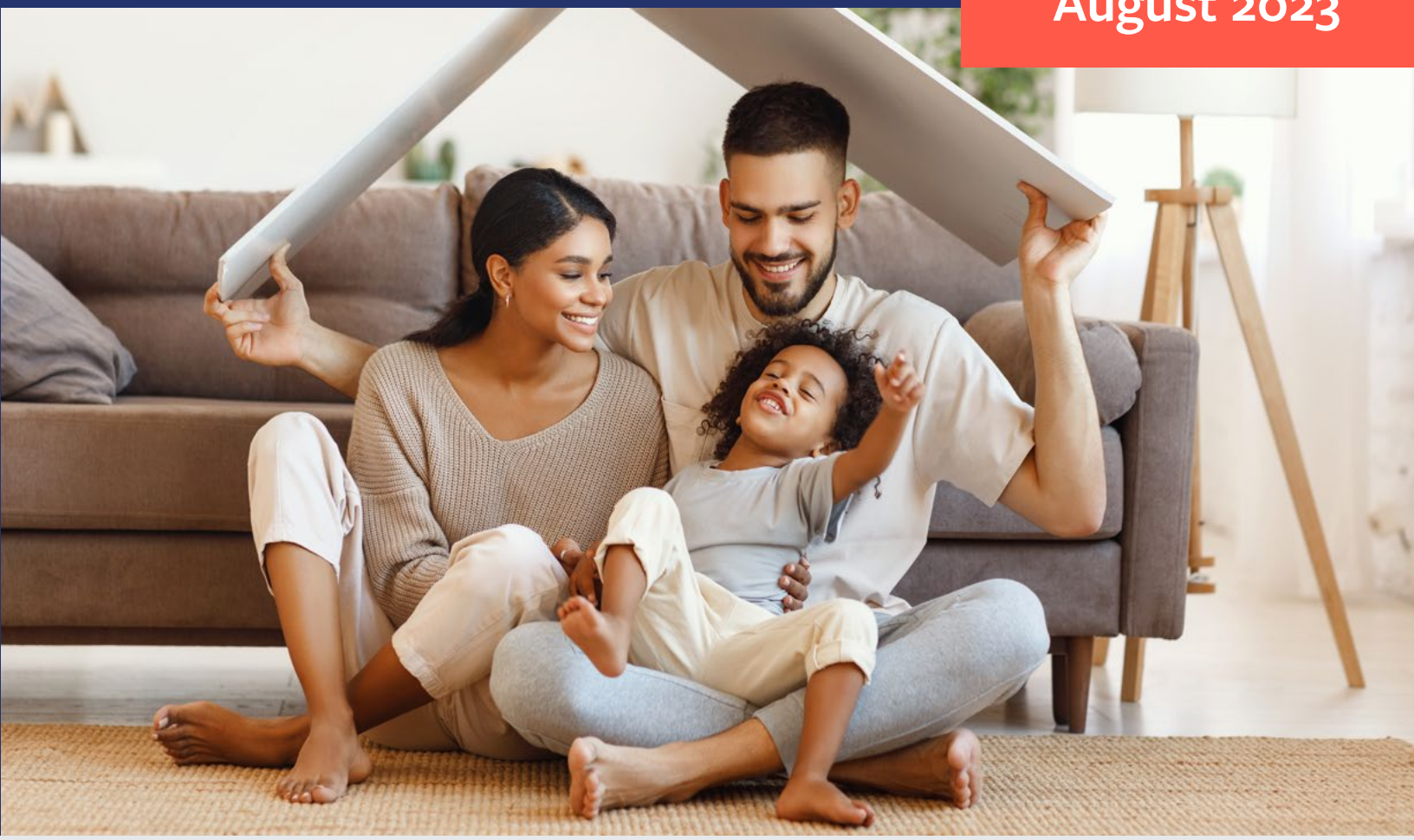
Privacy, Technology, and Fair Housing - A Case for Corporate and Regulatory Action

August 2023



NFHA NATIONAL FAIR HOUSING ALLIANCE

TECHEQUITY COLLABORATIVE

Authors:

Hannah Holloway, Director of Policy and Research, TechEquity Collaborative

Snigdha Sharma, MA., Senior Tech Equity Analyst, National Fair Housing Alliance

Samantha Gordon, Chief Program Officer, TechEquity Collaborative

Dr. Michael Akinwumi, Chief Tech Equity Officer, National Fair Housing Alliance

Contact Information:

- Samantha Gordon, Chief Program Officer, TechEquity Collaborative
  samantha@techequitycollaborative.org
- Dr. Michael Akinwumi, Chief Tech Equity Officer, National Fair Housing Alliance
  makinwumi@nationalfairhousing.org

Table of Contents

Who We Are

The National Fair Housing Alliance's (NFHA) Tech Equity Initiative is a multi-faceted effort designed to eliminate bias in algorithmic-based systems used in housing and financial services, increase transparency and explainability for AI tools, outline ethical standards for responsible tech, advance effective policies for regulating AI tools, and increase diversity and inclusion in the tech field. The goal is to have our "gold standard" of algorithmic fairness adopted by regulators, developers, and consumers of AI-based systems.

TechEquity Collaborative's Tech, Bias, and Housing Initiative examines the growing trend of companies entering the housing market that are promising speed, efficiency, and a modern approach to "traditional" rental or homeownership systems. These new companies are venture-backed and automating housing processes at a massive scale—and they play an increasingly influential role in the economy. As unprecedented capital investment flows into this space, venture-backed companies' winner-take-all approach to growth has the potential to exacerbate inequality in the housing market. The Tech, Bias, and Housing Initiative examines the growth of this industry and its potential harms and biases through comprehensive research, public policy advocacy, and recommendations for corporate practice.

As two connected and partnered organizations, we enter this discussion with different areas of expertise and interest. National Fair Housing Alliance (NFHA), a longstanding, national advocate for fair and accessible housing, enters through the NFHA Tech Equity Team that holds technical expertise in machine learning, computer science, financial modeling for greater access, and fair lending. Conversely, TechEquity Collaborative enters this discussion within our Tech, Bias, and Housing Initiative bringing expertise in housing policy and systems change, with an ongoing focus on illuminating the tech industry's emergence and growing presence in the national housing market.

Acknowledgments

We want to thank the many expert reviewers who guided us in the development of this paper. Specifically, we would like to acknowledge and thank the following individuals who provided feedback and guidance, pushed our thinking, and strengthened our critical lens for this complex set of issues, including:

- David Brody, Managing Attorney of the Digital Justice Initiative at the Lawyers Committee for Civil Rights
- Catherine Bracy, CEO and Founder of TechEquity Collaborative
- Imani Cherry, Attorney at Relman Colfax
- Michele E. Gilman, Venable Professor of Law and Associate Dean of Faculty Research & Development Director, Saul Ewing Civil Advocacy Clinic, Co-Director, Center on Applied Feminism.
- Stephen Hayes, Partner at Relman Colfax
- Debby Goldberg, VP of Housing Policy and Special Projects, National Fair Housing Alliance
- Maureen Yap, Senior Counsel, National Fair Housing Alliance
- Lisa Rice, President and CEO of the National Fair Housing Alliance

1. Introduction

What would you do if you discovered that your application for a mortgage or a rental unit was rejected—and that the denial was due to discrimination? You could, if motivated, appeal the decision with the bank or landlord, file a complaint with your local fair housing center, or hire an attorney to seek a remedy through the legal system. Eventually, though, such efforts would come down to whether or not you could prove that you were denied housing, credit, or a housing-related service because of a protected aspect of your identity such as race or gender. This could include information about whether the housing provider has a history of disproportionately screening out members of a certain group or groups in violation of the Fair Housing Act,[1] Equal Credit Opportunity Act,[2] or another anti-discrimination law.

