Article

Journal of Business and Technical Communication
1-15

DOI: 10.1177/10506519241280587
journals.sagepub.com/home/jbt

Technical Communication's Fight Against Extractive Large Language Modeling by Applying FAIR and CARE Principles of Data

Chris A. Lindgren[1], Erin Yunes[2], and Cana Uluak Itchuaqiyaq[3]

Abstract

This article assesses the data practices of Grammarly, the prominent AI-assisted writing technology, by applying data principles that advocate for empowering Indigenous data sovereignty. The assessment is informed by the authors' work with an Inuit tribal organization from rural Arctic Alaska that generated data and metadata about potentially sacred tribal activities. Their analysis of Grammarly's large-language modeling practices demonstrates how technical communication can hold businesses to

[1]Department of English, North Carolina State University, Raleigh, NC, USA
[2]Virginia Tech Libraries, Blacksburg, VA, USA
[3]Department of English, Virginia Tech, Blacksburg, VA, USA

Corresponding Author:
Chris A. Lindgren, Department of English, North Carolina State University, 131D Tompkins Hall, 2211 Hillsborough St, Raleigh, NC 27607, USA.
Email: calindgr@ncsu.edu

principled data practices created by Indigenous nations and communities that understand how to create more just futures.

Keywords
Indigenous data sovereignty, community engagement, social justice, large language modeling, AI-assisted technology

The web runs on an extractive capitalism that has maintained historical forms of oppression (Chun, 2021; Noble, 2018) and power inequities (Haas, 2012; Kong & Ding, 2024; McMillan Cottom, 2020). Web sites and applications convert personal user information and labor into assets as data to be used and sold with little-to-no regulation or equitable beneficial return to users beyond the use of services (Birch et al., 2021). Because AI-assisted writing technologies are part of this legacy, we ask, what critical tools can technical and professional communication (TPC) use to advocate for users by identifying and redressing extractive design practices of AI-assisted technologies that are inequitable? In this article, we argue that TPC is well-positioned to fight against power inequities by applying data principles that advocate for Indigenous data sovereignty (IDSov) in the context of AI-assisted technologies.

To address these power inequities, we extend existing calls for applying TPC's expertise in critical data (Atherton, 2021; Frith, 2017; Lindgren, 2024) and advocating for socially just language in large language models (LLMs; Byrd, 2023; Graham & Hopkins, 2022; Owusu-Ansah, 2023). Specifically, TPC has reframed *data* as processes of communication in order to demonstrate how data collection and uses are never neutral but instead enact the values, goals, and outcomes of the collectors. On the modeling front, TPC has begun to call for including historically marginalized groups to play a leading role in the design and control of LLMs. As Ghanian English scholar Owusu-Ansah (2023) has maintained, such groups are always "historically defined and not a definer" (p. 144). Byrd (2023) agreed, positing that marginalized peoples' participation in LLM development will likely be relegated to the use of AI technologies (p. 140). This extractive and colonizing relationship between companies and users is currently a feature of AI technologies, not a bug, so TPC must advocate to redress such power inequities.

We extend such calls for advocacy work in TPC by applying the data principles of FAIR (findable, accessible, interoperable, and reusable) and

CARE (collective benefit, authority to control, responsibility, and ethics) to assess the LLM data practices of the AI-assisted writing technology Grammarly. We define these principles and then support our analysis with insights drawn from our experiences conducting IDSov work during a project that is designed by, led by, and serves an Indigenous community. Our analysis of Grammarly's data and LLM practices shows how technical communication can hold businesses to principles created by Indigenous nations and communities that understand how to create more just futures.

Grammarly and Indigenous Data Sovereignty

Valued at $13 billion, Grammarly boasts over 30 million users, 3,000 educational institutions, and 50,000 teams logging on daily (Lytvyn & Erlichman, 2023). Grammarly's (2023) mission is "to improve lives by improving communication" by helping users to revise and correct grammar in their writing. Leveraging generative AI, it now claims to expedite the writing process by classifying tone or dialect inconsistencies and suggesting revisions that align with the preset writing goals and target audience. These AI features rely on Grammarly's LLM, which collects users' writing as data both within their word processor app and within any other approved external app in order to span users' writing technologies in general-purpose web browsers and word processors. Considering the extensive ecosystem of access that Grammarly has created to collect users' information at the individual level, we analyze the relationship between Grammarly and its users to tease out power inequities created by its data policies and practices: what user information is being captured, how it is used, who has access to it, and how it is processed.

