01

DATA

OPT-OUT

Data Refusal, Data Strikes and Data Leverage

Data leverage refers to strategies users can employ to influence systems that depend on their data, by altering their data-related contributions. Vincent et al., 2021 outline three strategies: withholding or ceasing contributions (*data strikes*), deliberately corrupting data (*data poisoning*), and intentionally sharing data with value-aligned platforms (*conscious data contribution*). Drawing on feminist and Indigenous scholarship, **Data refusal** frames non-participation as a political act that challenges the authority of data collectors.

DIFFICULTY	COLLECTIVE
●○○○○	Any

Vincent et al., 2021; Zong & Matias, 2024

02

DATA

INTERVENE

Data Poisoning

Data poisoning attacks involve perturbing input samples so that, when they are used for training AI models, they negatively affect the model's performance (Biggio et al., 2012). The goal of this strategy is *influence the model performance when "poisoned" data points are used during training*.

DIFFICULTY	COLLECTIVE
●●●●○	Medium-Large

Biggio et al., 2012

03

DATA

OPT-OUT

Unlearnable Examples

Unlearnable examples (Huang et al., 2021), adds error-minimizing noise to training data samples so that machine learning models cannot extract meaningful patterns from them. The goal of this noise can be seen as a way to *trick the model into believing there is "nothing" to learn from these example(s)*. These perturbations cause the model to treat the protected data as uninformative.

DIFFICULTY	COLLECTIVE
●●●●○	Small-Any

Huang et al., 2021

03 Unlearnable Examples

COMMUNITY LENS

Research Gap Examples of unlearnable examples can be seen in cases such as Sun et al., 2022 which protects open-source code from unauthorized training usage. However, no direct evidence of LGBTQIA2S+ community using tools like unlearnable examples in literature has been documented.

SCAN FOR SOURCES & MORE

02 Data Poisoning

COMMUNITY LENS

Research Gap While literature does not have a lot of evidence of LGBTQIA2S+ community using data poisoning style interventions to prevent harm in algorithmic decision making. Use of tools such as NightShade (Shan et al., 2024) has been documented by artists to prevent explotation of their art without artist consent.

SCAN FOR SOURCES & MORE

01 Data Refusal, Data Strikes and Data Leverage

LGBTQIA2S+ EXAMPLE

Data Refusal practices refer to the refusal of harmful data regimes rather than negotiation within them, a commitment articulated most directly in feminist data scholarship (FeministDataManifestNo). Within the LGBTQIA2S+ context, evidence of this can be seen in how trans activists handle data about their own communities. Stevens & Doğan, 2025 interview 16 activists working across community healthcare, media production, and policymaking, asking how trans activists use data in their activism (Stevens & Doğan, 2025). They find that participants' tactical approaches to data and data science are consistent with contemporary approaches to data refusal, while the study itself moves past refusal to theorize a broader trans data epistemology. Doğan et al., 2025 extends these tactics can benefit data advocacy and CSCW research and design (Doğan et al., 2025).

SCAN FOR SOURCES & MORE

04

DATA

OPT-OUT

Data Defences

Data defences (Agnew et al., 2024) enable data owners to *prevent large language models (LLMs) from inferring personally identifying information (PII)* from textual content. These defenses operate by embedding adversarial prompt injections, specifically crafted to prevent PII extraction, within the original text.

DIFFICULTY	COLLECTIVE
●●○○○	Any

Agnew et al., 2024

05

DATA

AUDIT

Data Archival & Activism

Data archiving and **Data Activism** can serve as an act of grassroots resistance. Currie & Paris, 2018 argue that *preserving data over the long term is itself an activist project*, discussing two literatures that have largely developed in isolation. They identify several shared affinities between archival activism and data activism: both respond to institutional neglect of marginalized perspectives, both seek to make overlooked issues visible and taken seriously in public discourse, and both push beyond standard ways of recording history and presenting statistical evidence. Related practices of counter-data production involve communities assembling their own case records and statistics not only to fill gaps left by official sources but also to reframe public narratives, influence policy, and support affected communities (D'Ignazio et al., 2025)

DIFFICULTY	COLLECTIVE
●●○○○	Small-Medium

D'Ignazio et al., 2025; Currie & Paris, 2018

06

MODEL

AUDIT

Confidentiality Attacks (Repurposing)

Confidentiality attacks compromise a platform's ability to secure its data and models. Platforms developing AI systems often treat their models as proprietary assets, making them accessible only through paid APIs. However, these models remain vulnerable to **model extraction attacks**, in which *an adversary with black-box access to a prediction API attempts to reconstruct the underlying model* by using its predictions to train a substitute model (Tramèr et al., 2016).