Increasingly, however, there is not a person making that decision, but an algorithm. These algorithms are built on troves of data to train them to make certain predictions independently, such as whether someone is likely to miss a mortgage or rental payment. The problem with this approach is that there are no consistent, concrete, accountable, or agreed-upon standards at a regulatory—or even industry—level to ensure that the data being used to build the algorithm is appropriately collected, trained, secured, and not creating a discriminatory outcome.

The truth is that these rejections—due to algorithmic decision-making and massive troves of personal data—are not hypothetical; they are already happening, usually without anyone realizing it. These real-world examples have been well documented by a variety of scholars, activists, journalists, and community members; take, for instance, Virginia Eubanks' detailed account of how algorithms entrench poverty and inequality in *Automating Inequality*;[3] or Cathy O'Neil's technical review of how math has been utilized to scale up harm in her book, *Weapons of Math Destruction*;[4] not to mention the series of advocate- and journalist-led investigations into the discriminatory outcomes of mortgage approval algorithms,[5] tenant screening technologies,[6] and price-fixing algorithms that artificially inflate tenants' rents.[7] The lack of agreed-upon standards for these technologies, coupled with questions about who is responsible for the outcome of the algorithms—the company who created the algorithm, the bank that used the algorithm, the landlord who made their decision upon the algorithm's prediction, or the data broker who sold the faulty training data—creates a murky picture of how to ensure consumers' rights are protected and, ultimately, that these tools do not further thwart each person's right to housing.

As these individual experiences grow and these methods proliferate throughout our society, many advocacy groups, policymakers, regulators, companies, and academics are focused on various pieces of this puzzle—whether that's tackling the massive amount of personal data that is often being collected without our consent, determining what information is appropriate for use when developing a model, establishing the best criteria to monitor and audit models, challenging the accuracy of these tools, or advocating to end the surveillance and privacy-invading practices that power many of these tools. There are many urgent areas to address as our world becomes increasingly digital and data-powered.

As stakeholders address discrete aspects of technology's impact on housing and other basic needs, there is a need to incorporate privacy, civil rights, and consumer protections into a unified, fair, responsible, and coherent approach. This has been challenging to date because of a variety of factors, including the lack of transparency from companies on how this data is collected, stored, secured, or processed and its impact on consumers. This lack of transparency is compounded by the power imbalance between consumers and companies, whereby consumers have little knowledge or recourse to control their own data and how it is used to make decisions that can impact their access to housing and other economic resources.

Advocates have urged better privacy protections, limited data collection, and legal requirements to test and ensure that these technologies do not increase discrimination or harm to consumers. Often the question in response is, how can we minimize data collection and protect privacy while ensuring that stakeholders have access to the data required to test these solutions for bias and discrimination? Are these goals conflicted?

Our paper explores these questions and makes the case that the right to privacy and the ability to collect information to ensure these tools are not discriminatory can be achieved together. We can do this by:

1. Adopting a baseline data minimization framework that requires all data collection to be narrowly tailored to a specific, justifiable purpose.
2. Learning from examples where these two tensions were not correctly calibrated, their consequences, and what they can teach us about the balance between privacy and civil rights protections when algorithmic decisions and emerging technology are operating in the real world.
3. Requiring that companies that have received consent to collect sensitive information for anti-discrimination testing apply Privacy-Enhancing Technologies (PETs) to ensure that what data is collected, is secure.
4. Strengthening our regulatory frameworks to more proactively integrate privacy, consumer protections, and civil rights—rather than treat them as discrete areas of the law—so we can improve enforcement, promote effective oversight, and strengthen people's rights in the digital era.

At the conclusion of this paper, we provide recommendations for how companies, policymakers, and regulators can maintain privacy while affirmatively furthering the right to housing for all people in an increasingly digital world. Our recommendations center on the three shifts that we believe are necessary to ensure that the balance of privacy and civil rights is appropriately applied to reduce harm and ensure access to housing for all.

- **Shift responsibility from the individual to companies and regulators** – Rebalance the burden for safety and harm reduction from the individual consumer to the people with the information, power, resources, and ability to effectively redress discrimination and other problems, namely those at companies and regulatory agencies.
- **Strengthen the review of these tools prior to their use on the public** – Technology and algorithms play a significant role in our daily lives and have great power in determining who gets access to housing and economic opportunities. Because of this, we believe that these technologies must meet critical business necessity, non-discrimination, and harm minimization standards prior to deployment and use on the public. It is the responsibility of both companies and regulatory agencies to ensure this happens, much like we require of other products that impact our safety and economic security.
- **Develop an intersectional approach to design and regulate tools and models** – These tools impact us in myriad ways that are often interconnected; for example,when your consumer data is used to train algorithms that later deny you housing, your rights to privacy, housing justice, and non-discrimination might all be at play. Regulatory bodies must require that protections are intersectional, broad, and nimble enough to apply across all sectors of modern life simultaneously. Additionally, companies must take a transparent and comprehensive approach to testing, monitoring, measuring, mitigating harms, and reporting about the impact of their technologies.

Within each of these shifts, we provide detailed recommendations to guide company behavior, policy development, and agency action. These recommendations mirror many of the principles outlined in the White House AI Bill of Rights[8] and provide specific guidance tailored to the goal of balancing privacy, civil rights, consumer protection, and access to housing and credit.

We believe that these steps, taken in concert, can produce a better environment for protecting our civil rights, privacy, and right to non-discrimination in housing and finance, as well as expand people's access to important opportunities. There are many exciting proposals emerging from advocates, policymakers, and governments that knit these concepts and ideas together. Our paper adds to that discussion by outlining how companies can improve outcomes now by applying data minimization, security, and anti-discrimination approaches, and how regulators can address these issues in an intersectional protections framework.

What is a data minimization framework?

Data minimization guides the behavior of companies and governments when they are conducting data collection and/or designing an automated system. Typically, someone using a data minimization framework would tailor their collection of information to what is strictly necessary to perform a given function, and would not retain the information after they have completed said function.[9]

Moreover, data minimization principles often require that a consumer give explicit consent for the tailored use of personal and highly sensitive data, and that companies offer written timelines and data handling policies.[10] We believe that for data minimization and discrimination testing to work, there must be a robust set of guidelines and protections to ensure that the use and scope of consumers' data collection follow data minimization principles. This includes requiring that data collected for one purpose cannot be used for a different purpose or context without conducting an assessment of new privacy risks, the civil rights of the data subjects, and implementing appropriate mitigation measures, which may include express consent. In later sections of this paper, we call for stronger policies, regulations, and enforcement to mitigate these potential harms.

Common Terms

Non-Discrimination: Non-discrimination refers to ending the practice of restricting individuals' access to housing and other critical economic needs on the basis of protected class status (or proxies for that status such as zip code or name), whether intentionally or unintentionally. A variety of federal and state laws outline existing non-discrimination protections.[11]

Privacy by Design: Referenced within the General Data Protection Regulation (GDPR), this term refers to building data protection and privacy rights into the design of a technology, rather than considering those protections as part of the technology's implementation.

Personal and Highly Sensitive Data: As outlined in the White House AI Bill of Rights, enhanced protections and restrictions for data and inferences related to sensitive domains including health, work, education, criminal justice, finance, and for data pertaining to youth should prioritize the consumer's privacy over the institution using that data. In sensitive domains, your data and related inferences should only be used for necessary functions, and you should be protected by ethical review and use prohibitions.[12]

Notice and Consent: "Notice and consent" refers to the current legal structure that undergirds how consumers "consent" to data collection that comes with using that website or service. These notices are generally very long, written by and for lawyers, and are not tailored to consumer comprehension. The term "consent" is misleading in this context and as outlined by many scholars, creates an extreme burden on consumers in a digital age where it is nearly impossible for consumers to have given any real, informed "consent" when using a service.[13]

Automated System: An "automated system" is any system, software, or process that uses computation as a whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities. Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing and artificial intelligence techniques, and exclude passive computing infrastructure. "Passive computing infrastructure" is any intermediary technology that does not influence or determine the outcome of a decision, make or aid in decisions, inform policy implementation, or collect data or observations, including web hosting, domain registration, networking, caching, data storage, or cybersecurity.[14]

Privacy-Enhancing Technologies (PETs): Technologies that allow consumers to protect the privacy of their personally identifiable information. PETs use varied techniques and methodologies to reduce an information system's access to personal data without minimizing or losing functionality.