IDSov is one lens for understanding imbalances in commercial data extraction, policy, and use practices that occur from participating in AI software like Grammarly. *IDSov* is the right of Indigenous people to own and control the data emerging from their communities, lands, and so forth, for which "Indigenous Peoples should have the right to 'opt into data structures' that support their aspirations, priorities and values, and decline to participate in data processes that do not" (Tsosie, 2021, p. 207). IDSov opposes current standard protocols limited to the "simple choice" to either "opt in" or "opt out" of data structures for collecting, processing, and storing data. The Global Indigenous Data Alliance (GIDA, 2023) offers FAIR and CARE to TPC as methodologies for advocating for and making changes that empower IDSov.

The FAIR and CARE Frameworks

FAIR is a heuristic to streamline data sharing and reuse by emphasizing the technical usability of data (Wilkinson et al., 2016). GO FAIR (2023) defined *technical usability* as "machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention)" (para. 1). The FAIR creators intended to "[bring] some clarity around the goals and desiderata of good data management and stewardship, and [define] simple guideposts to inform those who publish and/or preserve scholarly data" (Wilkinson et al., 2016, p. 1). But unfortunately, FAIR unknowingly perpetuates extractive data-mining practices.

CARE's exigence, however, according to the GIDA (2023), is to temper the voracity of FAIR in light of continuing power imbalances and unethical scholarly practices related to Indigenous data. CARE was developed by Indigenous global stakeholders in November 2018 (GIDA, 2023) to assess and guide the responsible and equitable collection and stewardship of Indigenous data. It responds to the widespread publication and use of FAIR governance of open data that ignore "power differentials and historical contexts" of data (GIDA, 2023, para. 1), especially in light of the history of extractive data practices affecting Indigenous communities.

Because of their shared focus on the good stewardship of data, FAIR and CARE are often combined in discussions of IDSov to promote community access to, control of, and benefits from Indigenous data. As Figure 1 illustrates, FAIR and CARE together reveal issues of access, usability, power,

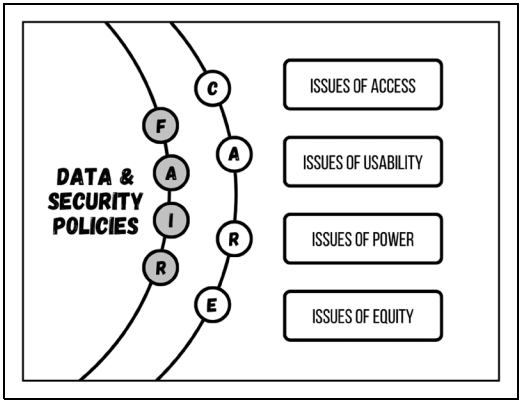


Figure 1. Data and security policies analyzed through the lenses of FAIR and CARE reveal issues of access, usability, power, and equity.

and equity in data and security policies. Our FAIR–CARE examination of user agency to "opt in" and "opt out" and participate in the development of Grammarly's data collection and modeling reveals areas of caution that users of AI software, especially such software that is suggested or assigned in educational or professional contexts, should consider.

We couple CARE with our work with an Inuit tribal organization from rural Arctic Alaska in a community-engaged research (CER) project that generates data and metadata about multifaceted, and potentially sacred, tribal activities. As part of this work, we are establishing a data management plan that is aligned with tribal policies and IDSov. In working with our tribal partner, the option to use generative AI technologies to assist with CER has surfaced in multiple team meetings, especially in relation to daunting yet crucial technical writing tasks, such as grant writing. The members of our CER team come from backgrounds representing both academic privilege and educational disparities, especially regarding technical writing. To help all team members work together on technical writing tasks from positions of strength, our team seriously considered the use of generative AI technologies, such as Grammarly. But as we considered these technologies alongside trying to establish a rigorous data management plan in light of IDSov, we recognized that we needed to analyze how our project's (and team's) data would be affected and potentially compromised by such technologies.

