches that incorporate reciprocity, recognition and context-sensitive terms, data governance risks reinforcing global inequality.

To illustrate these systemic issues and concerns, this article uses a set of vignettes, each of which highlights real-world consequences of applying SPOLs to datasets originating in Africa. These stories humanise the abstract problem of licensing inequity and demonstrate why reform is urgently needed. Further, this article proposes and explores a novel solution to the problem: the Nwulite Obodo Open Data Licence (NOODL Licence).[26] NOODL introduces a tiered licensing structure designed to address the inequities embedded in current SPOL frameworks. It seeks to enable differentiated access and benefit-sharing arrangements that align with the capacities and needs of diverse data users. The article further demonstrates how a novel, tiered data licensing framework such as the NOODL Licence can serve as an effective tool for redressing inequities embedded in traditional open data models, highlighting the implications for inclusive data governance and AI development in the Global South.

This study contributes to global debates on data ethics, legal innovation and equitable AI by examining a practical alternative. NOODL is a new initiative developed in 2024 by the authors of this article through collaboration with African dataset creators and other legal scholars. Unlike existing SPOLs, NOODL integrates contextual access controls, benefit-sharing obligations and culturally sensitive provisions into its framework. Its proposal represents a significant shift from merely identifying problems to offering actionable, locally grounded solutions. By critically analysing both the limitations of current SPOLs and the potential of NOODL, this article offers meaningful insights for policy-makers, dataset owners, developers and communities advocating for equitable AI. It builds a foundation for rethinking data governance models that are not only open but also fair, inclusive and culturally relevant in the African context.

[18] Szkalej, "Generative AI," 313.
[19] Artusio, "A Sui Generis Challenge."
[20] Khayyat, "Open Data Licensing," 231–233.
[21] Geiger, "Freedom of Artistic Creativity," 413–415.
[22] Margoni, "The Role of Licensing."
[23] Margoni, "The Role of Licensing."
[24] Siminyu, "Consultative Engagement of Stakeholders," 5.
[25] Widder, "Open for Business," 5; Siminyu, "Consultative Engagement," 5–6; Okorie, "Challenges of Copyright."
[26] See www.licensingafricandatasets.com

The rest of this article is presented in four sections. Section 2 maps the relevant properties of datasets reflected in key provisions of SPOLs, focusing on the six Creative Commons Licences, Apache 2.0, BSL 2.0 and Community Data License Agreement 2.0. This is followed in Section 3 by vignettes based on discussions with African data owners, illustrating inequities in the open data licensing environment. Section 4 proposes the NOODL Licence as a novel framework embodying tiered access, contextual justice and value-sharing mechanisms, which addresses the identified inequities. Section 5 concludes with key insights, policy implications and directions for further research.

2. Analysis of Standard Public Open *Data* Licences

In describing a licence, in *Thomas v Sorrell*, Vaughan J stated: 'A dispensation or licence properly passeth no interest, nor alters or transfers property in anything, but only makes an action lawful, which without it had been unlawful.'[27] Contextualised to African data and African datasets, a licence is permission to use data and datasets belonging to specific African person(s) such that, without the permission, the use would be infringing and/or unlawful. As such, a data licence arises because the person granting it has the power or legal right to do so. This power can emanate from various sources such as law, custom or usage.[28] This article focuses on the dataset owner's own volition in relinquishing some or all of the rights they have over the data/datasets and giving it to another. These rights may arise as a result of a broad range of regulatory frameworks: copyright law, data-protection law and mechanisms such as contracts, trade secrets, confidentiality measures and technological protection measures that are not explicitly statutory. To illustrate these bases for data licences, four projects undertaken by Masakhane Research Foundation – Masakhane MT, MasakhaNER, MakerereNLP and NaijaNLP – are helpful.[29]

Masakhane MT resulted in the production of a multilingual (six diverse African languages) scientific parallel corpus of African research, which could be used to, inter alia, create a machine translation tool for African universities to ensure accessibility of their publications. MasakhaNER involved the creation of monolingual new corpus, parallel datasets for various translated languages and the named entity recognition (NER) and parts of speech (POS) annotated datasets. In MakerereNLP, the outputs included text datasets and speech datasets for specified African languages, while NaijaNLP resulted in the creation of sentiment corpus, sentiment lexicon and hate speech lexicon. These datasets from these projects are licensable because of the legal rights inherent in them by virtue of either or all of copyright law, contractual obligations, confidentiality or technological protection measures (TPMs). Further, open licences have been utilised as the form of data licences with which to disseminate these outputs: Creative Commons License (CC-BY 4.0) for outputs from Masakhane MT, NaijaNLP and MasakhaNER, and CC BY-SA licence for outputs from MakerereNLP.

The approach of utilising open licences, including those not designed specifically for licensing datasets, is not limited to Masakhane projects. Across Africa and the rest of the world, open licences designed (mainly) for sharing software (e.g. MIT License), databases (e.g. Open Database License) and/or copyright-protected materials (e.g. Creative Commons Licences) have become increasingly popular for sharing and licensing datasets. These are SPOLs, by the adoption of which any data/dataset owner(s), holder(s) and/or custodian(s) can prospectively provide access to their data/dataset to any member of the public who wants to use the data/dataset on the terms of the relevant licence. Below, we highlight some of the key aspects of the following SPOLs: Creative Commons; Open Data Commons; Community Data License Agreement; MIT License; Apache License 2.0 and BSD 4-Clause License:

- *Licensed rights:* These are the rights the dataset owner has or exercises over the dataset. Generally, the licensed rights that may be explicitly granted or withheld are: the right to reproduce and share the dataset through any means that allows the public to access them; use of the datasets for commercial purposes; the right to circumvent TPMs or make modifications that allow the datasets to be accessed and used; moral rights, such as the right of integrity; publicity, privacy and/or other similar personality rights; patent rights; trademark rights; the right to collect royalties for the licensed rights; and sui generis database rights.

