The Politics of AI Systems are Inextricable from Their Supply Chains: Public Values Versus the Digital Political Economy

Ben Gansky
Arizona State University, School for the Future of Innovation in Society
bengansky@gmail.com

Abstract

This article narrates the trajectory of a conflictual relationship between a public interest collective and their AI technology vendor. Through an analysis of three years of email correspondence obtained via Freedom of Information Act request, I follow a team of public planners from the U.S. state of Oregon who insisted that their vendor and its technology be transparent and accountable; conflicts emerged. I argue that this case demonstrates fundamental limits on the ethics and contextual appropriateness of AI tools that are dependent on data supply chains entangled with the digital political economy. These limits are enacted by such legal mechanisms as trade secrecy protections and non-disclosure agreements, as well as by the structural complexity and recursiveness of AI/ML model and data supply chains. Negotiations over the politics of the digital tool in question became articulated as conflicts over the provenance of the tool's training data. By calling attention to how digital production relations are always situated within chains of dependencies, my analysis yields a more nuanced understanding about how the politics of AI-based tools are shaped in practice, and the terrain on which they might be contested and attempts made at reconfiguration and alignment with public values.

Introduction

This article narrates the trajectory of a conflictual relationship between a consortium of public planners and their one-time AI technology vendor. Scholarship on the political economy of digital technologies for public planning wrestles with how best to understand the outsize power vested in private technology providers — and what hazards such concentrations of technopolitical agency may pose to democratic societies. Taylor (2021, 898) identifies "problems arising from contracting… where commercial firms develop capacity that is complementary to that of the state and then become incorporated as contractors in the state's operations while still retaining a private-sector identity…" whereupon competing priorities emerge between private profit and public values (Fourcade & Gordan, 2020) and public value is captured for private purposes (Bates, 2012; Collington, 2019). Brauneis and Goodman (2018, 119) characterize the adoption of privately-developed algorithmic systems as a depoliticization strategy, either by 'objectivity-washing' or by concealing surface area for political decision-making behind the screen of 'trade secrets' so that "the politics of algorithms recede behind private hedges". The delegation of state functions to private technology providers is argued to reflect and amplify neoliberal agendas of intentional hollowing-out of state capacity (Collington, 2021; Morozov & Bria, 2018; Faulkner-Gurstein & Wyatt, 2023).

Limiting government transparency and accountability through the embedding of private technology providers within state infrastructures in some circumstances may not be a negative externality but instead a core strategic goal (Baykurt, 2022; Richardson, 2019). Even public servants looking to avoid the delegation of their authorities and the obfuscation of their work's politics, however, are at a disadvantage in dealing with private vendors: "Most local governments lack the expertise and wherewithal to deploy data analytics on their own. If they want to be 'smart,' they need to contract with companies, universities, and nonprofits to implement privately developed algorithmic processes… All of this work—the collection, analysis, and use of data—requires technical know-how and infrastructure that most governments lack… " (Brauneis & Goodman, 2018, 107-114).

The case I narrate and the analysis I develop here contributes to the AI and society literature in two ways. First, rather than focus on public employees bamboozled by better-equipped private tech providers, or on neoliberal warriors in bureaucrats' clothes borrowing governance opacity from their digital tools, I follow a team of public planners that is savvy enough to contest and negotiate terms with their private technology vendor, that asserts the priority of transparency for their algorithmic tooling, and that looks to secure back within public bounds the assets and capacity generated by their vendor. Conflicts emerge nonetheless; I argue that what we witness here are fundamental limits on democratic

accountability and transparency for AI tools dependent on data supply chains entangled with the surveillance economy. These limits are imposed by such mechanisms as trade secrecy protections and non-disclosure agreements, as well as by the structural complexity and recursiveness of the surveillance economy's data supply chains (Crain, 2018).

My second contribution is to underline how negotiations over the politics of the digital tool in question become articulated as conflicts over the dependencies of that AI tool: specifically, over the provenance of the tool's training data. I follow the multiple ways in which the planners construe the significance of the tool's data provenance; each iteration is consistent in that they draw attention to the data's (and therefore, the data-dependent tool's) embeddedness within supply chains. To the extent that digital objects embed and reproduce particular ways of knowing and relating (Benjamin, 2019; D'Ignazio & Klein, 2020; Alkhatib, 2021), the digital tools that integrate and rely on those objects risk inheriting their positionalities and modalities. Dependencies in turn have their *own* dependencies (Cox, 2019). Digital objects are enabled by, and recapitulations of, their digital supply chains; AI/ML systems especially (Dhaliwal, 2025). By calling attention to how digital production relations are always situated within chains of dependencies, my analysis yields a more nuanced understanding about how the politics of AI-based tools are shaped in practice, and the terrain on which they might be contested and attempts made at reconfiguration.

As the transit planners in this article assess the fitness-for-purpose of a machine learning-based simulation tool, they confront a truth of the contemporary digital: that both technical capacities and normative qualities are conveyed 'genealogically'.

The Case in Brief

This article emerges from analysis of over 1500 pages of email correspondence, obtained through public records requests, between transit planners in the Portland, Oregon metropolitan region in the United States and employees of the technology startup Replica. The correspondence, dating from mid-2017 to the end of 2020, traces the rise and fall of a pilot project in which Replica was awarded a sole source contract to develop a resident mobility simulation platform for the use of a consortium of planners from the Portland Bureau of Transit (PBoT), the City of Portland, and Metro, a regional planning organization whose leaders are directly elected by area residents. Unless otherwise noted, all quotes from emails among members of the planning group and between the planners and representatives of Replica are taken from the FOIA-obtained records.

Replica originated as a project within Sidewalk Labs, a sibling to Google in the Alphabet corporate octopus, before being spun out into an independent startup company in 2019, albeit with foundational investments and continuing board representation from Alphabet and the former Google chairman Eric Schmidt. Replica's software is based on agent-based models that simulate the mobility patterns of urban populations. What differentiates Replica's offerings from prior generations of agent-based models (which have been used in urban planning processes for generations [Johnson, 2020]) are their use of a wider variety of data sources, including data artifacts generated by the use of smartphones and obtained through telecom providers. As will be seen below, information about where Replica gets its data, and under what conditions the data are obtained, is a fiercely guarded commodity.

Replica trains its machine learning models (input-output hidden Markov models, or IOHMM, to be precise) on its varied data sources in order to produce interactive simulations of populations whose demographic and behavioral attributes they claim faithfully reproduce the aggregate movement patterns of residents in their subject cities — without transmitting information that could compromise the privacy of residents. Replica's claims of better, more comprehensive data that nonetheless protected the privacy of residents made it an alluring vendor for the Portland area planners. The lead representative of the Portland planners consortium was a Metro employee, a 'senior technology strategist' named Eliot Rose. At the outset of the project, Rose framed the partnership with Replica to a colleague by writing, "[o]ur residents are concerned about serious transportation safety and equity issues, and have asked us to get better data so that we can figure out how to best tackle these issues. At the same time, our residents are also increasingly concerned about privacy. This pilot is an opportunity to understand whether there are conflicts between the need for better data and the right to privacy, and if so, how we can address those conflicts".

