Check for updates

Original research article

BIG DATA & SOCIETY

Big Data & Society
October–December: 1–16
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517251389866
journals.sagepub.com/home/bds
Sage

Contesting data power at the margins: Contentious data imaginaries and social movement mobilization

Ngai Keung Chan[1,*], Chi Kwok[2,*], Renyi He[1] and Chris Chao Su[3]

Abstract

This article highlights the role of data imaginaries in shaping narratives about social movement mobilization. Weaving together critical data studies and social movement studies, it examines how activists imagine, negotiate, and resist the politics of datafication, which may further construct collective understandings of political opportunity structures and shape their repertoires of action. Empirically, we analyzed discussions about digital data and surveillance in Hong Kong's 2019–2020 Anti-Extradition Bill (Anti-ELAB) movement. These discussions took place on LIHKG, a Reddit-like online forum that served as a crucial site for framing and mobilizing the Anti-ELAB movement. The analysis identified three data imaginaries. The first envisioned CCTV, transaction, biometric, locational, and mobile phone data as everyday omnipresent surveillance infrastructures. Second, participants imagined how police could weaponize their personal data against them, reflecting deep institutional distrust. Third, participants envisioned how smart lamppost sensor and biometric data, and CCTV could be integrated into a trans-border regime of surveillance, rendering data surveillance an object of contention. Each imaginary articulated a distinctive framing of the threats of datafication and mobilized specific anticipatory tactical responses. Together, they conveyed unease about the present and possible futures of techno-authoritarian control. This study contributes to understanding how activists develop collective visions of data surveillance—data imaginaries—that enable them to identify threats, devise counter-strategies, and foster mobilization within datafied movements.

Keywords

Data imaginary, social imaginary, connective action, contentious politics, mobilization, Hong Kong

This article examines the role of data imaginaries in shaping narratives about social movement mobilization. We conceptualize data imaginaries as individual or collective visions about what data are and should be, whose interests they serve, and how they can be deployed for political purposes (Bucher, 2017; Ferrari, 2024; Kazansky and Milan, 2021; Lehtiniemi and Ruckenstein, 2019). Using the 2019–2020 Anti-Extradition Bill (Anti-ELAB) movement —one of the largest connective actions in Hong Kong's history—as a case, we examine how activists imagine, negotiate, and resist the politics of datafication, which may further construct collective understandings of political opportunity structures (POSs) and shape their repertoires of action.

Utilizing a mixed-methods approach, we analyzed a purposive sample of online discussions about digital data and the Anti-ELAB movement from June 2019 to May 2020. The discussions took place on LIHKG, a Reddit-like public online forum that served as a crucial site for framing and mobilizing the movement (Lee et al., 2022). The analysis identified three data imaginaries. The first envisioned CCTV, transaction, biometric, locational, and mobile phone data as everyday omnipresent surveillance infrastructures. Second, participants imagined how police could weaponize

[1]School of Journalism and Communication, The Chinese University of Hong Kong, Shatin, Hong Kong
[2]Governance and Bureaucracy Lab, Department of Government and International Affairs, Lingnan University, Tuen Mun, Hong Kong
[3]Division of Emerging Media Studies, Boston University, Boston, MA, USA

[*]The first two authors contributed equally and share co-first authorship.

Corresponding author:
Chi Kwok, Governance and Bureaucracy Lab, Department of Government and International Affairs, Lingnan University, Tuen Mun, Hong Kong.
Email: chikwok@ln.edu.hk

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

their personal data against them, reflecting deep institutional distrust. Here, we refer to the metaphorical use of weapons in framing this imaginary to capture how participants feared and speculated that personal data could be used punitively. Third, participants envisioned how smart lamppost sensor and biometric data, and CCTV could be integrated into a trans-border regime of surveillance, rendering data surveillance an object of contention. Each imaginary conveyed a distinctive way of framing the threats of datafication and mobilizing tactical responses aimed at confronting these perceived threats. While these imaginaries rendered data surveillance as thinkable threats and provided seemingly fixable responses, they conveyed unease about the present and possible futures of techno-authoritarian control.

This study contributes to understanding the contentious politics of data in three ways. First, it integrates framing theory from social movement studies with insights from critical data studies, showing how data imaginaries function as critical interpretive frameworks that activists use to shape perceptions of POSs. Second, it highlights how data imaginaries reveal the processes of meaning-making—including the work of translation, speculation, and materialization—rather than passive reflections of existing socio-political conditions within datafied movements. Third, it extends existing research on Hong Kong's Anti-ELAB movement by illustrating how data imaginaries serve as both interpretive lenses and generative forces that continuously reshape activists' tactical responses in the shifting political context.

Literature review

Contentious politics and mobilization: From collective to connective actions

Contemporary social movements operate within complex cultural and discursive fields that fundamentally shape their possibilities for action and meaning-making. Ferree et al. (2002) argue that movements face "discursive opportunity structures"—culturally-specific configurations that determine which ideas gain resonance in public discourse and which interpretive frames become available or legitimate. This cultural-contextual perspective moves beyond rationalist accounts of strategic framing to emphasize how movements' meaning-making efforts are deeply embedded in and constrained by broader cultural fields, power relations, and historically situated discourses. The effectiveness of movement messaging depends not merely on strategic calculation but on the environment's cultural receptivity and the discursive resources available to activists (Ferree et al., 2002; Koopmans and Olzak, 2004).

This emphasis on subjective interpretation and cultural context extends earlier work on POSs (Benford and Snow, 2000; McAdam, 1982; Tarrow, 2011) by highlighting how activists perceive and interpret opportunities, which matters as much as objective conditions themselves. Movements do not simply respond mechanically to structural openings; rather, they actively construct meaning around political possibilities and constraints through culturally-mediated processes (Gamson and Meyer, 1996; Kurzman, 1996). Some environments provide discursive resources that enable certain forms of collective action while constraining others, creating divergent movement trajectories based not only on material conditions but on the interpretive frameworks available to activists (Tarrow, 2011; Tilly and Tarrow, 2015).

This subjective dimension of opportunity structures has become even more pronounced in the digital age. The media environment itself has become part of these opportunity structures, introducing new layers of mediation and interpretive complexity (Tufekci, 2017). Cammaerts (2012) extends this thinking through the concept of "mediation opportunity structures," which captures how communication technologies shape both the material openings and the perceived possibilities that movements face. Digital platforms do not merely provide neutral channels for mobilization, but come embedded with particular logics, affordances, and constraints that activists must interpret and navigate through their own lenses.

This transformation is part of a broader "data revolution" (Kitchin, 2022) that has fundamentally altered how information flows, power is exercised, and activists understand their environment. Bennett and Segerberg's (2012: 739) theory of "connective action" illuminates how contemporary mobilization increasingly relies on personalized content sharing across digital networks, where meaning emerges organically from participants' own subjective understandings and lived experiences rather than through centralized organizational coordination. These personalized frames are not simply strategic choices but culturally embedded expressions facilitated by digital platforms that enable rapid diffusion through mechanisms like algorithmic recommendations and hashtag-driven spread (Tufekci, 2017).

Data become part of the objects of contention, enabling new forms of activism and political engagement (Beraldo and Milan, 2019; Lehtiniemi and Ruckenstein, 2019). Critically, activists now perceive the digital-data environment as integral to how they interpret the costs and opportunities of participation (Milan and Beraldo, 2024; Tufekci, 2017). Features of datafied environments—including algorithmic visibility, surveillance capabilities, and digital data trails—constitute a "datafication opportunity structure" (Kwok and Chan, 2024b; Ruppert et al., 2017; van Dijck, 2014). This represents not just a new set of objective conditions but also a subjectively mediated field requiring continuous interpretation. Activists must navigate these structures through their own frameworks and collective meaning-making processes, interpreting algorithmic logics, surveillance risks, and data traces in ways shaped by their specific cultural and political contexts (Ruppert et al., 2017; Van Dijck, 2014).