DIFFICULTY	COLLECTIVE
●●●●●	Small (skilled)

Tramèr et al., 2016

06 Confidentiality Attacks (Repurposing)

COMMUNITY LENS

Research Gap No direct evidence of LGBTQIA2S+ community use of model extraction as resistance has been documented in literature.

SCAN FOR SOURCES & MORE

05 Data Archival & Activism

LGBTQIA2S+ EXAMPLE

Data archival and activism is documented in Doğan et al., 2025, who study trans activist data practices through a restorative/transformative data science lens and find that recording occurs beyond counting, through archiving and documenting unstructured forms of data. Concrete instances include DigitalTransgenderArchive, a repository preserving and providing access to transgender history, and Felkner et al., 2023, a community-sourced benchmark generated from a community survey to measure anti-LGBTQ+ bias in language models.

SCAN FOR SOURCES & MORE

04 Data Defences

COMMUNITY LENS

Research Gap Data defenses are realized by Agnew et al., 2024, who automatically generate adversarial prompt injections that, appended to input text, reduce an LLM's ability to infer PII about the subject or reuse copyrighted content, with a public tool for protecting text before publication (wagnew3.github.io/LLM-Data-Defenses). No direct evidence of LGBTQIA2S+ community use of data defenses in literature has been documented.

SCAN FOR SOURCES & MORE

07

MODEL

AUDIT

Integrity Attacks / Adversarial Examples (Repurposing)

Integrity attacks aim to subvert or change the behaviour of algorithmic systems. Users wishing to correct algorithmic behaviour without relying on platforms may be motivated to repurpose these attacks to either *evade algorithmic decisions* or *modify data to get favorable outcomes*. Example of repourposing can be seen through **Adversarial examples** which are perturbed inputs designed to *"trick" machine learning models into making incorrect predictions at inference time* (Goodfellow et al., 2015; Szegedy et al., 2014).

DIFFICULTY	COLLECTIVE
●●●●●	Small (skilled)

Goodfellow et al., 2015; Szegedy et al., 2014

08

MODEL

INTERVENE

Availability Attacks (Repourposing)

Availability attacks aim to compromise the reliability of a system or hinder users' access to it, effectively degrading its normal functionality. Shumailov et al., 2021 introduce **sponge examples**, inputs crafted to *drastically increase energy consumption during inference*.

DIFFICULTY	COLLECTIVE
●●●●○	Small-Any

Shumailov et al., 2021

09

MODEL

INTERVENE

Algorithmic Collective Action (ACA)

Algorithmic Collective Action (ACA), introduced by Hardt et al., 2023, provides a framework for analyzing how *coordinated data modifications by groups of individuals can influence the behavior of deployed models*. Given a collective data-modification strategy, the ACA framework can determine the *minimum collective size required to achieve a desired level of success*, as well as *how success scales with collective size*.

DIFFICULTY	COLLECTIVE
●●●●●	Small (skilled)

Hardt et al., 2023

09 Algorithmic Collective Action (ACA)

COMMUNITY LENS

Research Gap Algorithmic Collective Action as formalized by Hardt et al., 2023, coordinated data modification that moves a deployed model toward a target with success scaling in collective size, has no documented LGBTQIA2S+ instance. The nearest documented collective practice is algospeak (Steen et al.), where LGBTQ+ creators collectively alter language to contest TikTok moderation that has suppressed queer content without guideline violations. Algospeak is could also be considered part of folk-theorisation.

SCAN FOR SOURCES & MORE

08 Availability Attacks (Repourposing)

COMMUNITY LENS

Research Gap No evidence of LGBTQIA2S+ community use has been documented in literature, consistent with the absence of any realized resistance use of sponge examples.