Over the three years of their partnership, the emails between the Portland planners and Replica took on an increasingly adversarial quality before Rose and the planners finally initiated the termination process for Replica's contract in mid-2020. This article contextualizes and narrates the tactics by which the planners attempted to discern, assess, and mitigate potential conflicts 'between the need for better data and the right to privacy'. Before narrating this relationship, the next section sketches the relevant context for this article's analysis. I describe the myriad dependencies on which the production of machine learning applications are based, and how those dependencies (and the dependencies on which those dependencies depend) 'pass on' both technical qualities and normative characteristics from their contexts of origin.

The Intensity of Inherited Dependencies in Machine Learning

The fundamental dependencies of machine learning models are compute (hardware, most notably/infamously graphics processing units, or GPUs; see [Hwang, 2018]) and training data. Numerous researchers have demonstrated how the characteristics of training datasets surface in the models trained on these data. Training datasets can problematically exhibit, *inter alia,* sampling distortions (over- or under-representation of certain sampled subgroups) (Buolamwini & Gebru, 2018; Koenecke et al, 2024). Pre-processing transformations in training data have been shown to embed, for instance, historical assumptions about racialized physical differences (Braun & Grisson, 2023; Vyas et al, 2020). Large language models, among others, are trained on and reproduce data rife with misogynist, racist, religiously intolerant, homophobic, and other hateful forms of speech and imagery (Birhane et al, 2023; Birhane et al, 2021; Crawford & Paglen, 2019).

As machine learning models are trained on these datasets and then in turn used in, e.g., facial recognition systems, voice transcription services, medical care early warning systems, clinical diagnoses, search engines, and generative AI programs, deficiencies of the training data result in, for instance, higher rates of false positives for dark-skinned individuals in facial recognition ML systems (Buolamwini & Gebru, 2018; Stevens & Keyes, 2021; Fatima et al, 2024), language models "more likely to suggest that speakers of AAE [African American English] be assigned less-prestigious jobs, be convicted of crimes and be sentenced to death" (Hofmann et al, 2024), speech-to-text models more likely to 'hallucinate' when transcribing the speech of those with speech/language disorders (Koenecke et al, 2024), and image generation services prone to render 'person' as light-skinned Western men and to produce oversexualized images of "women, specifically Latin American, Mexican, Indian and Egyptian women relative to other nationalities" (Ghosh & Kaliskan, 2023); see also (Noble, 2018).

The contents of models' training datasets structure the politics of these systems in use. This is well understood; less broadly understood is the fact that training datasets are themselves increasingly curated by models: "… there is a circularity inherent to the authoring of AI training sets. Because they need to be so large, their construction necessarily involves the use of other models, which themselves were trained on algorithmically curated training sets… There are models on top of models, and trainings sets on top of training sets" (Buschek & Thorp, 2024). So while training datasets are dependencies on which ML models rely, pre-existing models are the dependencies on which these dependencies rely. And those pre-existing models have dependencies on *their* training datasets.

What's even more head-spinning is that datasets and models are not static; new reference datasets are released in intervals (such as the Common Crawl dataset) and API-accessed models, such as the ChatGPT and Claude LLMs produced, respectively, by AI industry players OpenAI and Anthropic are subject to continuous, substantive, and secretive changes (see Offert & Dhaliwal [2025], Hockenberry [2021]). These changes are not necessarily for the better: peer-reviewed research has shown that "the behavior of the 'same' LLM service can change substantially in a relatively short amount of time… [with] significant drifts of their performance and behavior over a short time period on a range of different tasks" (Chen et al, 2024, 1).

In instances wherein a widely-used training dataset has come under fire for, e.g., lacking the consent of depicted, identifiable subjects (Harvey & Laplace, 2019), or for hosting hundreds of instances of known child sexual abuse material (CSAM) (Thiel, 2023) that dataset might be rescinded by its curators/distributors. However, researchers have demonstrated that once training datasets are made available for download, and particularly once they have been used to train machine learning models, rescinding or recalling hazardous and unethical datasets is of limited efficacy in halting their proliferation (and the proliferation of their 'descendants') (Peng et al, 2021).

Taking Accountability for Inherited Dependencies

How is a public interest organization to normatively reason and to make decisions about their digital tools, situated as they are in such (supply) chains of dependencies? This is a difficult question — not least because of the expertise required to parse the upstream chain of dependencies for their politics. It is difficult also because visibility up the supply chain is hard to come by. Training dataset opacity is the rule and not the exception (Widder et al, 2024; Gansky & McDonald, 2022). Even if a particular ML application publishes its training dataset (an extraordinarily uncommon occurrence), there is every likelihood that its training dataset (in other words, its key dependency) is itself dependent upon other models and their training datasets, which are very likely opaque.

The strategic use of the designation 'trade secrets' grants the privilege of opacity to software developers, even as many claim their rights to freely use scraped web content (such as the data within Common Crawl) under the banner of 'free use' (Katyal, 2019; White, 2025). If even a single node in the chain of dependencies cuts off visibility of its own dependencies, it does so for *their* dependencies, and so on. Visibility is a pre-requisite — necessary but not sufficient — for accountability.

Amid the pervasive opacity of digital supply chains, "anyone who depends on them may find themselves open to all sorts of unexpected and redirected entanglements" (Hockenberry, 2021, 644). Mitigating the risk of nasty surprises means that 'due diligence' is a complex, supply chain-encompassing endeavor, and one for which context-specific road maps have yet to be codified (Birhane et al, 2024; Sanchez-Graells, 2024).

This article describes the challenging political economic terrain on which a group of public planners, as public interest professionals, attempted to perform adequate due diligence to determine whether a digital tool was fit for purpose. Their attempts at traversing this terrain included demanding specific forms of dependency-visibility, negotiating over performance thresholds, and opening up room for adversarial experimentation on the tool itself. Ultimately, the terrain proved prohibitive: the journey was aborted and its labor declared lost.

Replica: Who & What

City planners have coupled computational techniques with urban data to practice their trade for over half a century (Batty, 2016). Since the 1980s, increasingly complex and sophisticated modeling techniques based on agent-based simulation have only allowed planners to more closely "mirror—rather than explain—the unpredictability of the real world" (Johnson, 2020, 433). Planners' desire to comprehend causality in urban mobility patterns has produced a market for techniques that might explain not just how residents traverse urban space, but why. The goal of making legible the 'hidden state' of residents' motivations for movement has led researchers and planners to embrace machine learning modeling techniques, as well as the incorporation of forms of data previously unexploited by planners, in order to infer these motivations.

Replica, a private software company, proposes that its products are capable of shedding light just so: not just on human mobility but human motivation for moving. Replica offers its platform (also named Replica) as "a fully calibrated, regional-scale, travel demand model offered via software-as-a-service (SaaS)" ("Replica Data Disclosure Documentation", 2019). Replica's customers are mainly government agencies engaged in the tasks surrounding spatial planning, particularly but not exclusively with regards to transit.