In the following sections, we use the FAIR and CARE frameworks to evaluate how Grammarly's policies and practices impact IDSov within the community setting, identifying areas for enhancing policy transparency and spotlighting power inequities in the company's governance model. Through this analysis, we provide crucial insights into concerns related to access, privacy, power, and equity. Specifically, we use the categories of FAIR and CARE to organize our analysis of various aspects of user agency within Grammarly's policies and practices. The goal of this analysis is to counter the extractive design paradigm by applying data principles that champion IDSov. As we examine Grammarly's LLM practices, we illustrate how these data principles can be used as tools to confront Western extractive norms and values and how generative AI businesses can contribute to a more equitable future by respecting mechanisms of IDSov.

FAIR

The three main principles of the FAIR framework are that data should be (a) findable and accessible, (b) interoperable, and (c) reusable. In this section, we elaborate on each principle and apply them to assess Grammarly's data practices.

Findable and Accessible

FAIR's principle that data must be findable and accessible coincides with the challenge in technical communication for the public to understand or interpret the LLMs in AI technologies such as Grammarly. The origins, processing, and applications of data in these LLMs remain inaccessible, which prevents others from verifying the source and integrity of the data, identifying biases, or ensuring the reliability and trustworthiness of commercialized AI. Companies like Grammarly have a vested interest in ensuring that their data are inaccessible so a product cannot be reproduced. Indeed, AI companies often limit or block computational systems from finding their proprietary data.

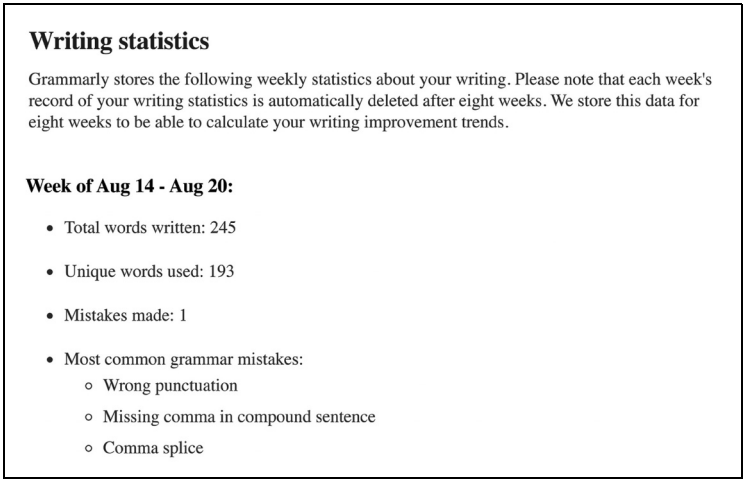
At the user level, finding and accessing user content data that the application has collected can be a difficult, time-consuming process. First, a user must navigate through the application's Terms of Services (ToS) to find what data have been collected and held by the company, where they can find their data, and how they can access the data. Some policies are prominently displayed and accessible on the company's website, but Grammarly has at least a half-dozen important policies distributed across its site with no search feature at hand. These ToS issues obfuscate users from knowing their data rights—a long-standing power problem between companies and users (Böhme & Köpsell, 2010; Fiesler et al., 2016). If businesses and individuals want to know more about the specific data that are sampled and used, they must request personal data reports from Grammarly; however, Grammarly only provides aggregated summary statistics in a PDF or text file (see Figure 2). Overall, Grammarly does not provide users with findable and accessible data or a means to understand the scope and scale of what user data it collects and retains.

Interoperable

Data are interoperable if they can be integrated within other technological systems and used by others for their goals. The data referred to in Grammarly's accessible user summary are not interoperable because the file's information is unstructured. Grammarly could instead offer a structured data set in interoperable file formats, such as CSV (comma-separated values). These data are interoperable because the provided values are not an agreed on provision that users get to define.

Reusable

The prior two principles of the FAIR framework ensure that data can be reused by people who did not create the data. The principle of reusable



Writing statistics

Grammarly stores the following weekly statistics about your writing. Please note that each week's record of your writing statistics is automatically deleted after eight weeks. We store this data for eight weeks to be able to calculate your writing improvement trends.

Week of Aug 14 - Aug 20:

- Total words written: 245
- Unique words used: 193
- Mistakes made: 1
- Most common grammar mistakes:
  - Wrong punctuation
  - Missing comma in compound sentence
  - Comma splice

Figure 2. Partial example of Grammarly's personal data report.

data assumes an open sharing of disaggregated data, so Grammarly's business values directly conflict with it. This principle also assumes a closer relationship between invested parties because the metadata of reusable data strive to document the data's exigencies and common goals (GO FAIR, 2023). If users from an Indigenous community wanted to better understand what data are collected and used by Grammarly's LLM, the company's aggregated per-document statistics, weekly email reports, and personal email reports do not meet the standards of the reusable principle. In sum, Grammarly's data practices obfuscate what Indigenous knowledges and practices they are collecting and using to develop their LLM without building the capacity to foster more equitable partnerships between the company and the community.