Data imaginaries as domains of contestation and mobilization

Just as activists subjectively interpret political and discursive opportunity structures, they develop collective visions about data technologies—data imaginaries—that shape how they perceive and navigate datafication opportunity structures. These imaginaries serve as cultural-cognitive frameworks mediating between objective datafied conditions and activists' strategic responses. The notion of socio-technical imaginaries draws attention to the deeply rooted visions and normative futures that people associate with technology, which in turn shape how they respond to it (Jasanoff, 2004, 2015). Unlike a discrete issue frame, an imaginary emerges through ongoing processes of collective negotiation and contestation about how technology relates to society. They are contested understandings of what is desirable or dangerous. Jasanoff (2004: 2-3) emphasizes this processual nature, noting that technological imaginaries "both embed and are embedded in social practices, identities, norms, conventions, discourses, instruments, and institutions," comprising "the building blocks of what we term the social." These elements are continuously reconstituted through the very processes of the imaginary-making they enable. Imaginaries thus remain emergent and contingent socio-technical constructs that are actively co-produced through the dynamic interplay between various actors, evolving as participants negotiate competing interpretations and adapt to changing circumstances (Jasanoff, 2004, 2015). This study adds to debates on collective and connective meaning-making, emphasizing how people envision possible futures with technology and how those visions become domains of struggle (Ferrari, 2024; Kazansky, 2021; Kazansky and Milan, 2021).

A growing body of work has examined how socio-technical imaginaries of data take shape across different domains. Marzecová and Husberg (2022) detail how a seemingly objective indicator like PM2.5 air quality data became embedded in competing imaginaries about environmental health, civic responsibility, and state accountability. Research on Latin America's data-driven governance has revealed imaginaries entwined with and enabling neo-colonial power and authoritarianism (Ricaurte et al., 2024). These cases demonstrate that data technologies carry multiple yet often competing imaginaries of society and governance, and that they can have material consequences for how data are used and regulated. Building on this insight, Ślosarski (2024) explores a "duality of care" in air quality data activism, showing how activists balance the demand for robust and credible data with an ethical commitment to care for the needs of communities. Debates over what counts as "good data" in that case reflect negotiations between different imaginaries of scientific objectivity versus communal well-being.

In parallel, research on data activism has shown how civil society engages with data both as a tool and as an object of contention (Lehtiniemi and Ruckenstein, 2019). Data activism refers to initiatives that repurpose or resist datafication for social change, which often arise in response to the harms and injustices of pervasive surveillance and data exploitation (Beraldo and Milan, 2019). Activists have leveraged open data and transparency to expose corruption, and created grassroots data projects to address community risks. Mattoni (2017) found how Italian and Spanish anti-corruption campaigns engaged in data extraction and data leaking to challenge institutional secrecy. Likewise, Gabrys et al. (2016) describe citizen-sensing projects in which communities gather "just good enough data" on air pollution to hold authorities accountable and reimagine environmental governance. Such efforts not only use data as evidence but also politicize data itself by reimagining what data means, who controls it, and how it should be used. Even ostensibly neutral open-data initiatives are imbued with context and power relations (Loukissas and Ntabathia, 2021). These studies show that individuals and collectives have the reflexive capacity to interpret, evaluate, and reconfigure the meaning of data (Bucher, 2017). By doing so, activists engage in what Beraldo and Milan (2019: 2) termed the contentious politics of data, "interfering with and/or hijacking dominant, top-down processes of datafication" and reappropriating data practices and infrastructures for alternative purposes.

Lehtiniemi and Ruckenstein (2019) contend that different imaginaries guide different modes of data activism. They distinguished between a technological imaginary, which centers on individual empowerment and technical solutions (e.g. the belief that giving people more control over personal data will automatically yield empowerment), and a socio-critical imaginary, which foregrounds collective rights, justice, and the structural power asymmetries of datafication. These divergent imaginaries illustrate how activists might envision data either as a personal asset to be managed or as a societal problem to be challenged. The technological imaginary frames the core problem as individuals' loss of control over data and thus promotes tools or legal reforms that enhance personal data autonomy. In contrast, the socio-critical imaginary identifies the problem in terms of systemic data injustices, such as corporate surveillance or governmental abuse, and calls for collective action, policy intervention, and structural change. Imaginaries involve processes analogous to framing—defining what is wrong, who is responsible, and what should be done—but activists bundle these elements into broader visions of social change.

Ferrari (2024) illustrates the multiplicity and contestability of data imaginaries in activism. Many of the digital platforms on which activists depend—often products of Silicon Valley—were never designed with dissent or collective action in mind (Ferrari, 2024). These platforms come bundled with a dominant imaginary equating technology with progress, individual empowerment, and freedom. Yet activists often cultivate counter-imaginaries (Kazansky and Milan,

2021), rejecting aspects of the prevailing utopian narrative even as they continue to use corporate social media for pragmatic reasons (Ferrari, 2024). As such, activists continuously reimagine the political possibilities and dangers of the very same technology in light of their own future-oriented goals, constraints, and local political contexts.

Meanwhile, social movements must contend with powerful opponents who deploy datafication against them. Activists grapple with new threats that are often invisible and uncertain, which places a premium on imagination and speculation in movement strategy. Owing to the opaque and evolving nature of data-driven repression, activists frequently develop threat models of what they imagine state or corporate surveillance might entail (Kazansky, 2021; Schopmans and Ebetürk, 2024), and organize defensive tactics around these anticipated scenarios. In other words, the ways that activists and citizens understand the dangers and opportunities posed by data infrastructures depends heavily on how they imagine those infrastructures will function.

Imaginaries inform activists' collective fears, hopes, and expectations about technology (Kazansky and Milan, 2021). While *narratives* are the specific stories, accounts, and discourses that activists articulate about datafication, *imaginaries* are the deeper collective visions that generate and organize these narratives. While prior literature tended to treat imaginaries essentially as a type of frame or narrative that activists deploy, we posit that an imaginary supplies an overarching worldview, that is, how knowledge about the issue is constructed and what future is anticipated, and an emotional tenor that goes beyond the immediate claims of a standard frame. Under conditions of deep datafication, these imaginaries become central domains of struggles in their own right.

Surveillance, socio-technical imaginaries, and digital activism

Contemporary social movements operate in "surveillance societies" (Zuboff, 2019), which creates new risks for activism that movement actors cannot ignore (Kwok and Chan, 2024a). Dominant institutions often justify expanded surveillance through an imaginary that equates data-driven monitoring with public safety and modern efficiency. Therein, narratives encourage people to entrust their personal data to governments and corporations under the assumption that watching equals protection (Lyon, 2017). However, activists engage in reworking, resisting, and countering dominant imaginaries of surveillance as part of their broader struggles (Kazansky, 2021; Kazansky and Milan, 2021).

A central feature of digitally-mediated activism is the need for anticipatory strategy under conditions of uncertainty (Kazansky, 2021; Kwok and Chan, 2024a). As noted earlier, activists often operate on imaginative conjectures about how and when data infrastructures might be monitored. They develop threat models of potential surveillance scenarios—essentially educated guesses about what a repressive adversary could be doing with data (Kazansky, 2021). Those threat models inform defensive tactics geared toward worst-case assumptions. Kazansky (2021: 2) terms these future-oriented precautions as "anticipatory data practices," which are practices concerned with imagined surveillance harms that have not yet materialized. Activists might presume that their social media chats are being intercepted, and thus switch to encrypted messengers preemptively, not because they have concrete proof of monitoring in each instance, but because an imaginary of pervasive surveillance compels them to err on the side of caution (Dencik et al., 2016). They continuously adapt their practices in an arms race against evolving surveillance. As Kazansky (2021: 2) put, "techniques of resistance must stay abreast of techniques of surveillance." This adaptive resistance is an integral part of activist skill sets, even though these efforts may not withstand sophisticated surveillance. The process of engaging in anticipatory countermeasures reflects how deeply imagination and material struggle interpenetrate in data and data-mediated activism.

Context, data, and methodology

Situating data imaginaries in context

From June 2019 to January 2020, Hong Kong witnessed unprecedented waves of mobilization amid diminished political opportunities, with hundreds of protest events occurring across the city (Cheng et al., 2022). While the movement was initially sparked by opposition to a proposed extradition bill that would allow the transfer of fugitives from Hong Kong to Mainland China, Macau, and Taiwan, it later evolved to demand "political accountability, institutional oversight and political development prescribed under the Basic Law" (Cheng et al., 2022: 642).

Research has examined the constitutive role of digital platforms in the Anti-ELAB movement, focusing on how activists enacted platform-specific affordances to create political opportunities and construct collective frames in the absence of centralized leadership. Specifically, the decentralized forum structure and real-time voting mechanism of LIHKG provided opportunities for anonymous users to deliberate on movement goals and tactics (Fong and Lee, 2023), while preventing a few opinion leaders from dominating the discussions (Lee et al., 2022). Additionally, Telegram's privacy-enhancing and polling features enabled activists to disseminate protest-related information (Su et al., 2022) and coordinate collective decision-making (Poon, 2024). Meanwhile, protesters and countermovement groups deployed doxing as a data weapon against each other (Li and Whitworth, 2024). As such, movement actors imagined alternative ways of experimenting with digital communication platforms for resistance.