The company began as a project incubated inside Sidewalk Labs. Sidewalk Labs, founded in 2015, was a subsidiary of Alphabet, making it a sister company to Google. The Replica team began working together within the Model Lab, a division of Sidewalk Labs. Model Lab began publishing blog posts on the subject of urban transportation modeling in early 2017 (Chim & Ory, 2017). It was shortly thereafter that the Replica group first demoed their platform to the consortium of Portland planners. In April 2018, the Model Lab team went public with Replica, which they advertised at the time as "a user-friendly modeling tool that uses de-identified mobile location data to give planning agencies a comprehensive portrait of how, when, and why people travel in urban areas" (Bowden, 2018).

Replica was created "to support the development of plans for Sidewalk Toronto" (Bowden, 2018) but by October 2018, following increasingly intense public scrutiny of the Sidewalk Quayside project (cf. McDonald & Wylie, 2019), Replica changed course. A spokesperson for Sidewalk Labs stated that "while Replica may be brought to Toronto in the future, 'the Replica team is now focused on developing their model for other cities in the U.S.'" (Oved, 2018).

In the middle of their partnership with the Portland planners, in September 2019, Replica was spun out into a separate corporate entity: "Replica the product has become Replica the company" (Bowden, 2019). Sidewalk Labs retained its connection to Replica with a seat on its board, and former Google chairman Eric Schmidt became a founding investor.

Broadly speaking, Replica uses multiple sources of data for a given region to construct a simulated population whose members proportionally match the demographics of the simulated area. They then assign mobility patterns to each simulated individual. Granular location data are one of the most revealing and potentially hazardous forms of data out there. Researchers have demonstrated how trivial it is to re-identify supposedly anonymized location data and how, in so doing, an individual's movement patterns can be reconstructed to reveal, e.g., the location of their home, workplace, hobbies, place of worship, etc. (Abbas et al, 2014; Brennan et al, 2023; Cyphers, 2022). Replica argues that by creating a synthetic population, they preclude the use of their platform and its data for identifying real individuals and revealing their movement patterns.

Replica (the platform) is a web-based interface for interacting with these populations of synthetic 'personas' generated by the company's modelers. In addition to viewing this population in motion, users can 'zoom in' to view not only simulated individual residents' origin and destination per trip, but the ascribed purpose of each journey: to eat, to shop, to go to work or school, for entertainment. Furthermore, users can simulate the effects of changes to the transit infrastructure in the region, such as the addition of a bus line or bike lane.

Privacy: Problematized & Pluralized

From the outset of their exchanges, both the planners and Replica agreed on one thing: preserving the privacy of residents was of the utmost importance. Among the Portland planners, discussions sparked about whether and how the

Replica project might occasion serious privacy risks from the very start of the project. The planners needed the Portland City Council to grant an ordinance in support of the project, which offered Replica a sole-source contract to develop a Portland version of its mobility simulation platform. Approval was far from assured.

In mid-December 2018, the Portland planners met to plan how to "highlight the benefits of this project while presenting the privacy concerns appropriately in context" to a local reporter doing a story on the project; the resulting story made them feel that "our work… to coordinate talking points paid off". The same day the story was published, the City Council approved the ordinance.

By January, however, press interest in the project was heating up, not cooling off: a reporter from the Wall Street Journal was working on an article that the planners feared might create controversy. Rose suggested that the planners might "[reach] out to the ACLU [American Civil Liberties Union] and potentially other local privacy advocates… to see if they have any privacy concerns that we may want to address in the contract?". Less than a week later, he followed up: "I heard… that the Smart City PDX [the City of Portland planning team involved in the Replica project] heard some concerns from advocates about the Replica project. Would it help at all for us to meet individually with some of those organizations to try and get out in front of their concerns? Our gov't affairs team suggested that when I briefed our council on this, but I defer to you on what's a good idea".

The team agreed to proactively reach out to local advocates, but before they could, investigative journalism outlet The Intercept published an article on Replica (Kofman, 2019) that caused one of the Portland planners to remark that "it's making cities look as if they're blindly entering into these agreements. To add to that, this article raises a new (but not unpredictable) concern that if SL can disaggregate and synthesize, why couldn't someone else un-synthesize and re-aggregate". Rose suggested that any response from the planners should reject any insinuation of naivete: "… we get pitched on other products derived from cell data all the time… the method that SWL [Sidewalk Labs] is using - simulating travel based on the data - is a more compelling way to protect privacy. It's what we've used for years to protect the confidentiality of our travel survey respondents. This article makes it sound like Replica is the first to offer any of this data and the first time we've considered any of these issues. Seems like we need to emphasize our experience here and how it's led us to work w/ Replica". Then, a new chord is sounded.

From Anonymity to Consent

"The issue that the article… raises about whether the data that informs Replica has been obtained with consent is one I want to dig into further. I feel like that's a real concern, and that it's also out of the league of local governments to address… we hear concerns about privacy from our citizens, and at the same time we hear a very strong demand for us to use a data-driven approach to make the streets safer. This pilot with Replica is an opportunity to understand if those goals conflict and if so, how we can balnce [sic] those competing concerns". Rose's colleague quickly replied, "To the consent issue… it's a point we've made in interviews (that Cities are being held to task for an issue that needs to get resolved Federally), but we definitely need to speak up about how we (Cities) are stepping into the void and doing everything we can to protect people's privacy".

Up to this point in late January 2019, 'protecting privacy', as a project requirement — and professional duty — articulated among the Portland planners and between the planners and Replica, had primarily meant 'preventing the identification of individuals whose location data contributed to Replica's model' ("The process of building Replica's, and therefore the Replica outputs, protect individual privacy, whereby it's mathematically impossible to re-identify an individual within Replica"). Now, a different reading of what 'respecting privacy' might mean is introduced: that the collection and use of the data on which the model depended needed to be consented to by the individuals represented. This reading of privacy was reinforced the following week, when during an unrelated public hearing on a new City privacy policy, Portland planners "fielded [concerns] from community groups about Sidewalk… People continued to raise the question… about whether we are endorsing the use of cell phone data that people feel has been collected without adequate consent by working with Sidewalk Labs".

Due Diligence Tactics and Attempts

To render a professional judgement on whether the platform and model were 'privacy-protective' in the sense of 'effectively anonymized', Rose asked for detailed information on Replica's synthetic data-generation process, but his requests were deflected by then-Replica Product Lead Nick Bowden. Bowden attempted to assure the planners of the efficacy of their anonymizing techniques by sharing a summary of the results of an independent privacy audit, only for Rose to respond that these summarized results effectively asked him to place his trust in the auditor, rather than in his and his colleagues' own analysis — which they were being precluded from performing without access to detailed methodological information. Rose asked again for the full report; Bowden responded "[w]e can't share the entire report, only in that it is a comprehensive security and privacy report which [sic] a substantial amount of proprietary and sensitive information related to our architecture, algorithms, and data processing".

By March the Portland group had a backup plan to determine whether Replica's anonymization would be fit for their purposes. They negotiated a contract provision that allowed them to 'red-team' the platform (cf. Singh et al, 2025) by attempting themselves to see whether they could identify individuals in Replica's synthetic, model-generated data.