1.1 Why Do We Need to Consider the Balance of Privacy and Civil Rights in Housing Technology?

Existing privacy and civil rights protections are critical; yet in practice, they do not speak to each other in ways that uphold the individual's rights to both privacy and non-discrimination. As is, policies that advance privacy at all costs (meaning ones that would allow for no data collection or estimation) can also provide cover to companies or systems that have unnecessarily disproportionately adverse outcomes for certain protected groups. On the other hand, if there are no requirements to monitor, measure, and set guardrails around the use of personal data, other harms can emerge. For example, a team of journalists was the first to uncover that the Facebook ad algorithm used personal data—collected for an entirely different purpose—to discriminate against Black users.[15] Nearly three years later, the U.S. Department of Justice charged Facebook with racial discrimination in their targeted housing advertisements.[16]

In 2020, Airbnb announced Project Lighthouse, an ongoing effort to study and address what it calls "the bookings gap"—or disparate experiences that people of color have while using the platform.[17] This came after repeated complaints from impacted users and hosts, advocacy, and external studies that found guests with distinctively Black names were 16% less likely to be accepted than those who applied using typical white names, and that Asian hosts received 20% less money per booking than their white counterparts.[18] As a result of pressure campaigns from Color of Change, Upturn, and other advocates—Airbnb's Project Lighthouse utilized an anonymized dataset that centered and prioritized privacy to determine how user profile photos and other pieces of information are impacting patterns of discrimination within their platform.[19] The result is an effort to balance privacy, non-discrimination, and centering the experiences of people impacted by technology in designing a more just model for their company. While it's not certain that strengthening this balance has solved the problem completely, the effort demonstrates that without an intentional and continual evaluation of the connection between privacy, discrimination, and user engagement, the likelihood of harm and unintended disparities is high. Airbnb has said they will utilize this research to further their efforts to end discrimination on their platform.

Existing law provides other examples of where some degree of data collection can further civil rights. The Home Mortgage Disclosure Act (HMDA) requires many financial institutions to maintain, report, and publicly disclose loan information that can help reveal discriminatory mortgage lending patterns. The data is modified before public release to protect applicant and borrower privacy.[20] The Consumer Financial Protection Bureau's recent rule implementing Section 1071 of the Dodd-Frank Act will create a similar system for small business lending data.[21] However, HMDA and 1071 operate on an opt-in system that requires individuals to voluntarily share their protected demographic information. The groups most likely to experience discrimination in the housing and lending systems are also those with little reason to trust that public systems will use sensitive information

in ways that help, rather than further harm. Given these tensions, it is not surprising that there is a significant number of records missing demographic information entirely from the dataset.[22]

Furthermore the tension between civil rights and privacy has been an ongoing debate—often between the banking industry, which would like to remove more data points under the guise of privacy concerns, and the advocacy community, which believes that HMDA data is critical for ensuring non-discrimination and equal opportunity.[23] It is important to note that HMDA was intended to be a tool for the public, including local officials and community-based organizations, to monitor the activity of lenders and assess mortgage lending patterns in their communities.[24] It is an addition to, not a replacement for, a robust regulatory framework. Industry has argued that releasing too many variables and data points may compromise the identity of individuals and allow them to be identified through publicly available data—however, without a great amount of data points it becomes difficult to test for discrimination in the approval or denial of mortgages. Certain key fields in the HMDA data, such as credit score, are redacted or modified—meant to protect individual privacy—which can present difficulties when testing for fairness and discrimination. One area that is particularly difficult to understand is the role that credit scores play in mortgage approvals or the potential availability of mortgage credit for multifamily buildings that contain affordable units. HMDA data can be used to affirmatively further fair housing by identifying barriers to credit and suggesting needed changes in underwriting, product design, pricing, and marketing. It may help identify particular lenders whose track records in serving borrowers and communities of color can be models for others to follow. In addition, it may be useful in determining how best to target down payment assistance dollars, including new funding currently under consideration in Congress. However, lack of access to specific data points and reporting thresholds have proved to be obstacles to understanding these patterns. A balanced approach that allows for fairness and discrimination testing by releasing the data with privacy-enhancing technologies—rather than simply deleting key fields in the data—offers a way forward.

Creating a balance between privacy and civil rights protections is critical to prevent these technologies from scaling up harm. In each of the examples outlined above it took massive amounts of action on the part of vulnerable and impacted communities, advocacy campaigns, and litigation to mitigate those harms. For these technologies to be deployed in a way that centers privacy and ends discrimination, we must require a set of protections that include a proactive and inclusive method for engaging people who will be impacted by the technology,[25] a robust consideration of privacy, proven methods to enhance privacy and data security, and a clear enforcement and regulatory regime to ensure that these protections are evenly applied and not reliant upon individual consumer action or advocacy alone.

1.2 What About the Government's Use of Data and Algorithms?

Our focus within this paper is on actions that can be taken by companies and governments to better regulate these technologies to increase privacy and protect our civil rights. We acknowledge that while our focus is on the emergence of these technologies within the space of housing and private companies, these technologies and their associated harms can and do exist within the public sector.

Governments have promulgated some of the worst abuses of consumer and civil rights through the rapid implementation of technology, often impacting some of our most vulnerable and marginalized citizens first.[26] There are many advocates and activists pursuing an end to facial recognition technology,[27] surveillance systems,[28] the use of algorithms in the carceral system,[29] the ways that algorithms punish poor people,[30] and more. We credit these experts and activists with a great deal of our understanding of the potential harms of these technologies. While this isn't the focus of our paper, we know that this paper would not be possible without the work that these groups have done to uncover harm and hold governments accountable for their use of these technologies.

2. How Private Companies Currently Use Technology to Improve Privacy

We will not be able to ensure that the right to privacy and non-discrimination are upheld through model enhancements or privacy-enhancing technologies alone, but we need corporate standards—in addition to legal requirements and regulation— to ensure that the models that process our data preserve privacy and minimize harm. Below we outline how companies can improve their methods for privacy and data security while simultaneously allowing for discrimination testing. Our hope is that coupling multiple approaches—both company standards, legal improvements, and strengthened regulations—can create a comprehensive set of protections to reduce harm to consumers and strengthen access to housing for all people.

In this section, we:

- Examine the benefits and limitations of methodologies that technologists use to preserve users' privacy
- Evaluate potential consumer harm against the marketed benefits of data collection for personalization algorithms
- Provide suggestions for incorporating discrimination testing and measurements within the use of privacy-preserving methodologies

There have been countless marketing claims that collecting massive amounts of personal data leads to a better consumer experience. Instead, we contend that consumers would be better served by having more control over when and how their data is collected, what it is used for, when it is sold, and so on.

We believe that a core method for achieving that goal is reducing the amount of personal data that is collected in the first place—while ensuring privacy and civil rights protections at the same time. Examining privacy-enhancing technologies allows us to simultaneously consider data minimization, privacy, and opportunities for studying the impact of these technologies on protected groups to ensure the technologies are not furthering harm.

While privacy-preserving methodologies are offered as technical solutions that use sensitive data but do not put user privacy in jeopardy, there are both benefits and drawbacks to these technologies—drawbacks that technologists must account for *before* implementation.