SCAN FOR SOURCES & MORE

07 Integrity Attacks / Adversarial Examples (Repurposing)

COMMUNITY LENS

Research Gap Facial recognition is a heavily documented site of LGBTQIA2S+ harm, particularly automatic gender recognition (Keyes, 2018, Scheuerman et al., 2019). One of the response pathways has been regulatory, such as LGBTQ groups joining the ACLU facial recognition letter (Tech). However, no direct evidence of community use of tools / repurposing integrity attack such as Fawkes or LowKey has been documented in literature.

SCAN FOR SOURCES & MORE

10

PLATFORM

AUDIT

Folk Theorisation

When users encounter an algorithmic system, they can form an experiential understanding of these systems based on their interactions. This understanding, termed **folk theorisation** (Karizat et al., 2021), could involve *trying various inputs, creating community knowledge, and trying to understand the working of the algorithms through trial and error*. Folk theorisation could be used to form a shared understanding of the system.

DIFFICULTY	COLLECTIVE
●○○○○	Any

Karizat et al., 2021

11

PLATFORM

AUDIT

Everyday Algorithmic Auditing

Users, during their everyday interaction with algorithmic systems, may notice and report inaccuracies, discriminatory behavior, or harms that algorithms cause on the platform. Shen et al., 2021 describe the process where users of a system *detect, understand, and interrogate harms of the system from their everyday interactions* as **everyday algorithm auditing**. It offers a user-driven solution for communities to observe and report harms they encounter in their regular interaction with algorithmic systems.

DIFFICULTY	COLLECTIVE
●○○○○	Any

Shen et al., 2021

12

PLATFORM

OPT-OUT

Platform Migration & Conscious Data Contribution

Platform migration involves users collectively moving to alternative platforms, either as protest or to seek environments better aligned with their values. Migration can function as a form of *data strike* (Vincent et al., 2019) when it withdraws user-generated content and engagement from an incumbent platform, while simultaneously serving as *conscious data contribution* (Vincent & Hecht, 2021) to a competitor.

DIFFICULTY	COLLECTIVE
●○○○○	Large

Vincent & Hecht, 2021

12 Platform Migration & Conscious Data Contribution

LGBTQIA2S+ EXAMPLE

Example of platform migration is can be seen in documented cases for queer communities in Pan et al., 2025, which traces users moving from TikTok to RedNote after the temporary US TikTok ban and the diffusion of the #wlw hashtag into RedNote.

SCAN FOR SOURCES & MORE

11 Everyday Algorithmic Auditing

LGBTQIA2S+ EXAMPLE

Everyday algorithmic auditing (Shen et al., 2021), where users surface harmful algorithmic behaviour through ordinary interaction, for instance Dennler et al., 2023, a Queer in AI participatory workshop critiquing and redesigning bias bounties from queer perspectives.

SCAN FOR SOURCES & MORE

10 Folk Theorisation

LGBTQIA2S+ EXAMPLE

Folk theorisation is a well documented LGBTQIA2S+ practice. Monea, 2023 shows queer TikTok users building intuitive folk knowledge of the platform's blackboxed algorithm and opaque moderation, then obscuring words and scenes to evade cisheteronormative censorship, and DeVito, 2022 documents how transfeminine creators theorise and navigate algorithmic visibility. Single-actor and expressive variants such as Chokly, 2024 also are examples here for folk theorisation.

SCAN FOR SOURCES & MORE

13

PLATFORM

AUDIT

Harm Reporting Infrastructure

Harm reporting infrastructure provides channels through which users can document and aggregate evidence of algorithmic failures. This includes platform-internal reporting mechanisms (e.g., flagging or appeals processes), as well as external databases (McGregor, 2020, AIAAICRepository), which catalogs real-world AI failures to prevent their recurrence. Recent work has proposed reporting-based frameworks for *identifying systematic algorithmic harms from individual user reports* (Dai et al., 2025). Unlike everyday algorithmic auditing, which describes an organic user-driven process, harm reporting infrastructure refers to the *systems and tools that facilitate the collection and aggregation of such evidence*.