But neither methodological transparency nor adversarial testing could reveal whether the collection of the data on which Replica depended was consented to. Back in September 2018, Rose had emailed Bowden that "we continue to be interested in any documentation you've developed about the data sources and methodology used in Replica. You sent a research paper that covers the use of cell data well, but doesn't discuss other sources in detail. I'm particularly interested in hearing more about how Replica incorporates Streetlight data, and how the end product differs from Streetlight". Bowden replied, "We actually use a combination of streelight [sic], safegraph, cell companies, consumer marketing data, google data, and census data. So, the first component of differentiation is in the composite of data across multiple sources". This litany of data sources would be the most detailed and concrete information the Portland planners would ever obtain from Replica regarding the provenance of their data.

By mid-December Rose pressed the Replica representatives again for methodological and data provenance transparency: "we would like to see more thorough documentation of the Replica data sources, methodology, and dataset prior to signing… [we still haven't seen] any description of the demographic data that are used by Replica, nor of the specific travel and location input data that will be used in our build of the model. This will help us… make sure that we're prepared to answer any questions that come up on our end about Replica's data sources as the agreement is finalized". Bowden replied, "We are happy to provide additional documentation on data sources and methodology. We do have NDAs in place with many of our data providers, which means we can provide a detailed description of the data used, how it's collected, our vendor auditing process, and other specifics, but cannot specifically name the vendors per our agreements with them". Rose pushed back, "I understand that you have NDAs in place, but the more specificity you can provide on data sources the better. It's particularly important to clarify whether Google data was used to building Replica or not. We've been assuming that based on our conversations and emails, but in your interview w/ OPB [Oregon Public Broadcasting; (Templeton, 2018)] you mentioned that Google data was not used in building Replica. That question will invariably come up as this moves toward approval".

Replica attempted to reassure the planners that their training data had been collected and used with full consent (and not from Google) with a 'data disclosure document' {Replica, 2018}, shared on New Years Eve, 2018. In his email to Rose, Bowden wrote that the document "contains detailed information about data sources and use of said data in the Replica building process. NDAs prevent us from sharing specific vendor names, but this provides a great deal of detail about the general sources and the process. Most importantly, Replica does not handle, process, or store personally identifiable information at any point". Bowden added, "We obviously consider [the data disclosure document] confidential and proprietary but want to make sure you all have a complete understanding of how Replica is created".

The document proclaimed that "Sidewalk Labs independently verifies and audits our providers to ensure stringent privacy protections are in place for data collection. This process also verifies that end-users have provided explicit consent and have the ability to opt-out of collection at any time. This verification process helps to ensure that Replica can confidently build a high-fidelity travel model while respecting individual privacy" (Replica, 2018). More specific (but no less 'trust us') is a section in the document titled "Summary of Vendor Audit Questions"; it begins, "Sidewalk Labs requires all data vendors to complete an audit prior to agreeing to work with the vendor. At any point the vendor fails to meet the privacy requirements, Sidewalk Labs terminates the relationship and no longer uses data provided. The audit is completed across several dimensions of the vendor operations" (ibid). There follows a list of steps in Replica's auditing process, including "Ensure data supplier has a current, clear, and conspicuous privacy notice that describes what data is collected, how it is used and shared, and user choices for opt-out… Ensure data supplier obtains opt-in consent for collection of any location data… Confirm data supplier has provided opt-out mechanisms and that user preferences for opt-out are propagated throughout the supply-chain… Confirm that suppliers make the data collected available to users…" (ibid).

By listing such generic requirements and procedures while failing to include any information, such as specific vendor names, that would have enabled the Portland planners to analyze for themselves whether Replica's data sources and methods were fit for their purposes, Replica was once again communicating, 'trust us'.

Negotiating Goalposts: Acceptance Criteria

By April 2019 the contract was finalized; by May all parties had signed it. The kickoff call for the project took place in April, before the contract had been executed. High on the list of priorities for that call was beginning the process of converging on a key element of the contract that had deferred up until that point: defining the *acceptance criteria* against which the performance of the Replica model would be evaluated (and if wanting, could give cause for the planners to terminate the contract).

After another month of meetings, Rose circulated a draft of the acceptance criteria among the planner consortium members for approval. He noted: "The acceptance criteria do not address some of the more detailed questions that we've heard from you… That's by design. The acceptance criteria are going to go in the contract, and we want them to give us cover to include all the data that we want to include in the validation testing while still allowing for some flexibility as we all learn more about the Replica dataset and process"

This strategic approach to crafting the contract with Replica gave the planners the ability to respond to new information (or reticence) from Replica. Staying at a higher level of abstraction in the acceptance criteria document enabled the planners, down the line, to determine thresholds for trustworthiness based on their evolving holistic assessment of Replica as a partner. This would come in handy.

The process of validation was delineated as follows: the planners assembled already-collected 'ground truth' data based on travel surveys, instrument readings, and observations of mobility events (such as pedestrian counts, public transit ridership counts, etc.). These ground truth data were split into two parts. The larger part was sent to Replica to "calibrate… meaning fit the model to the counts we see on the ground"; the minority remainder of the ground truth data, held back by the planners, were to be used to compare with the Replica model's simulated data in order to 'validate' its accuracy.

The acceptance criteria, then, included contractually-defined thresholds for similitude between the Replica modeled data and the held-back ground truth data. Each ground truth dataset (for instance, a particular travel survey, or a specific demographic census) was assigned a level of variance within which the Replica data must stay to be accepted. If Metro found in their testing that the modeled data was excessively different from the ground truth data, the contract stipulated that Replica would be given "a reasonable period of time using reasonable commercial efforts to correct any such nonconformity". If the corrections weren't acceptable to Metro, either the planners or Replica could opt to terminate the contract — and Replica would refund the planners in full.

In addition to quantitative acceptance criteria, the contract stipulated two qualitatively-assessed acceptance criteria: a review of the adequacy of Replica's privacy documentation ("including Replica methodology and results of third-party privacy audit") and "Inability to identify known individuals".

Replica and the planners negotiated the acceptance criteria addendum to their contract for six months, before both parties signed in September 2019. By then, the planners had been sharing ground truth data with Replica for months. It took until December 2019 before Replica was ready to unveil their newly-calibrated model for the Portland region; Rose organized a week of trainings and meetings among consortium members to introduce them to the Replica platform and to begin assessing the simulated data's usefulness.

Ending Things

Replica's model did not pass the acceptance criteria. In July 2020, a group of the Portland planners met with Nick Bowden (now CEO of the newly-independent Replica) and Replica's lead for the Portland project. After the meeting planner Rose followed up with an email in which he surfaced "the use cases that we submitted to Nick [Bowden] way back in 2017 (see below for original thread). It would help to hear your take on which of these use cases Replica would still be able to meet under the changes you're proposing". The changes proposed by Replica that Rose referred to were not changes to Replica's methods or data sources — they were amendments to the acceptance criteria: bids to move the goalposts.

At the end of July, Replica's lead for the Portland project emailed Rose with a series of "exercises for you all to walk through… [as demonstrations] that align with your use cases". She went on, "[w]e look forward to hearing from Metro how you all would like to proceed - with the amended acceptance criteria, a shift to the standard acceptance criteria [rather than the custom acceptance criteria negotiated and signed by both parties in September 2019] or sharing proposed amendments with Replica".