We cover examples from the three categories of Privacy-Enhancing Technologies (PETs):

- Those that focus on hiding or shielding data
- Those that limit access to parts of the data
- Those that generate data to limit personally identifiable information (PII)

Each of the privacy-preserving technologies present promises and drawbacks. A more comprehensive review of each technology's potential benefits, drawbacks, and opportunities for civil rights testing is outlined in **Appendix A**.

2.1 Privacy-Preserving Methodologies that Focus on Hiding or Shielding Data

2.1.1 Homomorphic Encryption

Homomorphic encryption is a technology that allows computations to be performed on encrypted data without first having to decrypt it. When the resulting computations are decrypted, they are the same as if they had been performed on the unencrypted data.

An example of how homomorphic encryption may be used in the context of housing or lending would be if a lending institution wanted to have personalized loan recommendations for their customers on the basis of their financial history, but did not want to expose the sensitive financial information that may be needed in order to make such personalized loan recommendations. A lending institution, for example, could calculate the debt-to-income ratio (or DTI) for each of its customers while also keeping the income and debt info of those customers confidential.

A lending institution may use homomorphic encryption to encrypt the income and debt data of customers separately so that it is not visible to anyone who is not authorized. The encrypted data can be sent to a data analysis service which is authorized to perform the necessary DTI calculation. The data analysis service can then use homomorphic encryption to calculate DTI on the encrypted data without needing to decrypt the sensitive information (the customers' debt and income information). Once the data analysis has obtained the encrypted result it can be sent back to the lending institution, which can then decrypt the result to obtain the actual DTI for each customer. This allows the lending institution to outsource carrying out the necessary calculations and offer their customers personalized loan recommendations while keeping their sensitive financial information safe.

However, depending on the homomorphic encryption scheme chosen, more resources and time (such as monetary costs and hours spent) may be required to perform operations on homomorphically encrypted data. Additionally, fewer addition and/or multiplication operations may be performed on the data. The limitation on the number of operations, for example, may not allow the use of algorithmic debiasing techniques that allow consumers to have fair and equitable access to housing and other economic opportunities. These resource and time constraints may push technologists to use weaker homomorphic encryption schemes that can compromise consumers' privacy.

2.1.2 Zero-knowledge cryptography or zero-knowledge proofs (ZKPs)

Zero-knowledge cryptography, or a zero-knowledge proof (ZKP), is a cryptography method in which one party (the prover) can prove to another party (the verifier) that a given statement is true without sharing any other information about the statement.

A recent example of zero-knowledge cryptography in the housing sector is in the use of "Identity Mixer" technology developed by IBM. This technology uses zero-knowledge proofs to authenticate the identity of people when they are in the process of buying a home. Traditionally, the homebuying process requires buyers to supply a significant amount of personal information to lenders, real estate agents, and other parties involved in the transaction. This personal information can often include sensitive information such as social security numbers, employment information, and bank account details. However, with identity mixer technology, a buyer can authenticate their identity without revealing sensitive information to third parties.

The buyer first obtains a digital certificate from a trusted third party such as a government agency or a credit reporting bureau that verifies their identity. The buyer then uses the identity mixer to generate a zero-knowledge proof that confirms their identity to the lender or other party without revealing additional personal information. Finally, the lender or other party can verify the zero-knowledge proof without learning any additional information about the buyer, ensuring that the buyer's privacy is protected. It is through using zero-knowledge proofs that the identity maker technology allows different buyers to authenticate their identity during the homebuying process without divulging sensitive personal information. This reduces the risk of identity theft and other privacy breaches while allowing for a secure homebuying process. Unfortunately, a major limitation of any ZKP protocol is that the information received by a receiver is likely still related to an individual; this means it will still be personal data, which means successful attacks on the protocol's implementation process can still expose individuals' sensitive data to leakage risks.

Zero-knowledge cryptography can allow for secure communications and transactions without revealing any sensitive information to the recipient so that sensitive data such as personal information, financial data, or health records can be shared securely and confidentially between parties without the risk of unauthorized access or exposure. ZKPs can also be used to enable data collection for testing algorithmic discrimination while maintaining the privacy of the individuals whose data is being collected. For example, if a researcher wanted to identify patterns of algorithmic bias, ZKPs could be used as a way to verify individuals' identities in the dataset to test for algorithmic discrimination without compromising any of their personal information.

2.2 Privacy-Preserving Methodologies that Limit Access to Parts of the Data

2.2.1 Secure multi-party computation (SMPC)

Secure multi-party computation is a technique for preserving privacy by multiple parties jointly undertaking a calculation over their separate inputs, all while keeping those inputs private.

An example of secure multi-party computation in the insurance sector is when an insurance company wants to calculate the probability of a policyholder making a claim that is based on sensitive personal information, such as the policyholder's medical history or genetic data. This may become difficult if the policyholder does not want to reveal such information to the insurance company due to privacy concerns. In this case, the insurance company can use secure multi-party computation to calculate the probability of a claim being made without seeing the sensitive personal information of the policyholder.

First, the policyholder's data, such as their medical history or genetic data, can be encrypted using their private key. The insurance company then also encrypts the statistical model they might use to calculate the probability the claim has of being made through their private key. The encrypted data is shared between the insurance company and the policyholder. The policyholder and the insurance company can then utilize multi-party computation to carry out the necessary computations on the encrypted data without ever revealing their private keys or the underlying sensitive information. Then once the computation is finally complete, the insurance company is able to present the result; they can share the probability of a claim being made without ever needing the sensitive personal data of the actual policyholder. Though SMPCs provide a possible solution, they may not be suitable for big data processing in real time, as they can be computationally intensive to implement. This makes it a less popular solution in the era of big data technologies and personalization algorithms, and, like homomorphic encryption, a methodology whose limited computational capacity will hinder its widespread application to the many algorithmic processes that intersect with civil rights.

Secure multi-party computation can simultaneously protect privacy and test for algorithmic discrimination because it enables collaboration between many different parties without revealing their sensitive data to each other. For traditional data analysis, the data is collected and analyzed centrally by a single entity, which can create a potential privacy risk since the entity can have access to sensitive personal data. However, with SMPCs, many parties can collaborate and perform analysis on their own data without revealing underlying data to others. Thus, for instance, SMPCs can allow multiple banks to analyze lending data for potential discrimination by utilizing multi-party computation to analyze the data collectively while ensuring each bank's data remains private. Through SMPCs, organizations can collaborate to test for algorithmic discrimination and improve decision-making processes without violating individual rights to privacy.

However, an SMPC protocol can be compromised if an attacker's capabilities and goals are not considered as part of the threat models in the design of its protocol.[31] There are significant and important drawbacks with SMPCs that should be reviewed; see **Appendix A** for more details.

2.2.2 Trusted Execution Environments (TEEs)

A trusted execution environment (TEE) is a method of securing data that prevents unauthorized external entities from altering data within a secure environment (the TEE), while also preventing code in the TEE from being modified by unauthorized entities.

An example of how trusted execution environments may be used in the lending sector would be if a bank wants to offer a loan to a borrower but needs to verify the employment history and income of the borrower. Oftentimes, the bank may request such data from the borrower and utilize it to make a decision to approve the loan or not. Yet this process can easily pose a risk to the privacy of the borrower's sensitive information. Therefore, a trusted execution environment can be used in this instance to securely store and process the sensitive information of the borrower without revealing the information to the bank or any other third party.

This would be done by first encrypting and storing the borrower's income and employment data in the trusted execution environment, which in this case would be within the borrower's device or trusted cloud device. The bank would then send a request to the trusted execution environment to access the data of the borrower. The trusted execution environment would then securely process the request and return the result to the bank without revealing underlying sensitive information. Through this privacy-enhancing technology, the bank can make a loan decision that is based on accurate data without risking the privacy of the borrower. TEEs assist in the need for data sharing in the lending sector while protecting the privacy rights of individual borrowers. They enable secure processing and storage for sensitive personal data and are able to provide a trusted computing environment that is separate from the rest of the system, making it harder for potential attackers to compromise the underlying data. However, one of the biggest drawbacks of TEEs is the security of confidential corporate data and databases as processing in shared environments may pose higher risks such as 'side-channel' attacks—an attack based on metadata from the communication between a TEE and other external computer resources—and leakage of cryptographic keys.