DIFFICULTY	COLLECTIVE
●●○○○	Medium-Large

Dai et al., 2025

14

PLATFORM

INTERVENE

User-Facing Resistance Tools

Mechanisms described in various other cards have been packaged into **user-facing tools** that lower the technical barrier to adoption. These tools *bridge the gap between academic research on adversarial techniques and practical community use*.

DIFFICULTY	COLLECTIVE
●●○○○	Any

15

PLATFORM

AUDIT

Algorithmic Abandonment

When users or media highlight discriminatory behavior, this can build public pressure on the platforms to correct this behaviour. In some cases, platforms might also decide to discontinue harmful algorithms. Johnson et al., 2024 define **algorithmic abandonment** as *a decision made by actors with jurisdiction over the system to discontinue the process of developing, deploying, or using the algorithm due to its (potential) harms*.

DIFFICULTY	COLLECTIVE
●●○○○	Large

Johnson et al., 2024

15 Algorithmic Abandonment

COMMUNITY LENS

Research Gap Hern Microsoft retires Face API features such as infer emotional states, gender, age, smile, facial hair, hair, and makeup.

SCAN FOR SOURCES & MORE

14 User-Facing Resistance Tools

LGBTQIA2S+ EXAMPLE

Several mechanisms are packaged into downloadable or web tools that lower the technical barrier: Glaze (Shan et al., 2023) and Nightshade(Shan et al., 2024) for artists, Fawkes (Shan et al., 2020) and the LowKey (Cherepanova et al., 2021) webtool for facial privacy, and the LLM Data Defenses tool (wagnew3.github.io/LLM-Data-Defenses) for text. These are general-purpose and available to anyone, not built for or specific to LGBTQIA2S+ users. The community relevance lies in who chooses to deploy them, for example trans users adopting facial-privacy tools against recognition to prevent automatic gender recognition, not in the tools themselves being community-specific.

SCAN FOR SOURCES & MORE

13 Harm Reporting Infrastructure

LGBTQIA2S+ EXAMPLE

Both the AI Incident Database (McGregor, 2020) and the AIAAIC Repository (AIAAICRepository) catalog harms affecting LGBTQIA2S+ people, for example AIAAIC's entry on the HRT Transgender dataset that exposes how trans people's data was collected without consent.

SCAN FOR SOURCES & MORE

16

PLATFORM

AUDIT

Demanding Less Discriminatory Algorithms

AI systems often show **model multiplicity** (Black et al., 2022), where several models reach *similar accuracy but make different decisions or behave differently overall*. When models perform equally well but some cause less harm to affected groups, communities can reasonably push for the less harmful ones to be used. Black et al., 2023 argue that service providers should have a *legal obligation to look for and adopt less discriminatory algorithms (LDAs)* that reduce unequal harm across groups.

DIFFICULTY	COLLECTIVE
●●●●○	Medium-Large

Black et al., 2023

17

PLATFORM

INTERVENE

Legal Contestation via Lawsuits

Legal contestation through litigation allows individuals and groups to challenge algorithmic systems in court, seeking remedies for harms that platforms are unwilling to address voluntarily. Unlike technical or community-driven mechanisms, this approach relies on existing legal frameworks (e.g., anti-discrimination, data protection, or consumer protection law) and the authority of courts to bind platform behaviour

DIFFICULTY	COLLECTIVE
●●●●○	Medium-Large

17 Legal Contestation via Lawsuits

LGBTQIA2S+ EXAMPLE

Examples of legal contestation can be seen through class action, ellisgeorge.com (Divino Group LLC et al. v. Google) where LGBTQ+ YouTube creators challenge Restricted Mode filtering and demonetization. GiggleTickleAustralia2026 (Tickle v Giggle), an Australian court finding gender identity discrimination by an AI-gated women-only app which discriminated against trans women. In term of advocacy, Tech records LGBTQ groups signing the ACLU letter demanding regulation of facial recognition.

SCAN FOR SOURCES & MORE

16 Demanding Less Discriminatory Algorithms

COMMUNITY LENS

Research Gap No direct evidence of LGBTQIA2S+ community use of model mutiplicity as legal resistance has been documented in literature.

SCAN FOR SOURCES & MORE