Rose responded: "Unfortunately, we can't accept either of the options that you've presented us. Based on what you're proposing, we don't feel like we can rely on Replica… The reduced scope of testing you're proposing, along with the limited information Replica has been willing to provide so far, would mean that we don't have enough transparency about how accurate and representative Replica's data is to use it with confidence in our projects". Rather than take Replica up on its offer to move forward with amended acceptance criteria, Rose countered, "[i]f you see a pathway to meeting the criteria in [the existing] agreement we're willing to discuss how we can move forward. Otherwise, since the data Replica has provided so far doesn't pass acceptance testing, we'd like to follow the process outlined in our current contract and terminate the agreement".

A back-and-forth over email ensued; Replica alleged that Portland had provided insufficient ground truth data for the categories of modeled data that had failed the acceptance criteria. The planners shot back that it took until March 2020 for concerns to be raised about the volume and quality of the supplied ground truth data — and that "it was clear from your comments and questions at that point that you had lost track of some of the data and documentation we had shared with you. Replica's behavior leads us to suspect that you didn't do your due diligence on the ground truth data when

we initially uploaded it and agreed on acceptance criteria, and that once you realized it was going to be challenging for you to meet those commitments you opted to focus on disqualifying criteria that you are having difficulty meeting rather than working with us to address them".

Rose goes on to note that the planners had asked Replica for "specific pieces of information that we need to be able to use Replica in a way that meets our standard for transparency, and you have not provided any of them". Among these was "[q]uantitative information on margins of error associated with Replica estimates… so we can [assess] whether Replica's estimates are accurate enough to be used in different planning contexts; and whether Replica adequately protects people's privacy. You have provided qualitative ratings that state whether Replica would recommend using data for different purposes at different scales. We need quantitative information so we can make that determination for ourselves".

Above, I have discussed how Replica's opacity around their data provenance became an article of contention between Replica and the planners. Early in their negotiations, this was registered as an issue of verifying that Replica's data sourcing was respectful (consentful) and protective of individuals' privacy (effectively anonymized). In his email initiating the termination of the contract, Rose mentions that the planners asked for and were denied "aggregate information on the home and work locations of cellular devices that Replica uses to build its dataset that we can use to understand whether Replica adequately represents certain groups, such as communities of color and rural residents". Here, Replica's lack of data transparency becomes registered as an issue not of verifying data subjects' anonymity or opportunities to consent, but of equitable representation. Rose wrote that the planners "need this information to feel confident relying on Replica given the concerns that Replica's lack of transparency so far have raised. You made a major change in your input data sources that invalidated most of the demographic acceptance criteria in our agreement without informing us. We planned to use demographic testing to ensure that the data we use is representative, and it's only because you changed your data sources that we are requesting additional information on representativeness"; the normative hazards of Replica's data practices are cast not only as an issue of opacity, but fluidity.

The Right to Discretion

Here is the crux of the conflict — who has the authority to declare whether a digital tool is fit for purpose? Who should assign thresholds for 'accurate enough', 'privacy protective enough', and other normative characteristics of the system? Each time Replica implicitly communicated another version of 'trust us', they precluded the exercise of the planners' discretionary agency and situated expertise: sharing a short summary of the results of a privacy audit rather than contractually-required details of analysis and methodology; declining to provide concrete information about data provenance and instead offering generic categories of data sources; and presenting "qualitative ratings" of the modeled data's fitness for purpose rather than the "quantitative information" the planners would need to "make that determination for [them]selves".

As described in Hong (2023, 6), "[t]o exercise discretion is… to judge which rules should apply to a given situation, and to define the situation as requiring a rule-based resolution in the first place (or not)". The haggling over acceptance criteria can be read as a contest of who has the right to determine what constitutes a sufficient degree of fidelity for the intended purposes for the modeled data — in other words, a conflict over whose discretion counts.

Replica's unwillingness (or inability) to transparently share their methods, data provenance, and validation testing results not only limited the planners' ability to determine the fitness for purpose of the data system on a case by case basis, but it circumscribed their ability to be transparent with their constituents: "For us, 'transparency'… means that we can share quantitative, objective information with the public on whether the data we use is representative, accurate, unbiased, and protective of privacy". Planners owe legal and professional duties to their region's residents, including transparency. The planners' legal duty of transparency as public employees, as codified in the federal Freedom of Information Act (FOIA) statute, is what enabled me — and local journalists — to analyze and write about their exchanges by accessing their email interactions with Replica through a public records request. Political philosopher Simone Weil writes that there are no rights without duties (Weil, 2001); to the degree that the planners honor their duties (*inter alia,* transparency, accountability) they have rights to make normatively consequential decisions about planning practices and tools that come to shape the lives of residents.

Replica and its employees have no such duties to the public. When journalists filed a series of FOIA requests to obtain the emails between the Portland planners and Replica, Bowden requested to intensively redact the disclosures, causing Rose to "encourage [Bowden] to take one more pass at this… and make sure that all the redactions [he was] requesting are defensible given Oregon law's definitions of confidential and trade secret… and remove those that aren't". Bowden doubled down instead. While at various points Replica had justified their opacity surrounding data provenance with the inhibitions imposed by their NDAs with data providers, here Bowden justified secrecy with Replica's need to protect themselves against "competing firms looking to uncover as much about our product and business model as possible" and claimed his trade secrets

included, among others, "[o]ur business model… References to other customers… Our technical process and procedure of combining several disparate data sets… Additionally, the specific sources of this data and how / when it is used is something we consider to be a significant trade secret…". Replica's duties, at the end of the day, are to their investors. At various points in the project, methodological and supply chain transparency became an arena for conflict between duties to the public and duties to investors; rather than continue to fight a losing battle, the planners chose to resign the game and seek more aligned partners elsewhere.

Discussion

"Our residents are concerned about serious transportation safety and equity issues, and have asked us to get better data so that we can figure out how to best tackle these issues. At the same time, our residents are also increasingly concerned about privacy. This pilot is an opportunity to understand whether there are conflicts between the need for better data and the right to privacy, and if so, how we can address those conflicts".

At the project's outset, Rose framed the partnership as a genuine experiment: not 'does this tool work?' but 'does the value of this tool outweigh the normative risks associated with its creation and use?' Was this question answered? At first glance, perhaps it was not; the partnership was ended before the question could be answered, because the delivered models under-performed their acceptance threshold. In theory a company and product 'exactly like Replica' but more performant could have resulted in an answer being provided, and in the affirmative.

But Replica's reasoning for why their data provenance and methodology should be obscured (both to the planners and to the public) was 'trade secrecy', justified by Replica's position in the market, competing for investor dollars, under the harsh gaze of would-be competitors. So the issue is not really about performance, but about situatedness. In other words, Replica became an invalid dependency for the planners not (just) because it wasn't dependable (fitness-for-use), but because its location within the digital political economic grid (startup, private, investor-backed) militated against its ability (or at least desire) to engage in trust-building transparency with planners — and the public. "The conflict between trade secrecy and a transparent and accountable democratic government is ultimately a clash of governing theory and values… trade secrecy and public accountability cannot easily coexist" (Levine, 2007, 157-158).