TEEs provide a way to protect privacy while allowing for data collection and testing of algorithmic discrimination by using TEEs to execute algorithms and models within a secure environment that ensures the code and data are protected from data tampering or malicious attacks—without compromising the privacy of the individuals or entities involved. If a financial institution used TEEs to securely sort and process sensitive data such as credit scores or income information, the institution could then use that data to develop and test lending models for discrimination within the TEE. This would ensure the models are protected from unauthorized access or modification from external unwarranted parties and individual privacy is protected.

Scalability can be an issue for big data processing due to limited memory and poor processing power. However, combining TEEs with other privacy-enhancing technologies may help overcome these limitations,[32] Additionally, the security of a trusted execution environment assumes that the environment is isolated, but trusted execution environments are not always isolated in practice. As a result, it is possible to release information from the environment.

2.2.3 Federated Learning (FL)

Federated learning (FL) is a machine learning method wherein multiple parties can train a single algorithm collaboratively, each with its own encrypted dataset.

A use case of federated learning in the housing sector would be when developing models to predict loan default. Federated learning can be utilized to allow many different lending institutions to collaboratively train a machine learning model to predict the chance of loan default while protecting user privacy and data. Each lending institution would train the model based on its own data, such as loan application information and repayment information, and would not need to share the data with external parties. Instead, through federated learning, they would only share the model updates with the other institutions. This would help protect individual user privacy while allowing for more accurate predictions of loan default across many different institutions. A federated learning approach would be particularly useful for smaller lending institutions that may not have access to larger amounts of data on their own. Federating learning can provide a larger pool of outcome data that is more representative and diverse through collaborating with other institutions while maintaining the privacy of their own customers. Though FL has its benefits, the cost of the collection and processing of data, and the limited computational ability of some devices are some of its disadvantages. In addition, FL alone may not be enough to guarantee data privacy because as models update, there may be traces of information left to infer users' personal and confidential data.

2.3 Privacy-Preserving Methodologies that Generate Data to Limit Identifiability of Individuals

2.3.1 Synthetic Data Generation (SDG)

Synthetic Data Generation (SDG) is a process that generates 'artificial' data using synthesis algorithms that replicate patterns and statistical properties of real data.

An example of Synthetic Data in the lending sector might be where banks need to decide on loan approval or denial for individuals. Credit scores based on the applicant's credit history, payment behavior, etc. are typically generated and used as main decision points in the lending sector. Unfortunately, sometimes there may not be enough data to generate an accurate credit score, which makes it difficult to obtain a loan. This is where synthetic generation can play a useful role.

Generative Adversarial Networks (GANs) are example techniques that can be used to generate synthetic data. A typical GAN consists of two components: a generator and a discriminator. The generator attempts to generate artificial data that looks real by learning from the patterns in real data. Once the artificial data has been analyzed, the discriminator can be used to sort the artificial data from the original dataset.

For example, in an experiment to investigate the role of credit score in mortgage underwriting, a GAN can be used to simulate the credit history of a person with similar characteristics to a typical applicant. If the applicant is a newly working professional with a limited credit history, synthetic data can be generated to simulate a credit history to other newly working professionals with similar characteristics such as income, education level, and geographic location. The synthetic data could then be used to generate a credit score for the applicant which would be utilized in the loan approval process. Through using synthetic data, banks may be capable of making more accurate loan approval decisions despite incomplete or limited data. Synthetic data may provide a way forward, but it is important to remember that the utility of SDGs is dependent upon artificial data approximating an accurate proxy for the real data. Additionally, assessing the accuracy of an SDG requires access to real data. The more accurate a synthetic dataset is, the greater its utility—and the higher the risk of exposing confidential or personal data.

For discrimination testing, if a dataset included protected class information such as data about one's race, ethnicity, gender, etc., synthetic data can be generated that mimics the statistical patterns of the original data while replacing the sensitive data with synthetic data points. This allows the original data to not be traced back to an individual and protect their privacy but also allows the algorithm to be tested for unfairness or discrimination. By analyzing the statistical patterns in the new dataset with synthetic data, one can determine if there are biased or discriminatory patterns in the original "real" dataset.

However, synthetic data do not represent real individuals; unless the model trained on synthetic data is used to make business decisions that are causing adverse impacts on consumers, it may be difficult to enforce any privacy laws or other civil rights laws on the basis of the trained model. This can make it extremely difficult and unrealistic to assess the model for fairness and other ethical principles; this is arguably the biggest limitation of artificial data.

2.3.2 Differential Privacy

Differential privacy is a method for measuring the degree of information an output of a computation process reveals about an individual.

A good example of differential privacy occurred in 2017 when the U.S. Census Bureau announced it would use differential privacy to protect the confidentiality of information provided to the agency.[33] As opposed to the Census Bureau's previous system, which utilized a technique based on swapping responses from easily identifiable units, the new differential privacy methodology added carefully structured random values ("noise") to every intermediate computation and then executed "postprocessing" algorithms to make the noise-injected data resemble the data produced by previous Census Bureau methods. Differential privacy allowed the Census Bureau to protect the individuals' privacy of those in the dataset, while still allowing the data to be used for analysis.

It is important to remember differential privacy is not applicable to every problem, and it requires access to real data. For example, differential privacy is not useful for individual-level analysis as its application to individual records will hinder an analyst from being able to gain information specific to individuals; thus differential privacy may only be applicable on a case-by-case basis.

While preserving the privacy of individuals in the dataset by adding noise, differential privacy can also be used to test for discrimination by analyzing the differential privacy bounds on different subgroups within the dataset. So for example, if the differential privacy bounds for a particular subgroup (i.e. Black individuals) are wider than for the overall dataset, it can indicate a higher risk of re-identification for that particular group and this may be a proxy of group discrimination.

3. How Do Our Existing Laws Help Us Balance this Tension?

The quality of our lives is already determined by emerging technologies that impact our privacy, consumer, and civil rights. And yet the rights we have to address those decisions are siloed: the Fair Housing Act and the Equal Credit Opportunity Act can protect someone whose civil rights have been violated in housing or credit and the Fair Credit Reporting Act provides certain protections related to information collected and furnished by consumer reporting agencies.

Moreover, we have no comprehensive federal privacy law. In its absence, a handful of states including California, Colorado, Connecticut, Virginia, and Utah, have privacy laws providing varying levels of protection for the personal information of residents. Not only is there a dramatic imbalance between our privacy laws (of which there are few and no national standards) and our consumer and civil rights protections—but both suffer from the siloed nature in which they have been created.

In this section, we compare existing laws against five principles necessary for a strong, cross-disciplinary protection framework. The laws span landmark civil rights, privacy, and financial reform acts, as well as consumer protections. In assessing how each addresses (or does not address) the principles side-by-side, it reveals where gaps exist in our current regulatory framework, and what future campaigns should consider to ensure that our rights to privacy, civil rights, and consumer protections are woven throughout each sector.