Our study extends existing research by spotlighting activists' data imaginaries as "tactical responses to perceived threats" (Kazansky and Milan, 2021: 366). Protesters anticipated that their personal data could be deployed for state surveillance, thereby turning data and data infrastructures into an object of contention (Kwok and Chan, 2024a). We argue that these anticipatory concerns reveal complex meaning-making efforts that render "surveillance threats 'thinkable' as a step toward identifying responses" and mobilizing counter-surveillance practices (Kazansky and Milan, 2021: 366).

Data collection

We considered LIHKG as a key site for the production of data imaginaries. We scraped and analyzed LIHKG discussions concerning digital data and surveillance from June 2019 to May 2020. Our research spanned an extended period, from the onset of the movement in June 2019 to its silencing in early 2020. We aimed to examine participants' discourses on datafication and resistance within a particular discursive space rather than to produce generalizable results.

We examined posts and comments that included the following keywords: "big data" (*da shu ju*), "surveillance" (*jian kong*), "CCTV," "smart city" (*zhi neng cheng shi*), and "smart lamppost" (*zhi neng deng zhu*). We included the keywords "smart city" and "smart lamppost" because there were protest events against such projects as surveillance technologies in August 2019. The original dataset included 165,517 posts and comments written in Chinese. We conducted purposive sampling to highlight posts that revealed how participants articulated and materialized their data imaginaries. We counted the number of unique replies to each post (e.g. "unique user count"; mean = 63.39; SD = 193.369) by attending to the unique LIHKG account ID. Although an individual could own more than one account, this operational measure allowed us to analyze the posts that involved active discussions by different members. We examined the titles of posts with at least 64 unique user counts (above the mean) and filtered out the irrelevant ones (e.g. posts that only mentioned CCTV cameras but not digital data). In total, the dataset comprised 76 posts and 42,035 records. Many comments included only emoji reactions and the word "push." While these comments might reveal user engagement that aimed to increase a post's visibility, they might not show how participants envisioned the datafication opportunity structure. We systematically removed empty and irrelevant entries, which resulted in a dataset comprising 76 posts and 29,982 comments.

Analytical approach

We used a mixed-methods approach, combining computer-assisted topic modeling with qualitative thematic analysis, to identify emerging themes about data imaginaries. Our analytic workflow largely followed Byrne's (2022) thematic analysis framework (see the supplementary material for further details about the research process). Although we present the two methods in sequence for clarity, they were applied iteratively, informing and refining one another throughout the process.

In the first phase, two researchers read the posts and comments independently to familiarize themselves with the materials. Given the relatively large number of fragmented texts written by different users, we used topic modeling to label the data and facilitate the discovery of discursive clusters relevant to our conceptual focus. Topic modeling is a productive complementary approach to "providing an overview or 'content map' of immense sets of documents, revealing small but systematic patterns and tendencies in the data" (Törnberg and Törnberg, 2016: 404). This step facilitated open-coding of the materials, thereby mitigating the risk of imposing overly "analyst-driven" interpretations (Byrne, 2022). The research team carefully read, assessed, and grouped the topics. Meanwhile, we conducted qualitative thematic analysis to reflect on the topic modeling results and use the identified topics to further refine the themes.

Topic modeling. Topic modeling was conducted using Latent Dirichlet Allocation (LDA), a generative probabilistic hierarchical Bayesian model designed to infer latent thematic structures from a collection of documents (Blei, 2012; Blei et al., 2003). We followed word segmentation and data-cleaning steps to refine the dataset to include 7754 unique records and 2757 distinct terms. We aimed to identify topic clusters that included distinct terms recurrently used by participants. The underlying intuition of LDA is that each document is a mixture of multiple topics, each characterized as a probability distribution over a fixed vocabulary (Buhin Pandur et al., 2021; Chen et al., 2019). The model operates through three steps: (a) documents are assigned a probability distribution over topics sampled using a Dirichlet distribution, (b) each word within a document is associated with a single topic determined by the same Dirichlet distribution, and (c) each topic is represented as a multinomial distribution over the vocabulary terms assigned to it. These steps allow LDA to model the thematic composition of documents, providing a structured approach to identifying and quantifying underlying topics.

We employed the stm package in R, with parameter estimation conducted using the Gibbs sampling algorithm. To ensure model quality and robustness, the analysis incorporated 1000 iterations and relied on model selection metrics like CaoJuan2009 (Cao et al., 2009) and Deveaud2014 (Deveaud et al., 2014), which assess model sparsity and topic coherence, respectively. These metrics were used to evaluate optimal topic solutions, yielding potential configurations of 17, 29, 36, and 48 topics. A topic contains "a list of words with different assigned probabilities" (Törnberg and Törnberg, 2016: 405) that requires further

qualitative analysis. While higher-topic solutions offered finer-grained segmentation, the 17-topic solution consistently showed a favorable balance between coherence and parsimony. This model provided sufficient conceptual differentiation while maintaining interpretability, making it suitable for inductive thematic analysis. Three researchers independently reviewed the topic-term matrix and top-ranked documents associated with each topic. The 17-topic model was selected not only based on the model statistics but also its semantic clarity, alignment with our conceptual goals, and usability for synthesizing broader discursive themes. Table 1 presents the top keywords associated with each of the 17 topics, along with the topic labels assigned during our analysis.

Qualitative thematic analysis. The qualitative thematic analysis treated the identified topics "as lenses for viewing a corpus of documents" (Törnberg and Törnberg, 2016: 407). Several topics featured top terms that included vulgar expressions, reflecting the discussions' informal and emotionally charged tone. We read the posts and comments iteratively to contextualize the recurrent data-related discussions and the underlying imaginaries, followed by a review of the 17 computationally generated topics. We theoretically operationalized data imaginaries by breaking them into the dimensions of problem identification and anticipatory tactical responses. Three key themes were grouped following deliberation among research team members. Table 2 summarizes the key themes and the related topics. While it is not necessary to include all topics into the three themes, we found that they aligned meaningfully with one or more of the three themes. The three themes were analytically distinct, ranging from the infrastructural layer of data surveillance to the relationship between data and state discipline to fears surrounding Hong Kong's techno-political future. Yet, they projected a shared concern about how data could be (mis)used, contested, and negotiated.

In short, our analysis involved an iterative refinement of topic labels and thematic categorization, informed not only by the topic modeling results but also by the close reading and qualitative coding of the posts. This combined approach helped map data imaginaries and contextualize the relationship between such imaginaries and the movement within the broader discursive field.

Critical data scholars have highlighted the risk of re-identifying sources, the traceability of online posts, and future-oriented harm to data subjects even when analyzing publicly available information (Chan and Kwok, 2025; Franzke et al., 2020; Markham, 2012). The digital traces of protesters could potentially be used as evidence against themselves. While our dataset does not include sensitive categories (e.g. race, ethnicity, or political beliefs), user IDs and direct quotes of comments might allow for potential re-identification. Hence, we deliberately discussed the key themes and reconstructed the quotes into composite accounts of forum interactions to reduce the risk of identifying individual users, while providing a rich account of participants' interpretive practices during the movement (Markham, 2012). All quoted materials were translated from Chinese to English.

Findings

Our analysis of LIHKG discussions identified three distinct yet overlapping data imaginaries during the Anti-ELAB movement, each with particular frames about the threats of data and data infrastructures and potential tactical responses to navigate such threats.

Data as imagined, everyday omnipresent surveillance infrastructures

The imaginary of data surveillance as a totalizing power—characterized by real-time monitoring, data interoperability, and inescapable traceability embedded in everyday life—abounded among forum participants. One illustrative example emerged from a post with over 500 unique user comments, alleging that the Mass Transit Railway (MTR) Corporation, a majority government-owned entity, was extensively installing facial recognition-enabled CCTV systems across railway stations. As one user articulated their concerns about the scope and capability of data surveillance: "This is no longer typical CCTV. The system sends the footage straight to the service provider. Then, the provider turns it into data for further processing and stores it in a database. Finally, the data is run through facial recognition software." While some questioned whether MTR's data collection is legally permissible by referencing existing privacy guidelines, skepticism prevailed regarding Hong Kong authorities' adherence to these guidelines, exemplified by comments like "they could move the goalposts however they like." One poster speculated that "MTR's CCTV could be directly sent to police officers." Additionally, participants noted the proliferation of CCTV in other public transportation systems and private housing estates. On the one hand, these comments echoed the reality that there were over 50,000 CCTV cameras in the city in 2019 (South China Morning Post, 2019). On the other hand, these comments lacked concrete evidence about the underlying socio-technical arrangements of monitoring and facial recognition systems.