- *Conditions/pricing model/licensees' obligations:* The fact that SPOLs are open does not remove the imposition of conditions of use on the licensees/dataset users. Indeed, it is the licensee's adherence to the conditions of use that makes the licensed use of the dataset valid. The commonly imposed conditions of use include attribution of some sort; the requirement to share any modifications to the datasets using the same or similar licence as the one used to access the licensed material in the first place (so-called share-alike obligations); and the requirement to include a statement of significant changes made to the licensed material.

[27] Para. 351.
[28] Mysoor, Implied Licences in Copyright Law, 12.
[29] Masakhane, "Ongoing Projects."

- *Data quality:* In SPOLs, licensors offer the licensed materials as is and make no warranties. Liabilities are also not accepted. In essence, no information is provided about the quality of the licensed material and no liability is accepted if the licensed material is found wanting in any way. This can be a problem for openness of AI models.[30]

- *Parties:* SPOLs indicate that the licence is between the licensor and each/any person using the work or exercising the licensed rights. This means all users, regardless of status, capacity and/or identity, are permitted to exercise the licensed rights.

3. Picturing and Defining Inequity in Standard Public Open Data Licences for African Datasets

3.1 Vignettes of Inequity[31]

Inequity is a by-product of SPOLs when used to disseminate African datasets. To demonstrate the lived consequences and the systemic challenges of applying SPOLs to African datasets, this section presents two vignettes, each of which illustrates a specific facet of inequity. The first vignette (Kelebogile, Omphile, Katlego, Masana) foregrounds issues of benefit-sharing, exploitation and infrastructural disparities. The second vignette (Nthenya and Lungati) emphasises sustainability, licensing dilemmas and the broader implications for community-centred innovation. Together, they provide grounded examples of how open licensing mechanisms can produce unequal outcomes for under-resourced communities.

3.2 Kelebogile, Omphile, Katlego and Masana

Consider these four African researchers, freelancers and techpreneurs: Kelebogile, Omphile, Katlego and Masana. Kelebogile is a content creator who creates content in Setswana. Her current project, creating a series of educational videos about indigenous plants, is proving a nightmare. The auto-correct function on her laptop continually 'corrects' Setswana words into English. Finding accurate pronunciation and spelling of some Setswana words is also hard because, even though Kelebogile learnt Setswana at home, she lives in Cape Town and does not hear Setswana spoken in her day-to-day interactions. Her aunt and her aunt's church group send her voice notes on WhatsApp pronouncing the names of various indigenous plants in Setswana. Omphile, a linguist, is collaborating with a tech startup developing an NLP model for Sesotho. The project is close to her heart – a way to preserve her grandmother's tongue. But the lack of standardised Sesotho orthography and the scarcity of annotated data are major roadblocks. She knows that a robust digital dictionary and funds to train the model on the nuances of different dialects would make a huge difference. Katlego is a translator and is currently working with an international news agency which has engaged her to translate a story about local innovation from English to Setswana. The challenge is not just the technical jargon; it is finding the right words to convey the spirit and cultural context of the original article.

These three meet at a conference and share their frustrations. Masana, a data scientist, is their mutual friend. Her work at the university and as a tech startup founder focuses mainly on African NLP, so she identifies with these frustrations. All four agree to collaborate on a project to create robust, human-annotated data and a comprehensive named-entity recognition (NER) and parts of speech (POS) corpus for 20 diverse African languages. They plan to undertake this project with other collaborators who speak and/or work on different African languages, including Igbo, Hausa, Twi, Swahili and Dholuo, in the African NLP space.

The challenges are immense: data scarcity; issues around copyright and contractual clearance needed to access any existing data/datasets; linguistic complexity; and getting the cooperation and trust of local and native speakers that can donate text and speech data in their language or review donated and/or annotated data to ensure accuracy. They debate translation strategies, navigate licensing agreements and meticulously review annotations. After months of hard work and with everyone contributing both time and money and a grant from a funder collective – the Akunnaonye Fund – Kelebogile, Omphile, Katlego, Masana and their collaborators across Nigeria, South Africa, Kenya, Uganda and Ghana succeed in creating robust, human-annotated data and a comprehensive NER and POS corpus for 20 diverse African languages. They all want people, especially researchers and tech startups like them across Africa, to easily access these outputs and advance African languages in AI innovation. Akunnaonye Fund advises that it has applied the CC0 licence to this work to ensure that everyone can easily and freely access the work.

[30] Widder, "Open for Business," 5.

[31] The cases presented here are based on the authors' personal discussions and regular engagements with data scientists, AI innovators, tech start-ups, (computational) linguists and tech enthusiasts across Africa at conferences, workshops, seminars, work dinners and so on between 2019 and 2024. The conversation and engagement remain ongoing.