CARE

The FAIR principles begin to illuminate how Grammarly adopts an extractive paradigm of user content data because users can see only summarized data about their use of Grammarly's services. But FAIR focuses on general appeals to openness and fairness without considering what historical issues and contexts warrant careful consideration; therefore CARE, which attends to these historical issues and contexts, is FAIR's important heuristic partner. But before we apply the CARE principles to Grammarly, we must

first remember that IDSov is specifically designed to support Indigenous peoples' self-determination of their data use. We use this framework not to co-opt IDSov for mainstream, academic, and corporate discussions of data; instead, we turn to IDSov as a teaching tool for considering how deeply relational data work is and how it impacts both the data's quality and people at a community and group level.

The GIDA clarifies CARE principles by breaking each of them down into three subprinciples (Carroll et al., 2022). Although some subprinciples easily transfer to non-Indigenous contexts, some do not. We aim to respect the boundaries of IDSov as we apply CARE here as a means to confront Western extractive norms and values that companies such as Grammarly adopt and perpetuate.

Collective Benefit

Collective benefit puts the onus on companies to adopt a different relationship with Indigenous communities and nations. This principle compels companies to ensure that they work with Indigenous communities to develop an equitable and agreed-on set of benefits defined by the communities. We use the following three subprinciples—inclusive development and innovation, improved governance and citizen engagement, and equitable outcomes (Carroll et al., 2022)—to assess Grammarly's modeling practices.

C1. Inclusive Development and Innovation. Carroll et al. (2022) suggested that research teams consider how their data protocols "actively support the use and reuse of data by Indigenous nations and communities" (p. 4). Grammarly maintains almost total access and control over a significant portion of user data, and users play little to no active role in creating and sharing reusable and interoperable data. Instead, Grammarly tasks teams to create what company linguists Ng and Nahorna (2022a) called "gold data" by hiring data labelers to tune their models. In describing this complex structural problem of extractive design and power, we do not mean to single out any particular Grammarly employee; rather, we cite Ng and Nahorna's findable and accessible information as representing Grammarly's modeling practices. Grammarly does not transparently communicate its collective and community approaches when explaining how potential users benefit from its capture of user data to develop its LLM. This issue conflicts with IDSov: Indigenous self-determination in relation to data.

C2. For Improved Governance and Citizen Engagement. This subprinciple asks how the data should inform and improve the relationships between user communities and institutional and commercial entities, such as

Grammarly. Because Grammarly's modeling centers on the more extractive work conducted by internal teams, individual users can only decide to use the product and generate personal data reports on user content. This relationship seems strange and troubling if we consider that Grammarly has access to vast amounts of highly personal, private user content data that it shares with other entities, such as external service providers (Grammarly, 2023).

C3. For Equitable Outcomes. Prospecting "gold data'' from users reinforces extractive practices of white-settler colonialists that steal land and its resources for the benefit of the few privileged enough to mine and model such data at a global scale. Based on the available information, Grammarly policies do not support IDSov, and its broad capture and use of user data impede data equity efforts.

Authority to Control

When users sign up for Grammarly, they relinquish control over how their data are leveraged by Grammarly and who has access to it. Grammarly is authorized to use user content for limited purposes, including refining its products, training the AI model, introducing new features, and promoting new services (Grammarly, 2023). As users employ the app on their desktop or browser plugins, Grammarly has access to emails, text messages, and documents—both individual and shared. When users agree to these terms, they may be sharing not just their own information with the company but potentially the data of their friends, family, and community.

A1. Recognizing Rights and Interests. Grammarly addresses the protection of personal data in its policies. Yet Grammarly appears to have a broader latitude in how it uses data sourced from user content. Grammarly's removal of personal identifying information (PII) from user content data can still result in misconceptions or improper use. Such considerations become even more vital for Indigenous and local communities, whose data might be of a sensitive or culturally sacred nature and have significance not meant to be shared with others outside of specific contexts.