Landscape of Existing Privacy Regulation/Policy

We compare federal consumer protections in housing with state-based and European Union privacy protections because there is no general federal United States privacy law. There are state-based civil rights laws that expand upon federal civil rights protections, but we do not dive into this for the purposes of brevity in this paper. Additionally, our partners at the Lawyers Committee for Civil Rights have highlighted that many state public accommodations laws are broader than federal law and could cover some types of discriminatory data uses. Seven jurisdictions apply their public accommodations laws to online platforms. Since their report was issued in 2020, Washington DC and Nevada have expanded their statutes on this issue. [34]

Landscape of Existing Privacy Regulation/Policy

	Federal Trade Commission Act (1914)	Gramm-Leach-Bliley Act (GLBA) (1999)	Dodd Frank Act/UDAAP (2010)	General Data Protection Regulation (GDPR) (2016)	California Consumer Privacy Act (CCPA) (2018)/California Privacy Rights Act (CPRA - 2020 expansion of CCPA)*	Fair Housing Act (1968)	Fair Credit Reporting Act (1970)	Equal Credit Opportunity Act (1974)	Home Mortgage Disclosure Act (1975)
	Jurisdiction: Federal	**Jurisdiction: Federal**	**Jurisdiction: Federal**	**Jurisdiction: European Union**	**Jurisdiction: California**	**Jurisdiction: United States**	**Jurisdiction: Federal**	**Jurisdiction: Federal**	**Jurisdiction: Federal**
Does it include a data minimization framework or principles?	N/A. The Act was created to protect consumers against unfair business practices - it did so by creating the Federal Trade Commission and intially, by enforcing the Sherman Antitrust Act and the Clayton Antitrust Act. It is notable for its role now as the primary enforcer of federal privacy laws and protections, which it does in part by enforcing the other laws in this chart.	No, GLBA does not limit the types of, or purposes for which, data can be collected. It utilizes an "opt out" framework for sharing "nonpublic personal information" with third-parties that allows entities to transmit data unless and until expressly told not to by data subjects.	No, Dodd Frank and UDAAP have recently been interpreted and updated to give consumers better protection and access to their information, but they protect and monitor the consumer data that financial institutions hold, rather than limiting what they can hold in the first place. Section 1033 of Dodd Frank requires financial service providers to make information available about products or services to consumers. This, however, does not relate to a consumer/data subject's personal data, rather information about financial products and services. In October 2022, the CFPB announced rulemaking on personal financial data rights related to section 1033 to require that financial institutions make consumer financial data available to data subjects or third-party entities at the subject's request. At time of publishing, rulemaking had not been finalized** The Act's UDAAP authority protects against unfair, deceptive, and abusive practices of financial institutions. It August 2022, CFPB announced that entities that do not adequately protect consumer data could be violating UDAAP.	Yes. Personal data must be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed. It must only be kept for as long as the data is necessary for the processing purposes.	No. CCPA/CPRA uses an opt out framework - rather than limiting what businesses can collect in the first place or ensuring privacy by right, it places the responsibility on individual data subjects to withdraw their consent from businesses for the sale and sharing of their personal information, or to request they limit their use of your sensitive information.	N/A. The FHA does not require or entail data collection	FCRA states that people requesting consumers information must have a valid need to do so, but it does not limit consumer reporting agencies to collecting only specific information. Instead it requires certain accuracy, fairness, and privacy standards of the information contained in consumer reports.	Regulation B of ECOA is limited data collection requirements for non-HMDA covered mortgage loans.	CFPB modifies the HMDA data they make public to protect applicant and borrower privacy, but it does not begin from a minimization framework as outlined in this paper.
What are its privacy & security protocols?	N/A. The Act was created to protect consumers against unfair business practices - it did so by creating the Federal Trade Commission and intially, by enforcing the Sherman Antitrust Act and the Clayton Antitrust Act. It is notable for its role now as the primary enforcer of federal privacy laws and protections, which it does in part by enforcing the other laws in this chart. The FTCA UDAP provisions can apply to privacy violations.	The GLBA requires that financial institutions implement security safeguards for consumer information that encompass administrative, technical, and physical protections. Technical safeguards include certain cryptographic and encryption standards.	No, Dodd Frank uses opt-out frameworks, which do not guarantee privacy by right. The Act's UDAAP authority has been interpreted to mean that failing to safeguard data could violate the prohibition on unfair practices; it outlines certain technical security measures that can mitigate risk of violating UDAAP: multi-factor authentication, password management, and regular software updates. While these show certain technical security standards, they do not align with the data standards outlined in this paper.	Yes. GDPR explictly requires appropriate technical measures such as pseudonymisation, encryption, and other data protection principles that safeguard data during the data processing stages.	No, CCPA/CPRA use typical notice-and-consent out-out frameworks, which enable data subjects to reclaim their data but does not guarantee ownership of the data by default. CPRA requires businesses whose processes of personal information presents significant security risks to perform annual cybersecurity audits. At time of publishing businesses did not need to be assessed against privacy perserving methodologies.	N/A	FCRA gives consumers certain rights related to the data held by consumer reporting agencies, but it does not give consumers say over whether those agencies have the information at all. Anyone with a designated valid need (landlords, employers, financial institutions, etc.) can access your information. In the case of employers, FCRA stipulates that consumers must give express written consent for access to their information. Consumers may request all information about themselves contained in consumer reports, and agencies must delete inaccurate or unverifiable information about consumers. The FCRA framework leaves agencies in charge of consumer data, with consumers/data subjects given discrete rights over how it is used, rather than giving consumers primacy in the use and distribution of their information.	N/A	Under HMDA, institutions disclose loan application details including date, loan type, property type, amount, location, applicant demographics, income, and approval information for each loan application, the data that is publicly available is aggregated by banking institution to exclude application date, property addresss, as well as applicant credit score and ethnicity.
Does it require notice & explanation?	N/A. The Act was created to protect consumers against unfair business practices - it did so by creating the Federal Trade Commission and intially, by enforcing the Sherman Antitrust Act and the Clayton Antitrust Act. It is notable for its role now as the primary enforcer of federal privacy laws and protections, which it does in part by enforcing the other laws in this chart.	Requires financial institutions to give customers and consumers a privacy notice that describes the institution's collection, disclosure, and protection practices, including the categories of collected data, disclosed data, and which third-parties information is shared with.	No, rulemaking at time of publishing on Section 1033 proposes requiring that financial institutions make personal consumer financial data available to data subjects or third-party entities at the subject's request but consumers currently do not have clear rights to their data under Dodd Frank.	Yes. Data subjects msut receive transparent information on the purposes and processing of their collected data, as well as information on the data storage period, how to access and delete one's data, how to submit a complaint, and the existence of automated decision-making. Data must be collected for "specified, explicit, and legitimate" purposes.	CCPA/CPRA ensures the right to know, delete, correct, limit, and opt-out of business' data practices. The right to know also includes the right to know how one's personal data is used.	N/A	The Act outlines when information can be shared (and with whom) and outlines when consumers must be notified when an adverse action is taken because of one of these consumer reports.	Requires notice and explanation of adverse action, 12 CFR 1002.9	HMDA data is submitted by financial institutions about applicants. Applicants can self report their ethnicity, race, and sex; in lieu of self-reported data, institutions must report demographic information based on visual observation or surname assumptions. Besides the option to self identify, there is no discretion or notice given to applicants about how their data will be submitted, and no opportunity to change the information.
Is there an explicit non-discrimination provision?	The Act created the Federal Trade Commission, and prohibits unfair competition and deceptive commerce practices. These protections are now commonly upheld through enforcement authorities given to the Commission in laws such as ECOA and UDAP. The UDAP provisions can apply to privacy violations.	The Federal Trade Commission may bring enforcement actions for Privacy Rule violations either through federal court or by examining stated privacy policies for deception or unfairness. The Consumer Financial Protection Bureau also has certain Regulation P enforcement authority under GLBA.	Dodd Frank created the Consumer Financial Protection Bureau, and gave it authority to protect against discriminatory lending and enforce federal fair lending laws. It also changed the disclosure requirements under HMDA, to better monitor for discriminatory lending patterns.	Yes. The protections do not neatly fit into common U.S. protected classes but GDPR prohibits the processing of data that reveals racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, biometric data, health data, and sex data.	CCCCPA/CPRA right to non-discrimination protects individuals from adverse action on the basis of enforcing or exercising their rights under CCPA, but it does not create additional discrimination protections specific to privacy rights.	Yes, the Fair Housing Act prohibits discrimination in housing on the bases of race, color, national origin, religion, sex and gender, familial status, and/or disability.	FCRA does not directly protect consumers against the discriminatory use of their data. Consumer data used in discriminatory ways would be protected by other laws, such as the Fair Housing Act in the case of housing discrimination.	Protects consumers from being discriminated against by lenders on the basis of race, color, religion, national origin, sex and gender, marital status, age, income assistance, and excersing one's rights under other consumer protections laws.	HMDA was enacted in 1975 in part to review the data to identify potentially discriminatory financial lending patterns. HMDA is effectively an accountability and supervisory tool to ensure non-discrimination in FHA and ECOA.
How does it handle enforcement?	The Act created the Federal Trade Commission, the primary enforcement entity for federal privacy laws and protections.	The Federal Trade Commission may bring enforcement actions for Privacy Rule violations either through federal court or by examining stated privacy policies for deception or unfairness. The federal financial regulators also have authority to enforce against the regulated entities.	Dodd Frank created new financial standards, and authorized several agencies to enforce them. The Consumer Financial Protection Bureau authority to enforce against unfair, deceptive, or abusive acts related to a consumer financial product or service. UDAAP largely relies on a consumer complaint and investigation process. The federal financial regulators also have authority to enforce against the regulated entities.	GDPR is an international regulation covering countries in the European Union and entities operating in those countries, enforced by authorities in each country known as Data Protection Authorities (DPAs). Violations are fined. The federal financial regulators also have authority to enforce against the regulated entities.	The California Attorney General enforces the CCPA and CRPA, with the California Privacy Protection Agency holding administrative and jurisdictional power to implement and enforce them. Enforcement of CPRA will not begin until July 2023. CCPA violations carry right-to-cure period and civil penalties up to $7,500 for each intentional violation.	FHA has private right of action and is enforced by various agencies including HUD and DOJ.	The Federal Trade Commission and. the Consumer Financial Protection Bureau have primary enforcement authority of FCRA, which they exercise through legal action against the consumer reporting agencies.	There is a private right of action under ECOA. The CFPB supervises lending and credit institutions for discrimination, and can take public enforcement action on covered entities. The FTC can enforce ECOA for non-CFPB entities.	Institutions that do not submit HDMA disclosures are subject to civil monetary penalties. That said, it is a transparency measure that publicizes potentially discriminatory patterns; it is up to the public to assess the data for those patterns to ensure that the data collected is being applied effectively. CFPB supervises and enforces for non-compliance.