Shortly after the release of the data and the NER and POS corpus, a global tech giant uses these resources to create a commercial AI-powered translation tool. The tech giant puts a lot of effort into marketing this tool, promoting its accuracy and advanced features. But the price is exorbitant in various African countries. Local schools, town halls and community centres, tech startups such as Masana's, and the individuals and researchers who had eagerly contributed to the project find that they are unable to afford it. There is a deep sense of betrayal. Hadn't they shared their stories, their language, their very culture, with the understanding that it would benefit their community?

The project led by Kelebogile, Omphile, Katlego and Masana had envisioned a tool that would empower their communities, not exclude them. The CC0 licence, meant to foster collaboration and accessibility, had inadvertently enabled their community's contributions to be exploited for profit, leaving them behind. The use of the CC0 licences meant all users of the tools were treated equally and granted the same and equal access to the tools. While the African NLP researcher, African tech startup and global tech giant all had access to the datasets because an open licence made that possible, their respective uses of the dataset depended mainly on the kinds of resources and infrastructure (including access to computers and existing development frameworks, among others) that they owned or (may) have had access to. But SPOLs disregard these contextual differences between users, resulting in inequity.

3.3 Nthenya and Lungati

Nthenya is from Kenya and, as a data scientist, understands the deep connection between her community and the land. She knows the local dialect, the intricate farming practices passed down through generations and the unique challenges faced by smallholder farmers. But she also knows that this knowledge, vital for agricultural innovation, is largely undocumented and inaccessible to researchers and policy-makers. Lungati is a computer scientist, and part of a team building a machine learning model to predict crop yields and optimise fertilizer use. He understands the technical requirements for a robust dataset – clean labels, standardised formats and diverse samples. But Lungati is removed from the realities on the ground. He struggles to bridge the gap between raw data and meaningful insights that could truly benefit farmers.

Nthenya and Lungati are brought together by a mutual friend, both driven by the desire to use their skills to address the challenges facing the agricultural sector in Kenya. They recognise the potential of AI, but also the critical need for data that accurately reflects the complex social and ecological landscape. Their project, *Mkulima*, aims to create a labelled dataset of crop images, annotated with information about plant health, soil conditions and farming practices, all grounded in the local context. They are fortunate to receive a startup grant to fund the first few years of the project. The grant is given with the expectation that the project will become sustainable over time and fund its continued existence. Nthenya's expertise is crucial for identifying key variables, ensuring culturally appropriate data-collection methods and translating local knowledge into standardized labels. Lungati's technical skills are essential for designing the data-collection platform, implementing quality control measures and preparing the dataset for machine learning applications.

They partner with local farming cooperatives, training community members to use mobile apps to capture images of their crops and record relevant information in their local language. Nthenya works closely with the farmers, explaining the project's goals and addressing their concerns about data privacy and ownership. She understands that trust is paramount, and that the data will only be valuable if it truly represents the farmers' experiences and perspectives. Lungati focuses on building a user-friendly platform that can handle the diversity of local dialects and capture the nuances of agricultural practices. He incorporates feedback from the community, ensuring that the technology is accessible and adaptable to their needs. The project is not without its challenges. They encounter issues with internet connectivity, language barriers and the time-consuming nature of manual annotation. But Nthenya and Lungati persevere and *Mkulima* becomes more than just a dataset. It provides a bridge between communities and technology, between local knowledge and scientific innovation.

At the largest gathering of African AI researchers, Nthenya and Lungati meet Kelebogile and others who share their frustrations about using SPOLs to make their work publicly accessible and available. Nthenya and Lungati are understandably anxious, which leads them to ask themselves questions: should they use the Creative Commons Attribution NonCommercial No-Derivatives Licence that required attribution from users and disallows commercial uses and modification or adaptations of the datasets? At least that would mean they will not suffer the same fate as Kelebogile – but will everyone still have non-commercial access to the datasets? Also, does it mean that everyone is restricted from improving the datasets or building tools from them? Should they use the Open Data Commons Attribution (ODC-BY) licence? That may mean they will suffer the same fate as Kelebogile, even though their authorship will be acknowledged by users. Come to think of it, how does using any of these licences help them fulfil their obligations to become financially sustainable over time? What about combining and using two or three of these SPOLs? That brings to mind what Lungati's friend from Nigeria often says about choices. Onyemaechi would say that when it comes to African datasets, these licences are like *mma abuo di n'ulo onye ubiam: nke nwere isi adighi nko;*

nke di nko enweghi isi.[32] Combining two or three SPOLs is not a workable solution – each licence has something to commend it, but also something that makes it a bad choice.

All that is needed for inequity to result from applying the SPOLs to these African datasets is for one entity with bigger resources to use the datasets to develop a proprietary AI system or use the datasets to sell another product or service. With several resources at its disposal, in the time that it takes a global tech startup to create a product or service based on the datasets, Nthenya and Lungati will still be struggling with writing and presenting proposals to access additional funding to scale their operations. SPOLs work differently for global tech giants: they are empowering, enabling innovation. But for people like Nthenya and Lungati, SPOLs hamper them compared with their developed country counterparts. Smallholder farmers could be priced out or, worse, they could be paying for a product they contributed to building; they could be exposed to unwittingly sharing land title information that could lead to land grab or loss because developing AI for sectors such as agriculture and climate change may sometimes require the use of Earth observation data or training datasets, which in some instances include land boundaries of farms, among other things.[33] This illustrates inequity in SPOLs for African datasets.

Together, these vignettes illustrate how, by ignoring contextual inequalities among data users, SPOLs can reinforce structural barriers rather than dismantling them. They underscore the need for a more responsive, flexible and equitable licensing approach that accounts for disparities in access, infrastructure, positionality, capabilities and community expectations of dataset users.