A2. Data for Governance. Carroll et al. (2020) stated that "Indigenous Peoples *must* [emphasis added] have access to data that support Indigenous governance and self-determination" (p. 6). Grammarly does not permit users to gain access to the data that they extract from their use. While much of our analysis has centered on Grammarly capturing user data to train its LLM, it is important to note that data about seemingly mundane tasks, such as how users at tribal organizations write emails in intercultural and intracultural ways, can be useful for tribes to understand

and analyze their own communication practices. Thus, capturing this data with no means for users to access and use it for their own self-determined goals is harmful.

A3. Governance of Data. In the same way that an LLM can train in a dialect by analyzing a large data set, which often comprises scraped data from the web, and pulling out linguistic patterns, an LLM can pull knowledge and information from regional text inputs, which could perpetuate harm to communities. Yet there is no communicated method to discover how Grammarly uses user content and who is using it. All users must permit Grammarly to store and display their content, as well as modify and develop new versions of it in order to improve and market their products. This authority is also extended to Grammarly's (2023) external service providers. Overall, users waive their authority to control their content, and their prospective data "only used in aggregate" can still be harmful if not tempered by CARE-informed governance policies.

Responsibility

Responsibility bridges the previously applied FAIR and CARE principles. It emphasizes relationships and requires that those who hold positions of power enact a culturally humble approach to data practices. As part of our own CER project, we defined a cultural humility framework (CHF) for community research partnerships (Itchuaqiyaq et al., 2023), which we also apply to commercial research and development. The CHF aligns with the responsibility principle of CARE because it calls for entities like Grammarly to understand how their scaling efforts are built on the imbalanced power dynamic between themselves and the communities from which they extract data.

R1. For Positive Relationships. Cultural humility promotes positive relationships. Mutual trust requires sharing data and insights and submitting to users' authorial rights, needs, and direction. As previous FAIR and CARE evidence demonstrates, Grammarly maintains a position of power over user communities rather than imagines equitably balanced authorial control by submitting to, or even sharing leadership with, users. Taking a humble approach to modeling is necessary to repair the historical disrespect of Indigenous sovereignty in research and data work.

R2. For Expanded Capability and Capacity. Since Grammarly controls access to all disaggregated data and LLMs, the company does not responsibly facilitate Indigenous self-determination. If Grammarly's mission "improves lives," then users and communities must lead them. Users,

such as the tribal members from our CER team, are interested in tools such as Grammarly to expand their own capability and capacities as researchers. But using these tools would put tribal knowledges at risk. Our tribal partners represent sovereign Indigenous nations, and for them, the benefit of easily writing grammatically correct sentences does not outweigh the cost of tribal sovereignty.

R3. For Indigenous Languages and Worldviews. Our CER project supports Indigenous languages and worldviews, and if the use of generative AI were ethical, it could assist our team with those efforts. But Grammarly's issues of access cannot foster responsible relationships. Our CER project supports Indigenous languages and worldviews, and the use of ethical generative AI could assist our team with those efforts. Yet, again, the risks to IDSov outweigh the potential benefits of using its tools. Cultural humility recognizes the responsibility that Grammarly has as stewards, not owners, of Indigenous knowledges and language practices.

Ethics

CARE defines an ethic in which Indigenous Peoples "inform the use of data across time in order to minimize harm, maximize benefits, promote justice, and allow for future use" (Carroll et al., 2022, p. 2). Grammarly (2023) states that it makes decisions about its services based on a set of ethical principles dubbed as EAGER: Ethical, Adaptable, Gritty, Empathetic, and Remarkable. Although the company attempts to define these values, the language is vague and highly interpretive based on a person's background, perspectives, and experiences. As Edenfield (2018) discussed, companies develop vague values and mission statements (Lytvyn, 2021) to construct a uniform culture that inspires order but remains flexible enough so that the companies can adapt to corporate challenges and expand. This strategy is employed by Grammarly. A Grammarly business statement (Lytvyn, 2021) states that the EAGER values underscore every business decision the company undertakes, but the ambiguity of the statements makes it difficult for the company to develop more actionable ethical data practices. Conversely, the FAIR and CARE principles can remedy this difficulty.

Considering Grammarly's operations, the responsibility for data ethics extends beyond individual employees and must be reflected in the corporation's policies and procedures. As Walwema et al. (2022) emphasized, "Ethical frameworks help us understand how to enact justice and identify the behaviors, actions, and policies that should be considered just or unjust" (p. 259). Grammarly claims that EAGER informs its mission to

foster inclusive and learning-centric environments for its team, offer empathetic user support with integrity, and create products that connect people. But as we will now discuss, users pay for a service that also collects and commodifies their knowledge and language practices in order to realize Grammarly's goals, definitions and methods of writing, and monetary gain.