The structure of digital supply chains, particularly with regard to AI systems, includes a number of centralized actors whose incentives — and methodologies — incline them to cut information flows and screen off visibility of their own upstream dependencies. These actors might be model developers pre-emptively defending themselves against accusations of intellectual property theft in their mass-scraped training datasets (Widder et al, 2023a). They might be data brokers looking to preserve the perceived uniqueness of their aggregated data products/services (Obar, 2020; Gansky & McDonald, 2022). They might be startups like Replica whose market competitiveness depends on a plausible narrative of uniqueness in terms of their data sources.

Then, too, the legal structures which scaffold the commercial exchange and aggregation of data are biased towards opacity. NDAs undergird much of the commercial data interchange ecosystem. The ubiquitous 'terms of service' contracts to which users of software 'agree' sometimes just by virtue of their visiting a webpage (Lemley, 2022) also function to obscure, rather than reveal, the normative salience of data interactions — in this case, by (ironically) preventing meaningful consent to data collection and use by formatting relevant information into functionally illegible forms and replacing what could resemble something like a widely understood meaning of 'contract negotiation' with 'click to continue'. "The scale of corporate power over people through data, combined with the ubiquity of the private sector in the public sphere, means that increasingly people do not have the ability to opt out of providing their data to firms— and that states have little option but to turn to technology providers for data on the public, as could be seen with the Google/Apple API developed in the Covid-19 emergency" (Taylor, 2021, 903).

The data subjects whose location data, app interactions, and other representations were sucked into the supply chains by which Replica obtained its data were — obviously — not offered opportunities to opt in or out of data collection and use by the data's collectors, Replica, or the Portland planners, for that matter. "The terms-of-service information people receive about data's lifecycle usually refer to 'research' and 'third parties' as if this constituted meaningful information based on which people could exercise rights over their data. Everyday life in the data economy, however, demonstrates that this kind of control and knowledge are an illusion" (Taylor, 2021, 905). The idea that 'clickwrap' contracts ('click here to accept the terms and conditions' or even better, 'use of this service implies acceptance of the terms and conditions') comprise real consent is generally understood to be a farce (Obar, 2020; Barassi, 2019; Solove, 2013).

Even setting aside the growing body of critiques of individualized consent as an effective mechanism for data governance (McNealy, 2021; Sexton et al, 2018; Tsosie et al, 2019; Viljoen, 2021) Replica's claim that their data were obtained consentfully is implausible. To give just one example, the 'data disclosure document' mentions that Replica "secures [mobile location] data from several vendors, including but not limited to; telecommunications compa-

nies…" (Replica, 2018). At no point did a Replica representative share with the Portland planners over email from which specific telecom they obtained their data, nor did they give more detail about the form in which that data was collected. However, in a 2020 NBER paper that used Replica data, the co-authors mention that for purposes of training a model to assign 'trip purposes', Call Detail Records (CDRs) were used (Akbarpour et al, 2020). It was not to the Portland planners but to a journalist that Sidewalk Labs spokesperson Dan Levitan confirmed that the telecom corporation AT&T was at least one of the company's data suppliers (Kaye, 2019).

Is it plausible that AT&T offered its implicated customers "a current, clear, and conspicuous privacy notice that describes what data is collected, how it is used and shared, and user choices for opt-out… opt-in consent for collection of any location data… [and that it made] the data collected available to users" (Replica, 2018)? Every time an AT&T customer in the Portland area made a call, sent or received a text, or connected to mobile data, that exchange was logged in a call detail record (CDR) that could well have become a datapoint in the mix of training data used by Replica to develop its models. While such customers might have been technically 'notified' (in a 'terms of service' contract unlikely ever to be read by human eyes) that their data could be sold or shared with other parties, Replica and the planning agencies were certainly not mentioned as potential users of such data.

If consent is truly a prerequisite to the normatively proper collection and use of sensitive data, Replica's 'vendor audit' is a paper-thin shield, a mere means of plausible deniability to participation in the surveillance economy. But/and: any available means by which an analyst might obtain such data as Replica is trained on (*inter alia,* CDRs, financial transaction data, granular location data) would require engagement with the commercial digital supply chain, which as I have argued is complex, opaque, and includes data that are absolutely obtained and processed without consent.

The political economic dynamics that encourage opacity in the data supply chain as a competitive strategy are the same political economics that incentivize and energize the data supply chain as in the first place — primarily, markets for data that can be used to programatically target ads to profitable consumers motivating the continuous, pervasive collection of personal data (Hwang, 2020; Zuboff, 2019). Even use-cases that stand entirely apart from advertising and marketing concerns — like public planning — still typically rely on data that may be generated 'incidentally' as 'traces' of mediated activities (Thylstrup, 2019) — but which are rendered 'valuable' by being reconfigured and brought into relation with other data 'exhaust' as part of their introduction and integration into data supply chains.

Conclusion

It should be evident by this point that the Portland planners were navigating challenging due diligence terrain. But they were far from feckless. Their tactics are instructive:

1. They refused the standard evaluation criteria offered by their vendor in favor of establishing their own (quantitative and qualitative) thresholds for validation and acceptance
2. They demanded direct access to aggregate data that they could use to validate data quality (e.g. checking the locations of represented residents to assess whether inputs were equitably distributed across different demographic groups in the metro)
3. They fought for visibility into the data sources to assess qualitatively for themselves whether the data was sourced in a consentful way
4. They advocated for their contractual right to probe the privacy protections assertedly enacted by the Replica system

In other words, they sought metadata transparency (provenance, aggregate summaries of geographic/demographic representativeness), specific tests and direct access to the results (custom acceptance criteria, quantitative margins of error, privacy audit with full results), and the agency to adversarially test the system in question. These are all opacity-mitigation tactics designed to facilitate the planners' decision-making around fitness for purpose *in the absence of systemic visibility*. But "transparency is not coextensive with accountability. It is merely a means. An algorithmic process is accountable when its stakeholders, possessed of meaningful transparency, can intervene to effect change in the algorithm, or in its use or implementation" (Brauneis & Goodman, 2018, 132). To what ends could the planners put their sought-after transparency?

In the email that began the termination process with Replica, Rose wrote, "Replica's conduct has not only prevented us from having the transparency that we need to rely on Replica's data; it's damaged our trust in the project and in Replica as a partner. We've given you opportunities to restore our trust by providing more thorough information; you've opted not to take those opportunities". But trust isn't about 'providing more thorough information'; 'more information' is needed precisely when there is an *absence* of trust. Trust is generated through accountability. The planners and Replica were accountable to different stakeholders. In the final determination, the interests of these stakeholders were mutually exclusive — and rather than fight for dominance, the planners disengaged to redeploy their resources elsewhere.