3.4 Defining Inequity

As the above vignettes illustrate, a key feature of SPOLs is the grant of licensed rights to everyone regardless of identity and status. Whether a potential dataset user is a large technology company (so-called 'Big Tech') with access to computers, development frameworks and other infrastructure needed to make quick and meaningful use of the datasets and maintain its dominance in the AI innovation space or a small group of researchers in a university in Kenya or a tech startup in Nigeria, SPOLs allow them equal and same access to the licensed datasets. SPOLs treat all dataset users as if they were the same, and only provide for different licence elements (attribution, non-derivative, non-commercial, share-alike) to address the owner's/licensor's preferences on a predetermined set of issues, resulting in inequity.

There is some discussion and conceptualisation of inequity in both scholarly and formal legal discourse.[34] As a concept, inequity has been defined as "inequality or difference resulting from systematic or systemic unfairness".[35] It is a qualitative measure that focuses on the intersection between inequality – lack of sameness – and injustice – systematic or systemic unfairness.[36] Identifying inequity involves examining and understanding the contexts of relevant stakeholders and where the inequality is located.[37] This article focuses on inequity as a concept that may apply in any human or social relations, specifically the use of SPOLs in sharing or granting use and access permission to African datasets and datasets from other under-resourced or historically marginalised regions.

The case of *Blind SA v Minister of Trade, Industry and Competition and 4 Others [Blind SA]*, decided by the Constitutional Court of South Africa (CCSA), offers a classic example of how inequity can be embedded in law while also presenting constitutional questions. In its decision, the CCSA highlighted the differences in the contexts of users of copyright-protected materials and noted that the inequality lies in the ability of users of copyright-protected materials to access copyright-protected materials.[38] Such a difference was one that the Constitution of South Africa prohibited from being a basis for discrimination. According to the court, in paragraph 67:

[32] This was one of late Elder Benjamin Onyemaechi Okoli's popular proverbs. The proverb is Igbo for saying that someone could have two useless or dangerous options because each option has both great and horrible qualities. The literal meaning is: 'Two knives in a poor man's house; the one that has a head is blunt, the one that is sharp has no head.'

[33] If these data are opened up without the necessary privacy-preserving measures, there is a risk of exposing farmers' locations, the size of their land and other sensitive information. Farmers are already hesitant to allow data scientists and AI developers to map without the necessary assurances of confidentiality and privacy. In contexts such as African countries, where land is very valuable, such exposure increases the risk of encroachment and land loss. To address these concerns, the Local Development Research Institute developed a toolkit to guide researchers on how to responsibly share their data for training AI and address these issues. See Local Development Research Institute, "AI Training Dataset Sustainability."

[34] Galanter, "Why the 'Haves' Come out Ahead," 100–104. Yuille, "Inequity as a Legal Principle," 859.

[35] Yuille, "Inequity as a Legal Principle," 868.

[36] Yuille, "Inequity as a Legal Principle," 859.

[37] Kuttner, "Equity vs Equality Graphic."

[38] *Blind SA v Minister for Trade, Industry and Competition* (2022) ZACC 33, 67.

> Legislation that protects the rights of copyright owners must take account of the differential impacts of such protection upon different classes of persons. In this case, how the requirement of authorisation bears upon the access that persons with print and visual disabilities have to literary works in comparison to the access enjoyed by persons without these disabilities. Where those with print and visual disabilities suffer great and particular hardship by reason of the requirement of authorisation, the requirement cannot be applied as if all persons who need access to literary works are similarly situated, when they are not.

With respect to SPOLs for sharing African datasets, the same conclusion is reached in this way. SPOLs that are used for sharing African datasets must take account of the differential impacts of such open access on different classes of persons – in this case, how the access to other resources and infrastructure needed for AI innovation bears upon the access that persons from Africa (and other developing countries) have to datasets compared with the access enjoyed by persons in developed countries. Those in the latter group have access to and control of other resources and infrastructure needed for AI innovation; when combined with African datasets, this puts them at a considerable advantage. Market dominance and open access to datasets cannot be designed and applied as if all persons requiring access to datasets were similarly situated when they are not – that would be inequitable.

4. Addressing Inequity: The NOODL Licence solution

In some ways, the inequities identified when existing standard open data licences are used to disseminate African datasets are an unintended feature of such licences. In a bid to create a simple, straightforward means for creators and developers to openly share their work, differences in the user ecosystem were ignored in the crafting of standard open licences. However, because the focus of these standard open licences has been on empowering the creator/developer to make a decision about how to share their work, the more appropriate view may be that the inequities are a by-product of standard open data licences specifically in the context of African datasets.