E1. For Minimizing Harm/Maximizing Benefit. This subprinciple demands that businesses never define Indigenous data as deficient or trapped in a Western colonialist historical narrative. Using Grammarly and data labeling are existing means for users to develop an LLM. This transactional relationship, however, does not minimize harm or maximize benefits for communities. Instead, it serves to prospect Grammarly's gold data in a way that minimizes its costs and maximizes its profits.

E2. For Justice. For LLMs to potentially redress colonial power imbalances, Grammarly must be led by Indigenous Peoples. Whether community representatives are sought after as leaders of LLM development is unclear and unlikely, but Grammarly's Ng and Nahorna (2022b) do recognize that "everyone has their biases" and discuss their approach to annotation practices that they claim will "eliminate bias" (31:45). They suggest that modelers can "ensure diversity" by "hiring a large enough number of annotators." Yet by "diversity" they do not mean a diverse population; rather, they are referring to diverse labeling tasks, such as not tasking labelers with correcting a single type of grammatical mistake. But as long as Grammarly teams retain control of the process, just LLMs cannot be developed.

E3:.For Future Use. Regardless of its EAGER principles, Grammarly falls short in truly minimizing harm and maximizing benefits for any users, let alone Indigenous peoples in relation to CARE, because it does not allow user access and control over what content is being sampled to enhance the product. For a just future, regulations and policies must replace the prospecting of gold data with CER practices that take up FAIR and CARE principles.

Conclusion: Advocating for FAIR and CARE-Ful LLMs

Grammarly, as a prominent player in the AI-driven landscape of language technology, serves as a generative example that illustrates how TPC can align with the principles established by both scientists and Indigenous nations and communities in order to support more equitable and just data futures. This analysis of Grammarly's LLM practices has demonstrated the use of the FAIR and CARE principles as an evaluation tool for navigating the complex terrain of data sovereignty. Although the FAIR and CARE principles were designed to provide guidance in establishing data protocols,

their use as an evaluative tool can help technical communicators and users articulate concerns about data sovereignty issues in order to advocate for more equitable, responsible, ethical, and equitable data practices and hold businesses and organizations accountable.

Technical communicators can partner with organizations such as the GIDA to hold businesses accountable for prospecting gold data. Grammarly is valued at $13 billion, and unfortunately, in corporate minds, the collective benefit of user content data is wrongfully limited to the usability of Grammarly's generative AI tool. Yet, as Carroll et al. (2020) stated, "Being CARE Full is a prerequisite for equitable data and data practices" (p. 8) because data work is deeply relational work. Currently, businesses that are prospecting gold data for their LLMs are building tall borders to block socially just relationships for change. Thus, we call for technical communicators to advocate for a different global web landscape, one in which businesses are held to an ethic beyond "if you don't like it, don't use it." FAIR and CARE together are one tool for making such a change.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Atherton, R. (2021). Missing/unspecified": Demographic data visualization during the COVID-19 pandemic. *Journal of Business and Technical Communication*, *35*(1), 80–87. https://doi.org/k876

Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by big tech. *Big Data & Society*, *8*(1), 1–15. https://doi.org/10.1177/20539517211017308

Böhme, R., & Köpsell, S. (2010). Trained to accept? A field experiment on consent dialogs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2403–2406. https://doi.org/bvx538

Byrd, A. (2023). Truth-telling: Critical inquiries on LLMs and the corpus texts that train them. *Composition Studies*, *51*(1), 135–142.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe,

R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, *19*(43), 1–12. https://doi.org/10.5334/dsj-2020-043

Carroll, S. R., Garba, I., Plevel, R., Small-Rodriguez, D., Hiratsuka, V. Y., Hudson, M., & Garrison, N. A. (2022). Using indigenous standards to implement the CARE principles. *Frontiers in Genetics*, *13*, 1–10. https://doi.org/k877

Chun, W. H. K. (2021). *Discriminating data*. MIT Press.