This imagined all-seeing surveillance implicated a broader ecosystem of socio-technical infrastructures that might enable or intensify the Hong Kong government's collection of citizens' data. For example, following a protest event at Hong Kong International Airport in August 2019, the police reportedly requested CCTV footage and Octopus card transaction records from some airport restaurants (Ling, 2019). Discussing this incident, a participant

Table 1. Topic modeling for the sampled LIHKG discussions.

Topic labels	Percent of documents	Percent	Topic number in LDA	Top 10 terms associated with each topic	Translated
Everyday resistance and resignation under surveillance in Hong Kong	540	6.96%	1	都可以, 有問題, 所有人, 一定會, 好多人, 大部分, 唔可以, 一個人, 打飛機, 冇問題	Anything goes, there's a problem, everyone, definitely, many people, the majority, not allowed, alone, a slang term used to describe people who are engaging in useless activities or living in their own bubble, no problem
Online sense-making and critique of political control	690	8.90%	2	懶人包, 通知書, 事務局, 中英聯合聲明, 第一手, 任何人, 多媒體, 合理的, 聯合國, 香港人	explainer post, notice, bureau, Sino-British Joint Declaration, first-hand information, anyone, multimedia, reasonable, United Nations, Hongkongers
State violence and the erosion of political rights under a surveillance regime	447	5.76%	3	示威者, 黑社會, 催淚彈, 尖沙咀, 個人資料, 地鐵站, 被發現, 近距離, 冇可能, 攻擊性	Protesters, triads, tear gas, Tsim Sha Tsui, personal data, MTR station, being discovered, at close range, impossible, hostility
Anger and dissent in a globalized surveillance society	602	7.76%	4	屌你老母, 唔怪得, 台灣人, 你老母, 集中營, 都可以, 全世界, 半個鐘, 八達通, 星期六	F*** your mother, no wonder, Taiwanese, your mother [vulgar], concentration camp, anything goes, the whole world, half an hour, Octopus card, Saturday
Digital surveillance and the suppression of free speech	541	6.98%	5	屌你老母, 係唔係, 有可能, 香港人, 身分證, 禁制令, 有限公司, 八達通, 分分鐘, 言論自由	F*** your mother, is it, possibly, Hongkongers, identity card, injunction, limited company, Octopus card, could happen any minute, freedom of speech
State punishment as a tool of surveillance enforcement	553	7.13%	6	唔可以, 你老母, 係因為, 黑名單, 都可以, 中國人, 習近平, 禁制令, 被禁止, 對唔住	Not allowed, your mother [vulgar], because, blacklist, anything goes, Chinese, Xi Jinping, injunction, banned, sorry
Distrust in institutions and their surveillance Networks	440	5.67%	7	立法會, 閉路電視, 記者會, 唔記得, 律政司, 集中營, 星期五, 自己人, 一字排開, 第一步	Legislative Council, CCTV, press conference, forgot, Department of Justice, concentration camp, Friday, insiders, lined up in a row
Identity tensions and mainland influence on surveillance in Hong Kong	473	6.10%	8	香港人, 大陸人, 屌你老母, 身有屎, 你自己, 一定要, 記者會, 去旅行, 中學生, 將軍澳	Hongkongers, mainland Chinese, F*** your mother, guilty, yourself, must, press conference, travel, secondary school students, Tseung Kwan O
The seeking of international support for resistance against surveillance	419	5.40%	9	委員會, 美國國會, 刻不容緩, 好多人, 多功能, 新聞稿, 講大話, 唔好意思, 此地無銀, 身有屎	Committee, US Congress, demanding immediate action, crowd, multi-functional, press release, lying, excuse me, a clumsy denial resulting in self-exposure, having done something wrong
Authoritarian silencing and foreign people's concerns about surveillance in Hong Kong	355	4.58%	10	有機會, 中聯辦, 外國人, 衝紅燈, 差唔多, 抑鬱症, 有效期, 有一次, 分分鐘, 偷偷地	Possibility, Liaison Office of the Central People's Government, foreigners, running a red light, almost, depression, expiration date, one time, could happen any minute, secretly
Everyday resistance and the inevitability of surveillance in Hong Kong	422	5.44%	11	香港人, 第一城, 一定要, 基本上, 一模一樣, 食花生, 全方位,	Hongkongers, City One, must, basically, exactly the same, sit on one's hand, all-encompassing,

(continued)

Table 1. Continued.

Topic labels	Percent of documents	Percent	Topic number in LDA	Top 10 terms associated with each topic	Translated
				第二個, 幾個月, 由頭到尾	another one, several months, from beginning to end
Totalitarian surveillance and public distrust	442	5.70%	12	一定要, 屌你老母, 佢老母, 解放軍, 講大話, 屋企人, 信唔過, 唔可以, 同一個, 大大力	Must, F*** your mother, his/her mother [vulgar], People's Liberation Army, lying, family members, cannot be trusted, not allowed, the same one, with great force
Pervasive surveillance and state violence in different neighborhoods	360	4.64%	13	共產黨, 將軍澳, 九龍灣, 返屋企, 真面目, 天水圍, 好似係, 唔需要, 大件事, 銅鑼灣	Chinese Communist Party, Tseung Kwan O, Kowloon Bay, go home, true nature, Tin Shui Wai, looks like, not necessary, serious incident, Causeway Bay
The importance of exposing surveillance to the whole world	467	6.02%	14	佢老母, 全世界, 連登仔, 廣東省, 行得正企得正, 缺一不可, 唔關事, 共產黨, 返上去, 可以話	His/her mother [vulgar], the whole world, LIHKG user, Guangdong province, I've got nothing to hide, indispensable, not related, Chinese Communist Party, go back up, it could be said
The role of advanced technologies in enabling state data collection	365	4.71%	15	一定要, 身份證, 英文版, 人權法, 去醫院, 如果有, 得唔得, 高科技, 可以話, 目擊者	Must, identity card, English version, Human rights law, go to hospital, if there is, does it work, high-tech, arguably, witness
Belonging and everyday disempowerment under surveillance	380	4.90%	16	缺一不可, 唔可以, 香港人, 屌你老母, 身分證, 無啦啦, 新加坡, 一個人, 基本上, 集思廣益	Indispensable, not allowed, Hongkongers, F*** your mother, identity card, all of a sudden, Singapore, alone, basically, collective wisdom
District-level data collection and monitoring in Hong Kong	258	3.33%	17	區議會, 油尖旺, 油尖旺區, 得唔得, 資訊科技, 辦公室, 政府部門, 攝影機, 十八區, 下一代	District Council, Yau Tsim Mong, Yau Tsim Mong district, does it work, information technology, office, government department, camera, 18 districts, the next generation

noted, "Compared with CCTV, which you could still avoid by covering yourself, using a personal Octopus card [at protest sites] is really easy to get you into trouble." This remark highlights two notable points. First, participants differentiated risks associated with bodily surveillance from those involving payment data. Second, it conveyed a sense of resignation about the difficulty of fully evading surveillance in everyday life.

Participants expressed concerns about Hong Kong's smart identity card—introduced in 2018—which is equipped with radio frequency identification (RFID) technology. They feared that the RFID-enabled identity card would allow the Hong Kong government to "read" their personal data from a distance without their knowledge, thereby enabling covert surveillance. These concerns underscored fear about the entanglement between public infrastructure and data surveillance, given the card's mandatory status for all citizens. Moreover, participants voiced anxiety that pro-government media outlets (e.g. Television Broadcasts Limited) and businesses might collaborate with the police by providing privately recorded footage about protest events.

This imaginary was shaped by distrust and fear of made-in-China technology. Examples include cameras in the elevators of buildings, WeChat, Douyin, and other Chinese mobile apps. In a post with over 200 unique user comments, participants found that WeChat seemingly linked their accounts to Facebook without permission. As one shared, "Honestly, why would a Chinese app like WeChat give you the option to link to Facebook, given that Facebook is banned in mainland China?" Although they had never enabled this feature themselves, the participant noted that it appeared to be active and speculated whether "this has something to do with [their own] prior visits to mainland China." Hence, participants were urged to disable linked accounts, uninstall WeChat and other

Table 2. Summary of the key themes.