Whatever the case, a solution that responds to the root cause of the inequity is necessary to restore equity. In this regard, the approach of the CCSA in *Blind SA* offers some guidance for addressing inequity caused by treating differently situated persons as if they were the same. According to the court:

> Sometimes, for the state to avoid unfair discrimination, it must treat people in the same way or make available the same entitlements. But sometimes what is required of the state is to recognise the differences between persons and to provide different or more favourable treatment to some, so as to secure non-discriminatory outcomes for all. This may appear paradoxical, but it is not, and cases of persons with disabilities well illustrate this principle.[39]

In the context of SPOLs for African datasets, equitable openness would involve a licence that continues the convention of empowering or enabling owners of African datasets to openly and easily share their datasets, but does so in a manner that responds to the differences in the contexts of users of African datasets.[40] While all dataset users may (and should) have access to and use permissions over the datasets, additional or different conditions of use and access, or both, should be imposed on better-resourced users of African datasets. To address the differences in the contexts of users of African datasets and the inequities that presents, SPOLs should take a broad(er) view of conditions of use. A SPOL that requires (certain) users to fulfil additional obligations as a condition for use may still be open insofar as some access and use is permitted. In this regard, exchange conditions calibrated to specific user contexts can address the inequities that arise from imposing the same conditions of use on all users. For African datasets, a SPOL is needed that gives African datasets owners a standardised way to grant differently situated members of the public the permission to use their datasets in a way that recognises and addresses such differences. From the dataset user's perspective, the presence of such a licence in relation to an African dataset answers two questions: What can a user of my status do with this dataset? And how can I contribute to growing the innovation enabled by this dataset? In this regard, the dataset user is recognised and integrated into the creative and innovation process.

Responding to the differences in the contexts of users of African datasets must also occur in a single licence document if the tradition of simplicity of sharing mechanism is to continue. Such approach also addresses that complexity of trying to decide between two or more licences to use in sharing African dataset across multiple users from different contexts.
Table 1 presents an analysis of SPOLs that are used for sharing and licensing datasets, including the NOODL Licence discussed more fully in this article.

[39] Para 68.
[40] Okorie, "Challenges of Copyright."

Table 1. An analysis of SPOLs that are used for sharing and licensing datasets

SPOLs	Licensed rights/permissions						'Pricing' model/Conditions imposed				Data quality		Parties	
	Copyright and similar rights	Modifications/adaptation	Circumvention of TPMs	Commercial use	Moral, privacy and personality-related rights	Patents and trade marks rights	Attribution	Share alike	Include copyright notice	Statement of significant changes	Warranties	Acceptance of liability	Data provider	Envisaged dataset user
CC BY 4.0	+	+	+	+	W	N	+	N	+	N	N	N	S	Anyone
CC BY-SA	+	+	+	+	W	N	+	+	+	N	N	N	S	Anyone
CC BY-NC	+	+	+	N	W	N	+	+	+	N	N	N	S	Anyone
CC BY-NC-SA	+	+	+	N	W	N	+	+	+	N	N	N	S	Anyone
CC BY-ND	+	N	+	+	W	N	+	N	+	N	N	N	S	Anyone
CC BY-NC-ND	+	N	+	N	W	N	+	N	+	N	N	N	S	Anyone
CC0	+	+	+	+	W	N	N	N	+	N	N	N	S	Anyone
Open Database License	+	+	+	+	W	N	+	+	+	N	N	N	S	Anyone
Open Data Commons Attribution License	+	+	+	+	W	N	+	N	+	N	N	N	S	Anyone
Open Data Commons Public Domain Dedication License	+	+	+	+	W	N	N	N	+	N	N	N	S	Anyone
Community Data License Agreement	+	+	+	+	W	N	N	N	+	N	N	N	S	Anyone
MIT License	+	+	+	+	W	N	N	N	+	N	N	N	S	Anyone
Apache License 2.0	+	+	+	+	W	+/N	+	N	+	+	N	N	S	Anyone
BSD 4-Clause License	+	+	+	+	W	N	+	N	+	N	N	N	S	Anyone
NOODL License	+	+	+	+	W	N	+	+	+	N	N	N	S	2 groups

+: Granted/Required N: No/Not granted S: Stated W: Waived

The proposal outlined above is embodied in the Nwulite Obodo Open Data Licence (NOODL Licence) developed in 2024 by the authors of this article, with the collaboration of other researchers including from the Data Science Law Lab.[41] It emerged in response to growing concerns that SPOLs, while valuable for enabling open access, were not designed to address context-specific inequalities in data use. NOODL Licence is a tripartite licence between owners of African datasets, users in or from Africa and developing nations, and users outside Africa and developing nations. It is a unified licence offering one document for all parties, addressing their specific needs and contexts. NOODL builds upon the structure of SPOLs but adapts them to explicitly recognise disparities in capacity, infrastructure and benefit-sharing. This section outlines the key features of NOODL and how it proposes to rectify the systemic issues identified in the previous vignettes.

Grounded in empirical insights drawn from real-world case studies and informed by comparative legal analysis, the development of NOODL responds directly to the central research question posed in this study: How can tiered licensing frameworks foster more equitable data-sharing and inclusive AI development?

[41] See also, Heikkilä, "Grassroots Effort."

The NOODL solution grants permission over same or similar rights over datasets to both kinds of users or for both groups of users, but the conditions differ slightly. Users in or from Africa and developing nations are required to share alike with others (people and organisations) from Africa and from developing nations any modifications or adaptations or derivatives from the datasets, under the same licence that applied to the original dataset – that is, the NOODL Licence. Where such users intend to make commercial use of the datasets, this has to be outside Africa and developing nations. Essentially, they cannot sell products or outputs from the datasets to the members of their community. Users outside Africa and developing nations have the same obligation, but are also required to provide some kind of benefit or interest as may be required by the specific community that is sharing or licensing the dataset in question. This benefit could be in the form of monetary compensation, access to specific infrastructure such as computers, access to human and physical resources or equal collaboration arrangements.