References

Abbas, R., Michael, K., & Michael, M. 2014. The regulatory considerations and ethical dilemmas of location-based services LBS: A literature review. *Information Technology & People*, *27*1, 2–20. https://doi.org/10.1108/ITP-12-2012-0156

Akbarpour, M., Cook, C., Marzuoli, A., Mongey, S., Nagaraj, A., Saccarola, M., Tebaldi, P., Vasserman, S., & Yang, H. 2020. *Socioeconomic Network Heterogeneity and Pandemic Policy Response* Working Paper 27374. National Bureau of Economic Research. https://doi.org/10.3386/w27374

Alkhatib, A. 2021. To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3411764.3445740

Barassi, V. 2019. Datafied Citizens in the Age of Coerced Digital Participation. *Sociological Research Online*, *24*3, 414–429. https://doi.org/10.1177/1360780419857734

Bates, J. 2012. "This is what modern deregulation looks like": Co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *The Journal of Community Informatics*, *8*2. https://doi.org/10.15353/joci.v8i2.3038

Batty, M. 2016. Classifying urban models. *Environment and Planning B: Planning and Design*, *43*2, 251–256. https://doi.org/10.1177/0265813516630803

Baykurt, B. 2022. Algorithmic accountability in U.S. cities: Transparency, impact, and political economy. *Big Data & Society*, *9*2, 205395172211154. https://doi.org/10.1177/20539517221115426

Benjamin, R. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Birhane, A., Prabhu, V., Han, S., & Boddeti, V. N. 2023. *On Hate Scaling Laws For Data-Swamps* arXiv:2306.13141. arXiv. https://doi.org/10.48550/arXiv.2306.13141

Birhane, A., Prabhu, V. U., & Kahembwe, E. 2021. *Multimodal datasets: Misogyny, pornography, and malignant stereotypes* arXiv:2110.01963. arXiv. https://doi.org/10.48550/arXiv.2110.01963

Birhane, A., Steed, R., Ojewale, V., Vecchione, B., & Raji, I. D. 2024. AI auditing: The Broken Bus on the Road to AI Accountability. *2024 IEEE Conference on Secure and Trustworthy Machine Learning SaTML*, 612–643. https://doi.org/10.1109/SaTML59370.2024.00037

Bowden, N. 2018, April 6. *Introducing Replica, a next-generation urban planning tool*. Sidewalk Labs. https://web.archive.org/web/20190204120144/https://medium.com/sidewalk-talk/introducing-replica-a-next-generation-urban-planning-tool-1b7425222e9e#expand

Braun, L., & Grisson, R. 2023. Race, Lung Function, and the Historical Context of Prediction Equations. *JAMA Network Open*, *6*6, e2316128. https://doi.org/10.1001/jamanetworkopen.2023.16128

Brauneis, R., & Goodman, E. P. 2017. *Algorithmic Transparency for the Smart City* SSRN Scholarly Paper 3012499. Social Science Research Network. https://doi.org/10.2139/ssrn.3012499

Brennan, S., Coulthart, S., & Nussbaum, B. 2023. The Brave New World of Third Party Location Data. *Journal of Strategic Security*, *16*2, 81–95. https://doi.org/10.5038/1944-0472.16.2.2070

Buolamwini, J., & Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Buschek, C., & Thorp, J. 2024. *Models All The Way Down* [Webpage]. https://knowingmachines.org/models-all-the-way

Chen, L., Zaharia, M., & Zou, J. 2024. How Is ChatGPT's Behavior Changing Over Time? *Harvard Data Science Review*, *6*2. https://doi.org/10.1162/99608f92.5317da47

Collington, R. 2019. *Digital Public Assets: Rethinking value and ownership of public sector data in the platform age*. Common Wealth.

Collington, R. 2021. Disrupting the Welfare State? Digitalisation and the Retrenchment of Public Sector Capacity. *New Political Economy*, *27*2, 312–328. https://doi.org/10.1080/13563467.2021.1952559

Cox, R. 2019. Surviving software dependencies. *Communications of the ACM*, *62*9, 36–43. https://doi.org/10.1145/3347446

Crawford, K., & Paglen, T. 2019. *Excavating AI*. -. https://excavating.ai

Cyphers, B. 2022, August 31. *Inside Fog Data Science, the Secretive Company Selling Mass Surveillance to Local Police*. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2022/08/inside-fog-data-science-secretive-company-selling-mass-surveillance-local-police

Dhaliwal, R. S. 2025. The Infrastructural Unconscious: Do Computers Dream of Carbo-silico Pipelines? In B. Siegert Ed., *Reckoning with Everything*. meson press.

D'Ignazio, C., & Klein, L. F. 2020. *Data Feminism*. The MIT Press.

Fatima, K., Schuckers, M., Cruz-Ortiz, G., Hou, D., Purnapatra, S., Andrews, T., Neupane, A., Marshall, B., & Schuckers, S. 2024. A large-scale study of performance and equity of commercial remote identity verification technologies across demographics. *2024 IEEE International Joint Conference on Biometrics IJCB*, 1–8. https://doi.org/10.1109/IJCB62174.2024.10744432

Faulkner-Gurstein, R., & Wyatt, D. 2023. Platform NHS: Reconfiguring a Public Service in the Age of Digital Capitalism. *Science, Technology, & Human Values*, *48*4, 888–908. https://doi.org/10.1177/01622439211055697

Fourcade, M., & Gordon, J. 2020. Learning Like a State: Statecraft in the Digital Age. *Journal of Law and Political Economy*, *1*1, 32.

Gansky, B., & McDonald, S. 2022. CounterFAccTual: How FAccT Undermines Its Organizing Principles. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1982–1992. https://doi.org/10.1145/3531146.3533241

Ghosh, S., and Caliskan, A.. 2023. "'Person' == Light-Skinned, Western Man, and Sexualization of Women of Color: Stereotypes in Stable Diffusion." In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 6971–85. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.465.

Harvey, A., & Laplace, J. 2019. *Exposing.ai: Duke Multi-Target Multi-Camera Tracking Dataset*. Exposing.Ai. https://exposing.ai/datasets/duke_mtmc/

Hockenberry, M. 2021. Redirected entanglements in the digital supply chain. *Cultural Studies*, *35*4–5, 641–662. https://doi.org/10.1080/09502386.2021.1895242

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. 2024. AI generates covertly racist decisions about people based on their dialect.

Nature, *633*8028, 147–154. https://doi.org/10.1038/s41586-024-07856-5

Hong, S. 2023. Prediction as extraction of discretion. *Big Data & Society,* 101, 20539517231171053. https://doi.org/10.1177/20539517231171053

Hwang, T. 2018. Computational Power and the Social Impact of Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3147971

Hwang, T. 2020. *Subprime Attention Crisis*. FSG.

Johnson, M. G. 2020. City in Code: The Politics of Urban Modeling in the Age of Big Data. *Open Philosophy*, *3*1, 429–445. https://doi.org/10.1515/opphil-2020-0115

Katyal, S. K. 2019. THE PARADOX OF SOURCE CODE SECRECY. *CORNELL LAW REVIEW*, *104*.

Kaye, K. 2019, May 28. *Portland quietly launches mobile location data project with Alphabet's controversial Sidewalk Labs*. GeekWire. https://www.geekwire.com/2019/portland-quietly-launches-mobile-location-data-project-alphabets-controversial-sidewalk-labs/

Kerr, M. 2019, January 28. *FW: Google's Sidewalk Labs Plans to Package and Sell Location Data on Millions of Cellphones* [Personal communication].