Theme	Problem identification	Anticipatory tactical responses	Related topics
Data as imagined everyday, omnipresent surveillance infrastructures	⇒ Data surveillance is embedded into different parts of our everyday lives (e.g. CCTV cameras, Octopus card, smart identity card, social media, etc.) ⇒ The intrusion of private sphere is anticipated to intensify because the state can expand its infrastructural capacity to collect, integrate, and exchange diverse forms of personal data across institutional and technological domains	⇒ Everyday resistance and evasion (e.g. using VPNs, prepaid SIM cards, wearing masks, etc.) ⇒ Raising awareness of surveillance through forum posts	1*, 7, 11, 12*, 13*, 15, 16, 17
Data as imagined weapons	⇒ Personal data can be repurposed as evidence to punish dissent and allow authorities to evade accountability ⇒ A sense of injustice arising from the opacity of data surveillance, which fuels deep distrust in the government and other institutions that might control data	⇒ Calls for greater transparency and data disclosure ⇒ Everyday resistance and evasion that might help to reduce the risk of being tracked.	1*, 2, 3, 5, 6, 12*, 13*
Data as imagined trans-border and global surveillance	⇒ Hong Kong would become integrated into mainland China's surveillance apparatus, especially its social credit system in the near future. ⇒ People in Hong Kong, like many in Western countries, are imagined to be subject to state surveillance.	⇒ Appealing to the international community for support	4, 8, 9, 10, 13*, 14

*Denotes that the topics were overlapped across themes.

Chinese apps, or use a separate phone for such apps. These technical precautions reflected an anticipatory concern that such Chinese apps could monitor users' conversations and potentially share their data with the Chinese government. However, not all participants could follow this advice, as some might need to use WeChat for work and family purposes. Others questioned the effectiveness of the suggested technical fixes. As one user wrote, "Even if it's disabled, I think it's very likely that they're still secretly storing all the data. To be safe, the linked Facebook account should also be deactivated."

Together, this imaginary envisioned a scenario in which the private sphere is increasingly subject to state scrutiny due to the state's expanding infrastructural capacity to collect, integrate, and exchange diverse forms of personal data across institutional and technological domains. This is exemplified by one user's comment:

> Many are afraid of how those in power might abuse artificial intelligence, CCTV, and social media. These technologies significantly lower the threshold for monitoring people. With just one click, they can analyze all the information about your entire life. In the past, surveillance required hiring a lot of people, but now it's really easy to implement surveillance.

This imaginary mobilizes participants to resist the seemingly all-encompassing regime of data surveillance through two tactical responses. The first centered on raising awareness and disseminating technical knowledge among participants, who encouraged others to use Telegram, which offers enhanced privacy through end-to-end encryption, as well as virtual private networks when sharing protest-related information, and to avoid uploading such data to cloud services, particularly those lacking comprehensive privacy policies and affiliated with Chinese companies. Protesters were also urged to be cautious about sharing excessive personal information and any evidence of their protest participation on social media and LIHKG. Furthermore, participants suggested that those who might participate in protests use a separate phone—one not linked to their personal accounts and equipped with a non-registered prepaid SIM card—and avoid using their personal Octopus cards or credit cards at protest sites to maintain anonymity.

The second tactical response can be understood as everyday resistance (Scott, 1989). Participants repeatedly reminded others that "the most important thing is for a critical mass of people to cover themselves." The target audience of this call included not only activists and protesters but also movement supporters. As they reasoned, "if everyone wears black and takes special care to cover up when they walk around the city, surveillance cameras won't be able to track those who actually participated in protest events." Additionally, participants proposed that activists and movement supporters could attempt to alert others of the presence of CCTV cameras or even obscure

the lens to degrade the quality of the image data being captured. Examples of the latter included "painting" facial recognition camera lenses, shining mobile phone flashlights, and pointing laser pens to temporarily blind such devices. While participants acknowledged that these tactics could not fully eliminate the risk of being monitored, they framed data surveillance infrastructures as threats and objects of collective contention. One poster argued that if people could take these small actions together, they could make a big difference *vis-à-vis* protecting themselves from data surveillance.

Data as imagined weapons

While the first theme centers on the imagined infrastructuralization of surveillance, the imaginary of data weapons foregrounds how Hong Kong police could use data as evidence to punish political dissent and evade their own accountability. Participants envisioned that their personal data—including e-payment records, identity cards, phone data, and social media activities, as discussed above—could be used as evidence against them. Noting the prevalence of sensor-based surveillance technologies, one shared, "If you show up at protest sites, they might scan your personal data from a distance and then later use it to frame you and settle the score." On the one hand, this quote calling attention to whether and how data might be used for repression remained largely obscure to protesters during the movement. On the other hand, it reflected the fear of being framed, meaning that data could potentially be misused to falsely implicate individuals, putting innocent citizens at risk. This sentiment was echoed by the following comments: "The biggest problem is, even if the data is sufficient to be used as actual legal evidence, it can still give the police a lead to dox you. Then they will find you and make you confess" (a discussion about the potential expansion of surveillance into Hong Kong), and "The real issue is about the people controlling the surveillance system. Have they all been vetted? They could sell your data, frame you, and follow you for no reason" (a discussion about 5G mobile networks, identity cards, and the smart city project in Hong Kong).

To make sense of the deep institutional distrust underlying this imaginary, it is important to consider how participants' discussions were grounded in the protest events. Participants often referenced two notable incidents in 2019: the Yuen Long mob attack on July 21 and the Prince Edward MTR station incident on August 31. The former involved a group of masked individuals attacking protesters and other civilians at Yuen Long MTR station. The police faced scrutiny for "arriving at the scene 39 minutes after initial reports" (Hong Kong Free Press, 2021).[1] In the latter incident, video footage showed that police forcefully entered the MTR station, deployed batons and pepper spray inside train compartments, and conducted arrests on-site (Dixon and Kilpatrick, 2019). While the government claimed that the police operation was targeted at protesters involved in "violent acts" (The Government of the Hong Kong Special Administrative Region, 2019) there were also allegations that non-protesters were inadvertently caught up in the incident (Dixon and Kilpatrick, 2019). Media access was restricted on-site, prompting widespread public concern and calls for the release of full CCTV footage to ensure transparency (Wong, 2019). Central to these two incidents was the contested role of data access, especially CCTV footage, in the pursuit of accountability. One post, which attracted a couple thousand comments, envisioned a scenario in which the MTR might share data with the police. Some commentators speculated that the police might have "monitored the CCTV but chose not to intervene" on July 21. To some degree, these comments align with the first imaginary concerning potential institutional collaboration in the operation of surveillance. Nevertheless, they serve a different symbolic function by highlighting how withholding or limiting access to CCTV data—information that should have been made public for accountability—shielded certain actors from accountability.

This identified problem became more pronounced in relation to the August 31 incident, largely due to MTR's refusal to release the complete CCTV footage. The denial of journalist access to the scene (Wong, 2019) fundamentally altered the mediation opportunity structure (Cammaerts, 2012) available to movement participants. Unable to witness events firsthand or access news-mediated representations, activists confronted a data vacuum that paradoxically became a generative space for collective meaning-making. This enforced opacity—where the state controls both the physical site and its mediated documentation—transformed the absence of verified information into a mobilizational resource. Online rumors sprang up, suggesting that "the MTR had assisted the police force in covering up the true number and nature of casualties from that night" (Wong, 2019).[2] Two illustrative examples came from a post discussing this incident. The first mobilized those affected by the incident to pursue legal action:

> The MTR may delete the CCTV footage very soon. This means that we only have a few days left to secure the evidence. Once this happens, neither the protesters nor regular passengers will be able to seek justice. (…) In theory, you could file a lawsuit against the police or the MTR and request the court to compel MTR to release the CCTV footage.

Another example, as the commentator noted, conveys "a personal theory" developed in the aftermath of the incident. Specifically, the MTR closed the Prince Edward Station for two days after August 31. In early September, the station, after its reopening, became a site of demonstrations, as participants sought to pressure the MTR to release the full CCTV footage. The MTR temporarily closed the station

on multiple occasions. Yet, the commentator observed that "none of these closures lasted as long as the first time." Hence, their "personal theory" was that "something very serious must have happened at Prince Edward Station, likely involving fatal or near-fatal injuries … The two-day closure could potentially have been used to clean up or remove evidence." This speculation could not be substantiated without access to relevant documentation, and falls outside this study's intent. Yet, the circulation of such speculation arguably provided the basis for participants to articulate a sense of injustice and facilitate mobilization.