With this approach, the NOODL Licence joins the league of SPOLs and retains the benefits associated with SPOLs while addressing the inequity arising from such licences. Individual creators still have a standardised way to grant permissions and legal clarity in terms of their rights over their datasets. Community creators are empowered, and are afforded the option of choosing how to share datasets. With the additional conditions imposed on specific user groups, community creators can receive support for community-driven innovation.[42] Dataset users continue to enjoy access to datasets under conditions that address their different contexts and positionality in the innovation ecosystem. AI and tech product developers have access to diverse datasets and clear guidelines on data usage. African NLP researchers and computational linguists have support for open data practices and enjoy enhanced collaboration and innovation, and dataset creators and communities enjoy fair compensation and benefits as well as protection against extractive uses of their datasets. Compensation here goes beyond monetary payment and can involve having a collaborative work arrangement where the people who created the datasets get to work on applications or models or products that are built on those datasets, together with whichever user is outside Africa or outside the developing nation. Funding organisations that require open licences for sharing outputs from funded projects have a simplified licensing process and an additional SPOL option that offers the assurance of community benefits.

The key difference between the NOODL Licence and other standard open data licences is that in the NOODL Licence there is one licence document: an arrangement between the licensor and differently situated licensees as described in the licence, which imposes additional use conditions on users outside Africa or developing countries. In essence, the NOODL Licence creates a tiered arrangement of African dataset users. What distinguishes NOODL from other SPOLs is its built-in differentiation between user groups based on geography and development context, and its incorporation of community benefit obligations – all within a single, unified licence document. It does this while preserving the core features of openness. While most SPOLs apply a uniform set of terms to all users regardless of context, NOODL operationalizes fairness by tying access conditions to users' geographic and economic positioning.

5. Conclusion and Future Work

Proposing a different approach to standard open data licensing and a different SPOL for African datasets represent an important early stage of the work being carried out in the licensing landscape in relation to AI innovation in Africa. Future work is necessary to propagate the new licence. Adding another licence to the mix – even one that addresses identified inequities with existing options – contributes to the proliferation of open data licences that may make it challenging and confusing for researchers and dataset owners, data-holders and/or data custodians to choose an appropriate licence. This problem does not mean the inequitable situation should be allowed to continue; instead, care should be taken and future work should focus also on providing comprehensive resources and guidance to the public on how to select appropriate licences. There is also a need for research and practical solutions towards the enforcement of the NOODL licence and a needed push towards an understanding of the value of African datasets and the bargaining power that value yields. This can help mitigate confusion and ensure proper usage. Furthermore, engaging the research community in discussions about best practices for data-sharing and licensing can help to establish clearer norms and expectations, reducing the likelihood of misuse. It is also important to adopt standard exchange conditions and undertake the creation of licence-compatibility tools that address the confusion arising from using diverse-licensed datasets.

The current structure of the NOODL Licence for responding to the differences in the contexts of African dataset users is based on geography and economic development status. However, this is only one option. In some other cases, a sector-based approach to differentiation is appropriate – different kinds of datasets being licensed differently. Furthermore, while the NOODL Licence is a viable alternative for funding organisations to consider adopting or recommending as *one of* the licences for grantees to use in sharing outputs from funded projects, one step towards fairness and equity in the data landscape must be for funders to be more liberal in allowing grantees to choose from a range of open licences rather than contractually prescribing a specific open

[42] Wairegi, "Just Because You Can"; Okorie, "Challenges of Copyright."

licence for grantees to use to share their outputs. Although this article focuses on African datasets, the broader framework of the NOODL Licence may also be applicable to similarly situated communities in other developing regions, where comparable inequities in data access and benefit distribution persist.

The call in this article for an equitable principles-based approach to open licensing therefore seeks to make equity, rather than equality, the foundation of open licensing of datasets in the African context. But it is recognised that, due to the current limits of technology, multiplicity of open licences for datasets may be difficult for persons without legal training to navigate and it may be difficult to enforce the terms of a publicly available open licence for datasets that are equally made publicly available. For example, once datasets are made publicly available using an open licence, licensors may not be able to trace, let alone proceed against, potential infringers and/or those who go ahead contrary to the licence terms. In these cases, it is proposed that government bodies establish and enforce a standard to which all AI developers should adhere in using datasets emanating from other persons.

Finally, while an alternative open licence can address some of the inequities in the data ecosystem towards building ethical AI and innovation generally, licensing is just one approach to dealing with the broader inequities in the global AI ecosystem and more work is needed to address other equitable approaches for a better AI ecosystem.

Acknowledgements
The work in this article received funding from Mozilla Foundation and Meta to the Data Science Law Lab as well as funding support from the University of Pretoria's Research Development Grant program to Chijioke Okorie. The authors acknowledge the input of Professor Vukosi Marivate, UP-ABSA Chair of Data Science, the Data Science Law Lab team members, members of the Data Science for Social Impact research group at the University of Pretoria and participants at various conferences and workshops where the work was presented, especially Deep Learning Indaba Dakar 2024. The authors take responsibility for all arguments, statements and errors and none of the funders and supporters accepts any liability in this regard.

Chijioke Okorie was Principal Investigator and Melissa Omino was Co-Investigator for the project leading to the NOODL Licence described in this article. Chijioke Okorie conceived the idea for this article and produced the manuscript, with contributions from Melissa Omino in refining the article. Both authors engaged in discussions throughout the project. The authors have reviewed all content carefully and assume full responsibility for the manuscript's final form.

Corresponding author: Chijioke I Okorie, Associate Professor at the University of Pretoria, South Africa, and the Principal Investigator of Data Science Law Lab. chijioke.okorie@up.ac.za

Bibliography

Artusio, Claudio and Federico Morando. "Creative Commons 4.0 Licenses: A Sui Generis Challenge?" *Conference for E-Democracy and Open Government* (2014): 269–275.