Edenfield, A. C. (2018). The burden of ambiguity: Writing at a cooperative. *Technical Communication*, *65*(1), 31–45. https://www.ingentaconnect.com/content/stc/tc/2018/00000065/00000001/art00005

Fiesler, C., Lampe, C., & Bruckman, A. S. (2016). Reality and perception of copyright terms of service for online content creation. *In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1450–1461). https://dl.acm.org/doi/10.1145/2818048.2819931

Frith, J. (2017). Big data, technical communication, and the smart city. *Journal of Business and Technical Communication*, *31*(2), 168–187. https://doi.org/gntw69

GIDA (2023, January 23). *CARE principles for Indigenous data governance*. https://www.gida-global.org/care.

GO FAIR (2023). *FAIR Principles*. https://www.go-fair.org/fair-principles/.

Graham, S. S., & Hopkins, H. R. (2022). AI For social justice. *Technical Communication Quarterly*, *31*(1), 89–102. https://doi.org/gn34vs

Grammarly (2023). *Terms of service*. https://www.grammarly.com/terms.

Haas, A. M. (2012). Race, rhetoric, and technology. *Journal of Business and Technical Communication*, *26*(3), 277–310. https://doi.org/gfxpfc

Itchuaqiyaq, C. U., Lindgren, C., & Kramer, C. (2023). Decolonizing community-engaged research. *Communication Design Quarterly*, *11*(3), 12–20. https://doi.org/k875

Kong, Y., & Ding, H. (2024). Tools, potential, and pitfalls of social media screening: Social profiling in the era of AI-assisted recruiting. *Journal of Business and Technical Communication*, *38*(1), 33–65. https://doi.org/k874

Lindgren, C. A. (2024). A stasis network methodology to reckon with the rhetorical process of data: How a data team qualified meaning and practices. *Technical Communication Quarterly*, 1–26. Advance online publication. https://doi.org/10.1080/10572252.2024.2306259

Lytvyn, M. (2021, June 25). *Leveraging corporate values to drive business growth and scalability*. Grammarly Business. https://www.grammarly.com/business/learn/leveraging-corporate-values-to-drive-business-growth-and-scalability/.

Lytvyn, M., & Erlichman, J. (2023, July 31). *We're ready to go public but don't see an immediate need to do so: Grammarly co-founder*. BNN Bloomberg. https://www.bnnbloomberg.ca/video/~2736916.

McMillan Cottom, T. (2020). Where platform capitalism and racial capitalism meet. *Sociology of Race and Ethnicity*, *6*(4), 441–449. https://doi.org/ghd79s

Ng, L., & Nahorna, L. (2022a, October 21). *Annotation best practices for building high-quality datasets*. Grammarly. https://www.grammarly.com/blog/engineering/annotation-best-practices/.

Ng, L., & Nahorna, L. (2022b, May 22). *Grammarly AI-NLP Club #15: How to collect high-quality data to power your machine learning systems [Video]*. Grammarly Ukraine. https://www.youtube.com/watch?v=Z8yqJz-Iq3s.

Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.

Owusu-Ansah, A. L. (2023). Defining moments, definitive programs, and the continued erasure of missing people. *Composition Studies*, *51*(1), 143–148. https://www.proquest.com/scholarly-journals/defining-moments-definitive-programs-continued/docview/2841570271/se-2

Tsosie (2021). The legal and policy dimensions of indigenous data sovereignty. In M. Walter, T. Kukutai, S. Carroll Raine, & D. Rodriguez-Lonebear (Eds.), *Indigenous data sovereignty and policy* (pp. 204–225). Routledge.

Walwema, J., Colton, J. S., & Holmes, S. (2022). Introduction to special issue on 21st-century ethics in technical communication: Ethics and the social justice movement in technical and professional communication. *Journal of Business and Technical Communication*, *36*(3), 257–269. https://doi.org/k873

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/bdd4

Author Biographies

Chris A. Lindgren, assistant professor at North Carolina State University, studies coding and data-driven work as rhetorical communication. He is a partner in The Rematriation Project and founded the Special Interest Group on the Writing and Rhetoric of Code.

Erin Yunes is a CLIR Postdoctoral Fellow in Community Data for the Rematriation Project at the Virginia Tech University Libraries. Her research focuses on broadband equality and access to digital cultural resources.

Cana Uluak Itchuaqiyaq (she/they) is an Iñupiaq tribal member of the Noorvik Native Community in northwest Alaska. She is an assistant professor of technical and scientific communication and the cofounder of the upcoming Equitable Research Center at Virginia Tech.