Participants tried to demand greater transparency from the relevant actors to address the identified problem. The targets extended beyond the MTR to other institutions perceived as complicit in or susceptible to police surveillance, including businesses, schools, and hospitals. While acknowledging the financial cost of civil litigation, one user noted that a potential tactical response was to "go through legal channels." However, as Wong (2019) argued, "Hong Kong's existing Access to Information regime … has no legal force and a refusal to disseminate relevant information to the public would, therefore, be very difficult to challenge in court." This might explain why participants' comments focused on criticizing the government as irresponsible, accusing it of deliberately overlooking police misconduct and ignoring widespread public anger. Given the perceived risk of data being weaponized, participants were urged to evade data collection by engaging in the aforementioned avoidance tactics.

Data as imagined trans-border and global surveillance

The third imaginary remained in what Kazansky and Milan (2021: 308) call its "speculative phase," meaning "it fulfills a diagnostic function, facilitating the initial identification of problems, causes, and potential harmful consequences" but focuses less on concrete socio-technical interventions. It envisioned a dystopian future in which Hong Kong would be fully subsumed into mainland China's surveillance regime, rendering data surveillance an object of contention.

We focus on discussions of the Hong Kong government's installation of "smart lampposts" in high-traffic areas. Equipped with sensors and CCTV cameras to collect data on traffic, weather, and air quality, the smart lampposts were introduced by the Hong Kong government as part of its Smart City initiative. In August 2019, protesters mobilized against the deployment of the smart lampposts. They used electric saws to cut those lampposts down and found that the provider of the lampposts' Bluetooth beacon was linked to a company involved in supporting China's digital surveillance apparatus (Kwok and Chan, 2024a). Through depictions like "surveillance lampposts" and "smart lampposts are watching you," some participants questioned the necessity of equipping lampposts with cameras, suggesting that these devices were intended to collect data on citizens, particularly activists. Others speculated that the surveillance lampposts could be linked to citizens' identity cards and CCTV cameras on public transport, turning Hong Kong into a site of "concentration camps" in the near future. While these examples hint at concerns about omnipresent surveillance infrastructures, they reflect worries about the erosion of civil liberties in Hong Kong under the influence of mainland China—one of the primary concerns documented in on-site protest surveys conducted in June 2019 (Cheng et al., 2022). As one participant commented about the lamppost installation: "The mass surveillance plan has already begun. They're trying to turn Hong Kong into Mainland China within a few years. By then, the streets will be completely covered with CCTV. We'll be totally screwed. Every move we make will be closely watched."

Other users echoed these concerns by linking smart lampposts and surveillance technologies to China's social credit system. In one post, they cited news about how China's social credit system might punish dissidents by restricting their access to travel and social services, and connected this with *1984* and *Black Mirror*'s "Nosedive" episode. One participant speculatively outlined a list of everyday practices, such as visiting LIHKG, criticizing the police, and jaywalking, that could "result in a loss of social credit points" and suggested "people with low scores won't be able to buy property, take transport, or travel." While the Hong Kong government stated in July 2019 that a social credit system would not be implemented, participants expressed distrust in the government. As one poster speculated, "I think the government will just roll out the social credit system under a different name. They'll say citizens' scores won't be deducted, but in reality, the scores will simply be hidden from us and transferred to mainland China." In contrast to those who envisioned a techno-dystopian future as inevitable, some participants countered the premises of smart city development: "To truly become a smart city, we need a mature legal system. Of course, these smart lampposts have their advantages and can even help track criminals. But now, even attending a protest could be illegal." They urged others to consider "whether we should accept this kind of technological development for our future."

These anticipatory concerns might reinforce the popular (mis)representation of China's social credit systems as unified instruments of authoritarian control. What is at stake here, however, is how this imaginary provides symbolic resources for framing data surveillance as objects of struggle. Participants acknowledged that the people of Hong Kong were not the only ones suffering from government-controlled surveillance, partly because Hong Kong is home to a significant number of foreign residents and because people in developed Western democratic countries faced similar problems. When participants discussed the

political threats of data surveillance in Hong Kong, they reminded others:

> The UK and the US are the two countries with the most serious privacy problems in advanced democracies. State surveillance in these countries has been widely discussed in the media and by the public. Only someone unaware of the facts would claim there's nothing to worry about in the UK and the US. Even democracies can use antiterrorism as an excuse to legitimize surveillance—just like the CCP (Chinese Communist Party).

Inherent in this comment is the inevitability of surveillance on a global scale. However, for participants, China's surveillance was more pervasive and intrusive, with a higher degree of control enforced by the ruling party, while the UK and the US had stronger legal protections for privacy and more robust mechanisms to limit government power.

Participants attempted to seek help from the international community. Many LIHKG users called on those proficient in English to expose data surveillance in Hong Kong on Twitter (now X), while mobilizing others to amplify the message's visibility through retweets. One post with more than 800 comments celebrated that a video showing a surveillance lamppost being pushed down had gone viral internationally and urged users to continue sharing it to raise global awareness of the movement. Another post suggested:

> If foreign governments like the US government believe that the smart lampposts in Hong Kong are being used to monitor foreigners or foreign tourists and realize that the digital traces are being uploaded to mainland servers, it is very likely that they will impose technological sanctions on Hong Kong or blacklist the relevant companies, which will add to the chaos of the CCP!

An underlying premise is that seeking power from outside becomes a viable way to resist the surveillance system enforced by the Hong Kong government. While protest actions aimed at international audiences or foreign governments were hardly new in Hong Kong, nor in other movements elsewhere (e.g. Beaugrand and Geisser, 2016), such tactics were elevated to an unprecedented scale during the Anti-ELAB Movement due to protesters' perceived political opportunities (Lee, 2023). One possible explanation for the frequent references to the US lies in the China-US trade war, which broke out in 2018 and escalated in 2019. This may have led participants to see sociotechnical opportunities to pursue international pressure on China.

Discussion and conclusion

This article examined data imaginaries as a contested terrain of collective struggle, shaping the collective perceptions of POS within the Anti-ELAB movement. It focused on how forum participants articulated the threats of datafication and anticipatory tactical responses during Hong Kong's Anti-ELAB movement. The analysis revealed three interrelated data imaginaries marked by notable tensions. Data infrastructures, which are critical elements of POS in data-mediated activism, were imagined as omnipresent and all-seeing due to the extensive collection of biometric, locational, sensor, and mobile phone data. This fostered a dystopian vision in which authorities could deploy digital surveillance to monitor and punish dissent. While this perspective echoed a dominant imaginary of the inevitability of surveillance, participants turned it into a point of contention by publicly exposing and contesting the infrastructure of data surveillance. They discussed not only the local pervasiveness of surveillance but also its global diffusion, drawing parallels between Hong Kong and other high-surveillance contexts. These imagined parallels helped transform fear into a counter-imaginary of shared struggle, galvanizing participants to seek international support for their cause. Echoing previous research (e.g. Ferrari, 2024; Treré, 2018), protesters devised various tactical responses, though they were largely constrained by a pervasive sense of being watched.

Our study advances the understanding of datafied social movements by offering three key contributions that help bridge critical data studies and social movement scholarship. First, while frame alignment processes (Benford and Snow, 2000; Snow et al., 1986) provide a useful analytical lens for understanding how data imaginaries operate, our analysis reveals that imaginaries function as more expansive meaning-making structures meant to shape activists' perception of POSs. Using frame alignment as an analytical tool, we observe how activists engaged in *bridging* (connecting surveillance technologies to broader authoritarian concerns), *amplification* (intensifying fears about data collection), *extension* (expanding surveillance concerns beyond dissidents to all citizens), and *transformation* (converting threat narratives into resistance opportunities). However, data imaginaries exceed the strategic dimensions of frame alignment by encompassing anticipatory futures, emotional registers, and collectively held assumptions about technological possibilities that persist across multiple mobilization efforts. Unlike discrete frames that activists consciously deploy to mobilize support, data imaginaries operate as deeper cultural-cognitive structures that shape how activists perceive the very nature of the datafication opportunity structure itself. The everyday omnipresent surveillance imaginary, for instance, does not simply frame surveillance as threatening—it fundamentally points to participants' understanding of what political opportunities remained available under pervasive datafication. This imaginary attempts to transform the POS not through strategic reframing but by altering activists' basic assumptions about the relationship between data, power, and possibility. Crucially, data imaginaries reveal how meaning-making in

datafied movements involves ongoing negotiation with technological futures that frame alignment does not fully capture. While frames typically address present grievances and mobilization, data imaginaries incorporate speculative dimensions about future surveillance capabilities and anticipatory resistance strategies. This forward-looking orientation could influence how activists assess political opportunities—not just as current openings for action but as evolving technological terrains requiring preemptive adaptation. Thus, data imaginaries do not merely reframe existing political opportunities but represent efforts to actively reconstruct people's understanding of what kinds of opportunities could exist within datafied authoritarian contexts.