Bartholomew, Mark. "Copyright and the Creative Process." *Notre Dame Law Review* 97 (2021): 357. https://scholarship.law.nd.edu/ndlr/vol97/iss1/8.

Braunschweig, Katrin, Julian Eberius, Maik Thiele and Wolfgang Lehner. "The State of Open Data." *Limits of Current Open Data Platforms* 1, (2012): 72.

Coleman, D. "Digital Colonialism: The 21st Century Scramble for Africa Through the Extraction and Control of User Data and the Limitations of Data Protection Laws." *Michigan Journal of Race and Law* 24 no 2 (2018): 417. https://doi.org/10.36643/mjrl.24.2.digital.

Craig, C.J. "The AI-Copyright Trap." *SSRN* (2024): 1–21. http://dx.doi.org/10.2139/ssrn.4905118.

Data.org. "Seeking Shared Incentives: Cultivating Cross-Sector Data Collaboration to Combat the Climate Crisis." https://data.org/events/seeking-shared-incentives-cultivating-cross-sector-data-collaboration-to-combat-the-climate-crisis.

de Castilho, Richard Eckart, Giulia Dore, Thomas Margoni, Penny Labropoulou and Iryna Gurevych. "A Legal Perspective on Training Models for Natural Language Processing." In *LREC 2018: Eleventh International Conference on Language Resources and Evaluation*. European Language Resources Association (ELRA), 1267–1274. Miyazaki, 2018.

Dixon, Martin. *Modern Land Law*. London: Routledge, 2023.

Dodds, Leigh. "Describing Dataset Archetypes." *Patterns* 1, no 9 (2020). https://doi.org/10.1016/j.patter.2020.100168.

Forsythe, Lynn M. and Deborah J. Kemp. "Creative Commons: For the Common Good." *University of La Verne Law Review* 30 (2008): 346.

Galanter, Marc. "Why the 'Haves' Come Out Ahead: Speculations on the Limits of Legal Change." *Law & Society Review* 9, no 1 (1974): 95–160. https://doi.org/10.2307/3053023.

Geiger, Christophe. "Freedom of Artistic Creativity and Copyright Law: A Compatible Combination." *UC Irvine Law Review* 8 (2018): 413–458.

Geradin, Damien and Miguel Rato. "Can Standard-Setting Lead to Exploitative Abuse? A Dissonant View on Patent Hold-Up, Royalty Stacking and the Meaning of FRAND." *European Competition Journal* 3, no 1 (2007): 101–61. https://doi.org/10.1080/17441056.2007.11428452.

Goss, Alex K. "Codifying a Commons: Copyright, Copyleft, and the Creative Commons Project." *Chicago-Kent Law Review* 82 (2007): 963. https://doi.org/10.26412/psr215.04.

Heikkilä, Melissa. "How This Grassroots Effort Could Make AI Voices More Diverse." *MIT Technology Review*, 15 November. https://www.technologyreview.com/2024/11/15/1106935/how-this-grassroots-effort-could-make-ai-voices-more-diverse.

Joshi, P., S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. "The State and Fate of Linguistic Diversity and Inclusion in the NLP World." *arXiv preprint* (2020). http://doi.org/10.18653/v1/2020.acl-main.560.

Khayyat, Mashael and Frank Bannister. "Open Data Licensing: More than Meets the Eye." *Information Polity* 20, no 4 (2015): 231–252. https://doi.org/10.3233/IP-170406.

Kuttner, Paul. "The Problem with That Equity vs Equality Graphic You're Using." *Cultural Organizing*, 1 November. http://www.socialventurepartners.org/wp-content/uploads/2018/01/Problem-with-Equity-vs-Equality-Graphic.pdf.

Labastida, Ignasi, and Thomas Margoni. "Licensing FAIR Data for Reuse." *Data Intelligence* 2, nos 1–2 (2020): 199–207. https://direct.mit.edu/dint/article/2/1-2/199/10013/Licensing-FAIR-Data-for-Reuse.

Lessig, Lawrence. "The Creative Commons." In *Copyright Law*, edited by Benedict Atkinson, 335–347. London: Routledge, 2017.

Local Development Research Institute (LDRI). "AI Training Dataset Sustainability Toolkit." 2023. https://developlocalorg.github.io/ml-toolkit.

Loren, Lydia Pallas. "Building a Reliable Semi-Commons of Creative Works: Enforcement of Creative Commons Licenses and Limited Abandonment of Copyright." *George Mason Law Review* 14 (2006): 271.

Mahelona, Keoni. "A Practical Guide to Creating Your Own Stewardship License." 2024. https://blog.papareo.nz/a-practical-guide-to-creating-your-own-stewardship-license.

Maracke, C. "Creative Commons International: The International License Porting Project – Origins, Experiences, and Challenges." *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 1 (2010): 4–18.

McMahon, Aisling. "Accounting for Ethical Considerations in the Licensing of Patented Biotechnologies and Health-Related Technologies: A Justification." In *Patenting Biotechnical Innovation*, edited by Naomi Hawkins, 163–195. Cheltenham: Edward Elgar, 2022.

Miller, Paul, Richard Styles and Tom Heath. "Open Data Commons: A License for Open Data." *LDOW* 369 (2008). http://events.linkeddata.org/ldow2008/papers/08-miller-styles-open-data-commons.pdf.