Koenecke, A., Choi, A. S. G., Mei, K. X., Schellmann, H., & Sloane, M. 2024. Careless Whisper: Speech-to-Text Hallucination Harms. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1672–1681. https://doi.org/10.1145/3630106.3658996

Kofman, A. 2019, January 28. *Google's Sidewalk Labs Plans to Package and Sell Location Data on Millions of Cellphones*. The Intercept. https://theintercept.com/2019/01/28/google-alphabet-sidewalk-labs-replica-cellphone-data/

Lemley, M. A. 2022. *The Benefit of the Bargain* SSRN Scholarly Paper 4184946. https://doi.org/10.2139/ssrn.4184946

León, L. F. A., & Rosen, J. 2020. Technology as Ideology in Urban Governance. *Annals of the American Association of Geographers*, *110*2, 497–506. https://doi.org/10.1080/24694452.2019.1660139

Levine, D. 2007. Secrecy and Unaccountability: Trade Secrets in our Public Infrastructure. *Florida Law Review*, *59*1, 135. https://scholarship.law.ufl.edu/flr/vol59/iss1/2

McNealy, J. E. 2021. *An Ecological Approach to Data Governance* SSRN Scholarly Paper 4164112. https://doi.org/10.2139/ssrn.4164112

Metro. 2019, September 23. *Amendment—Exhibit 1-A. Contract No. 935980 Contract between Metro and Replica, Inc.*. Metro.

Morozov, E., & Bria, F. 2018. *Rethinking the Smart City: Democratizing Urban Technology*. Rosa Luxemburg Stiftung. https://rosalux.nyc/rethinking-the-smart-city/

Noble, S. U. 2018. *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Obar, J. A. 2020. Sunlight alone is not a disinfectant: Consent and the futility of opening Big Data black boxes without assistance. *Big Data & Society*, *7*1, 205395172093561. https://doi.org/10.1177/2053951720935615

Offert, F., & Dhaliwal, R. S. 2025. *The Method of Critical AI Studies, A Propaedeutic* arXiv:2411.18833. arXiv. https://doi.org/10.48550/arXiv.2411.18833

Ory, D. 2017, May 25. *A first step toward creating a digital planning laboratory is populating it* [Blog]. Sidewalk Labs. https://web.archive.org/web/20190204120307/https://medium.com/sidewalk-talk/a-first-step-toward-creating-a-digital-planning-laboratory-is-populating-it-beeb87d485f1

Oved, M. C. 2018, October 12. *Sidewalk Labs use of cellphone data in proposed U.S. deal raises concern in Toronto*. Toronto Star. https://www.thestar.com/news/gta/sidewalk-labs-use-of-cell-phone-data-in-proposed-u-s-deal-raises-concern-in-toronto/article_1346fd08-bc64-5148-b451-115db0545da7.html

Peng, K., Mathur, A., & Narayanan, A. 2021. Mitigating Dataset Harms Requires Stewardship: Lessons from 1000 Papers. *Proceedings of the 35th Conference on Neural Information Processing Systems NeurIPS 2021 Track on Datasets and Benchmarks*. Neural Information Processing Systems NeurIPS.

Replica. 2018. *Replica Data Disclosure Document*.

Richardson, R. 2019. *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*. AI Now Institute. https://perma.cc/888D-XDCX

Sanchez-Graells, A. 2024, September 30. Procuring AI is a risky activity. We need more regulation to make it work in the public interest. *Open Contracting Partnership*. https://www.open-contracting.org/2024/09/30/procuring-ai-is-a-risky-activity-we-need-more-regulation-to-make-it-work-in-the-public-interest/

Sexton, A., Shepherd, E., Duke-Williams, O., & Eveleigh, A. 2018. The role and nature of consent in government administrative data. *Big Data & Society*, *5*2, 2053951718819560. https://doi.org/10.1177/2053951718819560

Singh, R., Blili-Hamelin, B., Anderson, C., Tafesse, E., Vecchione, B., Duckles, B., & Metcalf, J. 2025. *Red-Teaming in the Public Interest*. Data & Society Research Institute. https://doi.org/10.69985/VVGP4368

Solove, D. J. 2013. Introduction: Privacy Self-Management and the Consent Dilemma. *Harvard Law Review*, *126*7, 1880–1903. https://harvardlawreview.org/print/vol-126/introduction-privacy-self-management-and-the-consent-dilemma/

Song, P., Xue, L., Georgia State University, Rai, A., Georgia State University, Zhang, C., & Fudan University. 2018. The Ecosystem of Software Platform: A Study of Asymmetric Cross-Side Network Effects and Platform Governance. *MIS Quarterly*, *42*1, 121–142. https://doi.org/10.25300/MISQ/2018/13737

Stevens, N., & Keyes, O. 2021. Seeing infrastructure: Race, facial recognition and the politics of data. *Cultural Studies*, *35*4–5, 833–853. https://doi.org/10.1080/09502386.2021.1895252

Taylor, L. 2021. Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector. *Philosophy & Technology*, *34*4, 897–922. https://doi.org/10.1007/s13347-020-00441-4

Templeton, A. 2018, December 12. *SimCity In Stumptown: Portland To Get Model Powered By Cellphone Data For Planning*. Oregon Public Broadcasting. https://www.opb.org/news/article/cell-phone-location-data-portland-google-privacy/

Thiel, D. 2023. *Identifying and Eliminating CSAM in Generative ML Training Data and Models*. https://doi.org/10.25740/kh752sm9123

Thylstrup, N. B. 2019. Data out of place: Toxic traces and the politics of recycling. *Big Data & Society*. https://doi.org/10.1177/2053951719875479

Tsosie, K. S., Yracheta, J. M., & Dickenson, D. 2019. Overvaluing individual consent ignores risks to tribal participants. *Nature Reviews Genetics*, *20*9, 497–498. https://doi.org/10.1038/s41576-019-0161-z

Viljoen, S. 2021. A Relational Theory of Data Governance. *The Yale Law Journal*.

Vyas, D. A., Eisenstein, L. G., & Jones, D. S. 2020. Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms. *The New England Journal of Medicine*.

Weil, S. with Eliot, T. S.. 2001. *The Need for Roots: Prelude to a Declaration of Duties Towards Mankind* First Edition. Routledge.

White, M. 2025, March 14. *"Wait, not like that": Free and open access in the age of generative AI*. Citation Needed. https://www.citationneeded.news/free-and-open-access-in-the-age-of-generative-ai/

Widder, D. G., & Nafus, D. 2023. Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society*, *10*1, 20539517231177620. https://doi.org/10.1177/20539517231177620

Widder, D. G., Whittaker, M., & West, S. M. 2024. Why 'open' AI systems are actually closed, and why this matters. *Nature*, *635*8040, 827–833. https://doi.org/10.1038/s41586-024-08141-1

Wylie, B., & McDonald, S. M. 2019, August 13. Sidewalk Labs, the Candidate. *Bianca Wylie*. https://biancawylie.com/sidewalk-labs-the-candidate/

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* 1st edition. PublicAffairs.