Second, our study highlights the importance of attending not only to what datafication *means* to movement participants but also to how they actively engage with it. This emphasis brings to light the anticipatory and relational politics of datafied movements (Kazansky, 2021). Activists in our case exercised interpretive agency by developing data imaginaries that entailed threat modeling and risk assessment (Kazansky, 2021). Participants critiqued the imagined assemblage of public data infrastructures (e.g. smart lampposts) and used those critiques as rallying points during protests. They enroled themselves, potential allies (e.g. overseas supporters), opponents, and the broader public into "anticipatory modes of doing and being" (Kazansky, 2021: 9) via data imaginaries. The meaning-making process unfolds through three interconnected mechanisms. First, *translation*, in which activists transform abstract datafication concepts into concrete, comprehensible threats by drawing on familiar cultural references (e.g. *1984* and *Black Mirror*) and local experiences of data infrastructures, making invisible surveillance infrastructures cognitively accessible. Second, *speculation*, in that, while facing opacity in surveillance systems, activists collectively theorize about potential risks and capabilities, as seen in their discussions about MTR's CCTV systems and smart lampposts. This speculative work fills information voids with collectively constructed interpretations that, while unverifiable, provide actionable frameworks for resistance. Third, *materialization*, in which activists convert these imaginative constructs into tangible practices—from mundane acts like wearing masks to dramatic interventions like destroying lampposts—thereby transforming abstract imaginaries into embodied repertoires of action. This meaning-making process unfolds within unequal power relations, as activists' capacity to reshape datafication is constrained by the power and resources wielded by states and tech companies. Even so, our findings demonstrate how participants negotiated the meaning of technology on their own terms. Additionally, our findings complicate Lehtiniemi and Ruckenstein's (2019) typology of technological versus socio-critical imaginaries. Rather than oscillating between individual-oriented technological fixes and collective-oriented structural critiques, as their typology suggests, Hong Kong activists demonstrated a contextually-driven hybridity that moves beyond such categorization. They simultaneously deployed VPNs for immediate protection while articulating these tools as part of broader resistance against authoritarian surveillance—not as separate imaginaries but as integrated responses to acute political threats. This integration reveals how activists under authoritarian pressure forge imaginaries that are neither purely technological nor socio-critical, but reconstructed through the exigencies of high-risk resistance. While Lehtiniemi and Ruckenstein (2019) advocate for merging imaginaries to achieve more robust data activism, our study shows how such merging occurs organically and necessarily when activists confront imminent datafied repression, producing new forms of politically-embedded technological practice that transcend their original typology.

Third, our analysis reveals a dynamic and co-constitutive relationship between data imaginaries and perceptions of POS that extends the current theorization of both concepts. It also adds to existing research on the Anti-ELAB movement, which generally focuses on the use of platform-specific affordances and considers data-related threats as a kind of discrete frame. Rather than functioning as static interpretive frameworks, data imaginaries in the Hong Kong case continuously reshaped—and were reshaped by—activists' evolving assessments of political possibilities and constraints. When the government introduced smart lampposts ostensibly for urban management, activists' existing surveillance imaginaries transformed this initiative from a potential constraint into a mobilizational opportunity. The physical destruction of lampposts became both a symbolic rejection of datafied control and a mechanism for reconfiguring the POS itself—generating international attention and creating new alliances around shared concerns about techno-authoritarianism. This demonstrates how data imaginaries do not simply interpret existing opportunity structures but actively reconstruct them through anticipatory action and symbolic contestation. The imaginaries' evolution from speculative threat assessment (surveillance as omnipresent) to tactical opportunity identification (lampposts as targets for collective action) to transnational mobilization (surveillance as a globally resonant concern) illustrates a recursive process: imaginaries shape how activists perceive opportunities, these perceptions inform tactical choices, and the outcomes of these tactics subsequently transform the imaginaries themselves. This dynamic interplay suggests that in datafied movements, the boundary between meaning-making and opportunity structures becomes increasingly fluid, with data imaginaries serving as both interpretive lenses and generative forces that actively produce new political possibilities through their enactment.

While this study offers invaluable insights into how activists engage with, reframe, and resist data surveillance through their interpretative practices, it is not without

limitations. First, although the purposive sampling strategy allowed us to spotlight active forum discussions related to digital technologies, data, and surveillance, it may have limited the diversity of perspectives (e.g. emphasis on users who might have stronger opinions about the movement) and excluded activists who strategically chose *not* to participate in online discussions. Consequently, the analysis does not fully account for more covert forms of resistance that occur outside publicly accessible digital spaces. Second, we focused on how activists anticipated data-related problems and articulated tactical responses. Alternative frames may thus exist within the dataset, and more broadly, within forum discussions about the Anti-ELAB movement. Employing additional qualitative methods, such as in-depth interviews and ethnographic fieldwork, would help address these limitations by providing a more nuanced and situated understanding of activists' imaginaries and resistance.

To conclude, data imaginaries constitute a key mobilizational domain in which strategic actors continuously reshape understandings of socio-technical opportunities and threats. In this domain of contestation, activists collectively negotiate what data means for their cause, what dangers it poses, and what countermeasures or new possibilities it affords. Examining the construction, contestation, and evolution of data imaginaries offers a novel analytical window into the intersection of datafication and social movement dynamics.

ORCID iDs

Ngai Keung Chan https://orcid.org/0000-0002-5848-3098
Chi Kwok https://orcid.org/0000-0003-3754-2447
Renyi He https://orcid.org/0000-0003-3194-8032
Chris Chao Su https://orcid.org/0000-0003-4591-1062

Ethical approval

The study did not involve human subjects, and therefore, ethical approval was not required.

Informed consent

The method section has reflected on the ethical considerations of the study.

Funding

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

Conflicting interests

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

Supplemental material

Supplemental material for this article is available online.

Notes

1. According to Hong Kong Free Press (2021), "official accounts of the incident changed from defensive, to conciliatory to often conspiratorial" from 2019 to 2021. For example, the police later claimed that the response time was "within 18 minutes" in August 2020.
2. Hong Kong authorities denied these online rumors and emphasized that no one died during the incident (Wong, 2019).

References

Beaugrand C and Geisser V (2016) Social mobilization and political participation in the diaspora during the "Arab Spring.". *Journal of Immigrant & Refugee Studies* 14(3): 239–243.

Benford RD and Snow DA (2000) Framing processes and social movements: An overview and assessment. *Annual Review of Sociology* 26: 611–639.

Bennett WL and Segerberg A (2012) The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society* 15(5): 739–768.

Beraldo D and Milan S (2019) From data politics to the contentious politics of data. *Big Data & Society* 6(2): 1–11.

Blei DM (2012) Probabilistic topic models. *Communications of the ACM* 55(4): 77–84.

Blei DM, Ng AY and Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.

Bucher T (2017) The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20(1): 30–44.

Buhin Pandur M, Dobša J, Beliga S, et al. (2021) Topic modelling and sentiment analysis of COVID-19 related news on Croatian Internet portal. In: Slovenian KDD Conference on Data Mining and Data Warehouses.

Byrne D (2022) A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity* 56: 1391–1412.

Cammaerts B (2012) Protest logics and the mediation opportunity structure. *European Journal of Communication* 27(2): 117–134.

Cao J, Xia T, Li J, et al. (2009) A density-based method for adaptive LDA model selection. *Neurocomputing* 72(7-9): 1775–1781.

Chan NK and Kwok C (2025) Data, human subjects, and research ethics in transformative contexts. *AI & Society*. Epub ahead of print 3 July 2025. DOI:10.1007/s00146-025-02457-w.

Chen Z, Su CC and Chen A (2019) Top-down or bottom-up? A network agenda-setting study of Chinese nationalism on social media. *Journal of Broadcasting & Electronic Media* 63(3): 512–533.

Cheng W, Lee FLF, Yuen S, et al. (2022) Total mobilization from below: Hong Kong's freedom summer. *The China Quarterly* 251: 629–659.

Dencik L, Hintz A and Cable J (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society* 3(2): 1–12. Epub ahead of print 24 November 2016.

Deveaud R, SanJuan E and Bellot P (2014) Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique* 17(1): 61–84.