Moody, Glyn. "A Blatant No from a Copyright Holder Stops Vital Linguistic Research Work in Africa." 2023. https://walledculture.org/a-blatant-no-from-a-copyright-holder-stops-vital-linguistic-research-work-in-africa.

Mysoor, Poorna. *Implied Licences in Copyright Law*. Oxford: Oxford University Press, 2021.
Ncube, Caroline and Isaac Rutenberg. "Intellectual Property and Fourth Industrial Revolution Technologies." In *Leap 4.0: African Perspectives on the Fourth Industrial Revolution*, edited by Zamazima Mazibuko-Makena and Erika Kraemar-Mbula, 393. MISTRA, 2021.
Okorie, Chijioke. "Beyond Intellectual Property Protection: Other AI IP Strategies for the African Context." In *Artificial Intelligence and the Law in Africa*, edited by Caroline Ncube, Desmond Oriakhogba, Tobias Schonwetter and Isaac Rutenberg, 153–174. Johannesberg: LexisNexis, 2023.
Okorie, Chijioke and Vukosi Marivate. "How African NLP Experts are Navigating the Challenges of Copyright, Innovation, and Access." 2024. https://carnegieendowment.org/research/2024/04/how-african-nlp-experts-are-navigating-the-challenges-of-copyright-innovation-and-access?lang=en.
Okorie, Chijioke and Omino Melissa. "Licensing African Datasets." https://licensingafricandatasets.com.
Rahmatian, Andreas. *Copyright and Creativity: The Making of Property Rights in Creative Works*. Cheltenham: Edward Elgar, 2011.
Sadiq, Shazia and Marta Indulska. "Open Data: Quality Over Quantity." *International Journal of Information Management* 37, no 3 (2017): 150–154. https://doi.org/10.1016/j.ijinfomgt.2017.01.003.
Schindler, Sirko, Marta Dembska, François Bonnarel, Mark Lacy, Michael Johnson and Vandana Desai. "Data Documentation Beyond Provenance: Metadata, Research Data Management, FAIR Principles." In *Data-Intensive Radio Astronomy: Bringing Astrophysics to the Exabyte Era*, edited by Eleni Vardoulaki, Marta Dembska, Alexander Drabent and Matthias Hoeft, 397–418. Cham: Springer, 2024.
Senftleben, Martin. "Generative AI and Author Remuneration." *International Review of Intellectual Property and Competition Law* 54, no 10 (2023): 1535–1560. https://doi.org/10.1007/s40319-023-01399-4.
Shackelford, Scott J. and Andrew H. Raymond. "Building the Virtual Courthouse: Ethical Considerations for Design, Implementation, and Regulation in the World of ODR." *Wisconsin Law Review* 3 (2014): 615. https://wlr.law.wisc.edu/wp-content/uploads/sites/1263/2014/07/3-Shackelford-Raymond-Final-Online.pdf.
Siminyu, Kathleen, Jade Abbott, Kọ́lá Túbọ̀sún, Aremu Anuoluwapo, Blessing K. Sibanda, Kofi Yeboah, David Adelani et al. "Consultative Engagement of Stakeholders Toward a Roadmap for African Language Technologies." *Patterns* 4, no 8 (2023): 100820. https://doi.org/10.1016/j.patter.2023.100820.
Szkalej, Kacper and Martin Senftleben. "Generative AI and Creative Commons Licences: The Application of Share Alike Obligations to Trained Models, Curated Datasets and AI Output." *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 15 (2024): 313. https://www.jipitec.eu/jipitec/article/view/415/417.
Te Mana Raraunga. "Principles of Māori Data Sovereignty Brief #1." 2018. https://www.otago.ac.nz/__data/assets/pdf_file/0014/321044/tmr-maori-data-sovereignty-principles-october-2018-832194.pdf.
Theriault, Claire. "First, Do No Harm: Five Years of Book-Industry Data Sharing with BookNet Canada SalesData." 2010. https://summit.sfu.ca/_flysystem/fedora/sfu_migrate/11493/etd6339_CTheriault.pdf.
Truong, Hong-Linh, G.R. Gangadharan, Marco Comerio, Schahram Dustdar and Flavio De Paoli. "On Analyzing and Developing Data Contracts in Cloud-Based Data Marketplaces." In *2011 IEEE Asia-Pacific Services Computing Conference*, 174–181. IEEE, 2011.
Wairegi, Angeline and Melissa Omino. "Just Because You Can Doesn't Mean You Should: The Case for and Against Utilisation of Artificial Intelligence for Indigenous Knowledge Curation and Preservation in Kenya." In *Artificial Intelligence and the Law in Africa*, edited by Caroline Ncube, Desmond Oriakhogba, Tobias Schonwetter and Isaac Rutenberg, 189–204. Johannesberg: LexisNexis, 2023.
West, Allen. "Little Victories: Promoting Artistic Progress Through the Enforcement of Creative Commons Attribution and Share-Alike Licenses." *Florida State University Law Review* 36 (2008): 903. https://ebin.pub/discrimination-copyright-and-equality-opening-the-e-book-for-the-print-disabled-cambridge-disability-law-and-policy-series-1107119006-9781107119000.html.
Yuille, LaKisha K. "Inequity as a Legal Principle." *University of Kansas Law Review* 66 (2017): 859.

Legal Citations

Blind SA v Minister for Trade, Industry and Competition and Others (2022) ZACC 33.
Thomas v Sorrell (1672) Vaugh 330.