Current Protections are Insufficient

The table above illustrates that the key provisions of each category of protections are often disconnected: the protections in privacy laws pertain to the right to enforce privacy rights without discrimination or retaliation *for exercising those rights* (as opposed to protected class status discrimination). Conversely, privacy protections and enforcement in civil rights statutes are lacking, as evidenced by landlords' reliance on arrest and conviction records despite rulings that such reports disproportionately screen out members of certain classes,[35] or the inaccuracies that follow housing applicants around despite the right under FCRA to correct erroneous consumer information. These two pieces must be connected—privacy and non-discrimination—in order to ensure that new technologies do not further harm vulnerable communities attempting to access—or stay in—housing.

Moreover, the existing protections place a high burden on individuals to take action to enforce their rights. Most privacy laws use opt-out models, meaning entities have a right to collect our information unless and until we say they cannot. In instances of both privacy and housing infringements, individuals can pursue action against bad actors but as these decisions become increasingly digitally enabled, the ability for a housing applicant to know and gather evidence to prove that the decision was discriminatory becomes more difficult.

In the next section, we outline recommendations for a federal privacy approach that incorporates agile civil rights and consumer protections that can stand up to the increasingly digitized systems in which we live.

4. The Future We Envision

Often we think of these technologies and the ways they currently operate as inevitable and ongoing. The moment is ripe to envision a new future, wherein we collectively recognize the scale, the importance, and the potential—both good and bad—of these technologies and what they mean for housing justice and economic opportunity. Reckoning with the imbalance of power between individuals, companies, and regulators is a critical first step in building a better future.

To tackle the imbalance of power, there are three major shifts that we think are immediately required to ensure that our civil rights and our privacy are protected as new technologies emerge and existing technologies grow. These shifts can and should be applied at a company and a regulatory level.

Shift #1: Strengthen the review of these tools prior to their use on the public

Principles that Minimize Harm to the Public: Pre-Deployment Review and Approval

Given the significant role that technology and algorithms play in our daily lives and economic opportunities, technologies must meet design and harm mitigation standards prior to their deployment on the public, much like we require safety testing of products such as pesticides and pharmaceuticals before they are marketed.

How can regulators strengthen review of the tools?

- Require an algorithmic impact assessment[36] [37] prior to the implementation of a new technology or prior to the implementation of a model that would impact access to housing, employment, credit, healthcare, legal rights, or other essential needs. These assessments should be reviewed by a regulatory authority prior to the technology being deployed to the public to document potential model harms, data use, and how the algorithm mitigates those harms.
- Pair pre-deployment assessments with ongoing audits of the algorithms throughout their use. The audits must test outcomes as well as the model and data inputs on an ongoing basis, to monitor for harms and identify risk and harm mitigation opportunities.
- Ensure that data collected for one purpose cannot be used for a different purpose or context without assessing for new privacy risks and implementing appropriate mitigation measures, which may include express consent and regulatory approval.

Principles that Minimize Harm to the Public: Notice and Explanation

Notice and explanation principles underscore the rights of all consumers to their personal data including the right to be informed, the right to correct, delete, and ultimately the right to control when, how, and by whom our data may be used. Notice and explanation principles require clear, digestible information, reasonable response time periods, and the end to deceptive or manipulative practices that allow for wide-ranging consent practices that favor the developer and create unreasonable barriers for consumers to be able to exercise their rights.

How can regulators ensure notice and explanation standards?

- Prohibit companies from obtaining consent in ways that are misleading, manipulative, unnecessarily expensive, or overly burdensome to a consumer.

How can developers implement notice and explanation standards?

- Ensure consumers have the right to be informed and are provided adequate time to exercise their right to opt out when their personal data is being collected, processed, transferred, sold, or utilized in a decision-making process.
- Require that when user data is collected, processed, or utilized in a decision-making process that users have a right to an explanation of the decision-making process and how their data was used in that decision-making. Provide consumers with detailed information to allow them to exercise existing rights under civil rights statutes as part of that explanation.
- Document the source of data that is used in models or in training databases to verify its justifiable purpose and to support explainability that is required under accountability and enforcement.
- Implement the right to access, correct, port, and delete personal data to allow consumers control and agency over the use of their personal data.
- Ensure consumers have the right to withdraw previously given consents, the right to opt out of covered data transfers to third parties, and the right to opt out of targeted advertising, including by global opt-out mechanisms.

Principles that Minimize Harm to the Public: Public Participation

The importance of public and civic participation is threaded throughout our legal and regulatory processes. However, many of our current mechanisms for public participation are exclusive to those with the resources, time, and agency to access them; and often public participation is requested after a decision has been made—providing citizens with a small window to tweak a decision, rather than inform its design. Given that housing is a core human need and the growth of technology and digital regimes in determining who can access housing and capital, public participation is critical in the design of these technologies and systems. These principles, case studies, and lessons learned are outlined in Michele Gilman's paper *Beyond Window Dressing: Public Participation for Marginalized Communities in a Datafied Society.*[38] Many of the lessons and principles shared in that report are reflected below:

How can regulators and developers engage stakeholders?