Dixon R and Kilpatrick RH (2019) 'I thought I was about to die': Eyewitnesses describe brutal beatings by Hong Kong police. *The Los Angeles Times*, September 2. Available at: https://www.latimes.com/world-nation/story/2019-09-02/hong-kong-police-violence-protesters-eyewitnesses

Ferrari E (2024) *Appropriate, Negotiate, Challenge: Activist Imaginaries and the Politics of Digital Technologies*. Oakland, CA: University of California Press.

Ferree MM, Gamson WA, Gerhards J, et al. (2002) *Shaping Abortion Discourse: Democracy and the Public Sphere in Germany and the United States*. New York: Cambridge University Press.

Fong IWY and Lee FLF (2023) Disagreement resolution on digital communication platform in a self-directed political consumerism campaign. *Information, Communication & Society* 26(15): 3035–3053.

Franzke AS, Bechmann A, Ess CM, et al. and the Association of Internet Researchers. (2020) Internet research: Ethical guidelines 3.0. Available at: https://aoir.org/reports/ethics3.pdf

Gabrys J, Pritchard H and Barratt B (2016) Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society* 3(2): 1–14. Pub ahead of print 1 December 2016.

Gamson WA and Meyer DS (1996) Framing political opportunity. In: McAdam D, McCarthy JD and Zald MN (eds) *Comparative Perspectives on Social Movements: Political Opportunities, Mobilizing Structures, and Cultural Framings*. New York: Cambridge University Press, 275–290.

Hong Kong Free Press. (2021) *Shifting narratives: From 'violent attack' to 'gang fight'—How the official account of the Yuen Long mob attack changed*. Hong Kong Free Press, July 21. Available at: https://hongkongfp.com/2021/07/21/from-violent-attack-to-gang-fight-how-the-official-account-of-the-yuen-long-mob-attack-changed-over-a-year/

Jasanoff S (ed.) (2004) *States of Knowledge: The Co-Production of Science and Social Order*. London: Routledge.

Jasanoff S (2015) Future imperfect: Science, technology, and the imaginations of modernity. In: Jasanoff S and Kim SH (eds) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press, 1–33.

Kazansky B (2021) 'It depends on your threat model': The anticipatory dimensions of resistance to data-driven surveillance. *Big Data & Society* 8(1). Epub ahead of print 29 January 2021. DOI:10.1177/2053951720985557.

Kazansky B and Milan S (2021) "Bodies not templates": Contesting dominant algorithmic imaginaries. *New Media & Society* 23(2): 363–381.

Kitchin R (2022) *The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures*, 2nd ed. Los Angeles: Sage.

Koopmans R and Olzak S (2004) Discursive opportunities and the evolution of right-wing violence in Germany. *American Journal of Sociology* 110(1): 198–230.

Kurzman C (1996) Structural opportunity and perceived opportunity in social-movement theory: The Iranian Revolution of 1979. *American Sociological Review* 61(1): 153.

Kwok C and Chan NK (2024a) Datafication and implicated networks of demobilization: Social movement demobilization in datafied societies. *Social Movement Studies* 23(3): 377–390.

Kwok C and Chan NK (2024b) Human-automated collectives: Automating communication for social movement mobilization. *New Media & Society* 26(9): 4992–5012.

Lee FLF (2023) Proactive internationalization and diaspora mobilization in a networked movement: The case of Hong Kong's Anti-Extradition Bill protests. *Social Movement Studies* 22(2): 232–249.

Lee FLF, Liang H, Cheng EW, et al. (2022) Affordances, movement dynamics, and a centralized digital communication platform in a networked movement. *Information, Communication & Society* 25(12): 1699–1716.

Lehtiniemi T and Ruckenstein M (2019) The social imaginaries of data activism. *Big Data & Society* 6(1). Epub ahead of print 7 January 2019. DOI:10.1177/2053951718821146.

Li YT and Whitworth K (2024) Coordinating and doxing data: Hong Kong protesters' and government supporters' data strategies in the age of datafication. *Social Movement Studies* 23(3): 355–372.

Ling YD (2019) [Airport assembly] Police issue letters requesting submission of CCTV footage: Lawyer warns businesses face risk of civil claims. HK01, August 23. Available at: https://www.hk01.com/%E7%AA%81%E7%99%BC/366978/%E6%A9%9F%E5%A0%B4%E9%9B%86%E6%9C%83-%E8%AD%A6%E7%99%BC%E4%BF%A1%E8%A6%81%E6%B1%82%E4%BA%A4cctv%E7%B4%80%E9%8C%84-%E5%BE%8B%E5%B8%AB-%E5%95%86%E6%88%B6%E6%89%BF%E5%8F%97%E6%B0%91%E4%BA%8B%E7%B4%A2%E5%84%9F%E9%A2%A8%E9%9A%AA (in Chinese).

Loukissas YA and Ntabathia JM (2021) Open data settings: A conceptual framework explored through the map room project. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2): 1–24.

Lyon D (2017) Surveillance culture: Engagement, exposure, and ethics in digital modernity. *International Journal of Communication* 11: 824–842.

Markham A (2012) Fabrication as ethical practice. *Information, Communication & Society* 15(3): 334–353.

Marzecová A and Husberg H (2022) And then came this number PM2.5: Atmospheric particulate matter, sociotechnical imaginaries, and the politics of air quality data. *Environment and Planning C: Politics and Space* 40(3): 648–665.

Mattoni A (2017) From data extraction to data leaking: Data-activism in Italian and Spanish anti-corruption campaigns. *Partecipazione e Conflitto* 10(3): 723–746.

Mcadam D (1982) *Political Process and the Development of Black Insurgency*. Chicago, IL: University of Chicago Press.

Milan S and Beraldo D (2024) Data in movement: The social movement society in the age of datafication. *Social Movement Studies* 23(3): 265–284.

Poon M (2024) Achieving organizationality through authorship affordances — A communicative episode of telegram polling from 2019 Hong Kong. *Social Movement Studies* 23(3): 391–412.

Ricaurte P, Gómez-Cruz E and Siles I (2024) Algorithmic governmentality in Latin America: Sociotechnical imaginaries, neocolonial soft power, and authoritarianism. *Big Data & Society* 11(1): 1–6. Epub ahead of print 6 February 2024.

Ruppert E, Isin E and Bigo D (2017) Data politics. *Big Data & Society* 4(2): 1–7. Epub ahead of print 3 July 2017.

Schopmans H and Ebetürk IT (2024) Techno-authoritarian imaginaries and the politics of resistance against facial recognition technology in the US and European union. *Democratization* 31(5): 943–962.

Scott JC (1989) Everyday forms of resistance. *The Copenhagen Journal of Asian Studies* 4(89): 33–62.

Ślosarski B (2024) Data phronesis and the duality of care in the air quality data politics. *Information, Communication & Society* 27(4): 758–773.

Snow DA, Rochford EB Jr, Worden SK, et al. (1986) Frame alignment processes, micromobilization, and movement participation. *American Sociological Review* 51(4): 464–481.

South China Morning Post (2019) Hong Kong's police force said to have access to facial recognition AI tech - but are they using it? *South China Morning Post*, October 23. Available at: https://www.scmp.com/tech/innovation/article/3034141/hong-kongs-police-force-said-have-access-facial-recognition-ai-tech

Su CC, Chan M and Paik S (2022) Telegram and the anti-ELAB movement in Hong Kong: Reshaping networked social movements through symbolic participation and spontaneous interaction. *Chinese Journal of Communication* 15(3): 431–448.

Tarrow SG (2011) *Power in Movement: Social Movements and Contentious Politics* (3rd ed.). New York: Cambridge University Press.

The Government of the Hong Kong Special Administrative Region. (2019) LCQ1: Incident in Prince Edward station. Available at: https://www.info.gov.hk/gia/general/201911/13/P2019111300525.htm?fontSize=1

Tilly C and Tarrow S (2015) *Contentious Politics* (2nd ed). New York: Oxford University Press.

Törnberg A and Törnberg P (2016) Combining CDA and topic modeling: Analyzing discursive connections between islamophobia and anti-feminism on an online forum. *Discourse & Society* 27(4): 401–422.

Treré E (2018) *Hybrid media activism: Ecologies, imaginaries, algorithms*. New York: Routledge.

Tufekci Z (2017) *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven, CT: Yale University Press.

Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

Wong V (2019) *The '831' Prince Edward MTR incident proves Hong Kong urgently needs access to information reform*. Hong Kong Free Press, September 22. Available at: https://hongkongfp.com/2019/09/22/831-prince-edward-mtr-incident-proves-hong-kong-urgently-needs-access-information-reform/

Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.