- Create an inclusive, proactive stakeholder engagement process that centers people who are likely to be impacted by the technology.
- Require public participation by law and develop clear enforcement mechanisms to ensure public participation.
- Create an oversight body with clear requirements for impacted community representation and ongoing accountability if the technology is approved for release.
- Utilize a participatory design process for technology development that brings user experiences and goals into the system's design.[39]

Shift #2: Rebalance the responsibility for redressing harm from the individual to companies and regulators

In lieu of robust federal enforcement action, discrimination and privacy protections often place the burden on the individual to understand their rights, suspect their rights have been violated, collect evidence to support that concern, and then file the complaint with a regulatory body or company to redress harm. The processes to enforce one's rights can often be time-consuming, costly, and difficult for a non-lawyer to decipher. This dynamic requires that an individual, who has fewer resources than a company or regulator and often must request their data from the very companies they have filed complaints against, be the key actor in redressing the potential harms of these products.

In a digital world, where algorithms often operate without transparency or clarity for customers and users, it is unreasonable that an individual who did not develop the algorithm, has no rights to access the model data or training protocols, and has no ability to compel that information from the company should shoulder the responsibility for proving they have been harmed. Additional policy and company changes, outlined below, will strengthen an individual's rights to privacy and non-discrimination. Shifting the onus of proving non-discrimination to companies and regulators will help ensure that the state of our protections is maintained by those with the greatest power and responsibility to do so.

Principles that Shift Responsibility: Data Minimization

Most privacy laws use opt-out models, meaning entities have a right to collect our information unless and until we say they cannot. In contrast, baseline data minimization requirements limit the processing, storage, transfer, and collection of data from the outset, without consumers needing to take any action.[40] Multiple privacy laws, including the GDPR and CPRA, have required entities to practice components of a data minimization framework to enforce data privacy and integrity. Additionally, the White House AI Bill of Rights stresses the importance of data minimization.[41]

How can developers and policymakers ensure data minimization?

- Limit the collection, use, and storage of data to what is necessary to conduct a legitimate business, including ensuring compliance with privacy, anti-discrimination, and other consumer protection laws.[42]
- Outline a taxonomy for what constitutes a legitimate business use.

Principles that Shift Responsibility: Privacy by Design and Data Security

Privacy by design and data security principles move away from a notice and consent framework. In a notice and consent framework, individuals either authorize (or deny) data collection for each website they visit, often skipping over the long, contractual terms outlining what that consent entails. Privacy by design and data security, on the other hand, require that systems are designed with data minimization and privacy as a foundational rule. These systems limit their data collection to justifiable business necessities. They do not sell data without an individual's express consent and utilize heightened privacy-preserving methodologies for all data, in particular enacting stringent security protections around highly sensitive data.

How can regulators ensure privacy by design?

- Engage directly impacted people and implement their feedback.
- Prohibit surveillance or monitoring systems.
- Prohibit the use of data for reasons other than the purpose for which it was collected and received consent.
- Require pre-deployment approval that requires a company to prove that any special-case surveillance monitoring systems are legally justified; exceptions should be regularly audited and reviewed for ongoing approval.
- In special case exceptions for surveillance, companies must also provide pre-deployment impact assessments for approval to ensure that their surveillance or monitoring systems do not discriminate based on protected class status. Furthermore, they must conduct regular discrimination testing to ensure that their tools are not discriminatory in their impact.
- Require heightened data security measures for all data, in particular sensitive data, including government identifiers, biometric and health data, geo or location data, private images, and information about minors.

How can companies ensure privacy by design?

- Engage directly impacted people and implement their feedback.
- Utilize privacy-enhancing technologies to further data security.
- Conduct a data audit to understand what datasets already exist in the organization, which do not fit a business need, and which hold sensitive data.[43]
- Implement a Data Loss Prevention (DLP) strategy at the company level and utilize storage with built-in data protections.[44]
- Implement strong authentication and authorization protocols to verify user credentials and ensure that user privileges are applied correctly.
- Restrict the ability for employees to hoard data and ensure that there are protocols in place to flag unusual data storage and transfer patterns to flag potential data theft.
- Conduct periodic independent evaluations to assess the need of storing each instance of sensitive or personal data.
- Develop and implement a data deletion policy that mirrors both the principles outlined in the data security recommendations as well as those in the data minimization section that calls for the deletion of data after a reasonable amount of time.

Shift #3: Develop an intersectional approach to design and regulation of the tools and models

Recognizing that these tools affect us in ways that are not limited to a particular industry or type of protection (e.g. when your consumer data is used to train algorithms that later deny you housing, your rights to privacy and non-discrimination might all be at play), regulatory bodies must require that protections are intersectional and nimble enough to apply across all sectors of modern life simultaneously.

Currently, the Federal Trade Commission plays an important role in upholding privacy protections. The Consumer Financial Protection Bureau could, with greater authority granted by Congress, also handle many of these concerns and act as the agency responsible for redressing harm associated with a personal data breach or unfair use of personal data. The Department of Housing and Urban Development could then work with these agencies to create and enforce rules addressing the unique ways that privacy and data use affect housing outcomes.

Alternatively, there have been calls to establish a Digital Platform Agency that would regulate the behavior of tech companies and digital actors across a variety of industries.[45] Whatever the mechanism, responsible agencies must have broad authority and the resources to appropriately investigate, examine, pursue corrective action, and regularly measure the impact and efficacy of protections. Additionally, these agencies must be primarily accountable to the consumers who have a right to privacy, due process, knowledge of how their personal data is being used, and recourse for harm.

Finally, in order for agencies to be successful in a broad application of these rights, companies must also understand their responsibilities and how to design, test, and monitor key areas of concern, the methods for which are outlined in the preceding explanation of necessary shifts.

Principles for an Intersectional Protection Framework: Non-Discrimination

Non-discrimination protections within enhanced digital privacy regulations are critical, as technologists and private companies are applying our data to determine everything from who is visiting abortion clinics[46] to which job candidates move forward to the next round. In the absence of meaningful protections over who has our data and how they are able to use it, our information can become the basis for how decision-making systems amplify bias and discrimination. Privacy law must incorporate material anti-discrimination protections, including the right to both understand how data systems treat certain groups differently, as well as support for the right to hold companies accountable for harm. It's important to note that while many companies will fixate on the accuracy of their model or data, accuracy becomes irrelevant if the outcome is discriminatory or disparate in its application.

How can regulators create an intersectional framework?

- Require pre-deployment algorithmic discrimination impact assessments and ongoing audits on actual impact.*
- Continue to make clear in statute and regulation that irrespective of intent, if a company is deploying a tool that has a discriminatory or disparate impact they are liable for that impact.
- Identify new ways to work across agencies to create, clarify, and uphold people's rights that exist at the intersections of data privacy and housing justice.

How can companies create an intersectional framework?

- Engage stakeholders in the audit and evaluation process[47] and require ongoing audits of disparate impact as products are in the market.
- Publicly disclose and report on audit results and methods, as well as steps taken to minimize discrimination and bias within the product design and deployment.
- When evaluating models, go beyond examining accuracy or predictive capacity and employ impact assessments to choose the method that produces the smallest likelihood to further disparate impact.

**See Shift #1: Pre-Deployment Assessment and Approval.*

Principles for an Intersectional Protection Framework: Accountability and Enforcement

The current maze of regulatory authority, discrete protections for consumers in specific sectors, and overall lack of privacy protections in digital markets has resulted in a fractured and inadequate apparatus for everyone. In order to ensure data minimization and strong consumer protections we need an expansive set of policies to enact those rights in addition to a well-equipped and well-resourced regulatory agency that proactively and consistently engages to reduce harms that can result from these products.

How can lawmakers strengthen accountability and enforcement for regulatory agencies and the public?

- Expand or create an agency with the mandate, resources, and technical capacity to effectively implement comprehensive privacy statutes—that incorporate clear consumer and anti-discrimination protections—and hold decision-makers, industries, and governments accountable for violations of those laws.
- Equip the agency(s) with the technical capacity to rigorously test PETs in the production systems of private companies and conduct audits of those systems.
- Create complaint and response programs at regulatory agencies to investigate consumer-sourced reports of data discrimination and privacy violations. Ensure that the agency has the resources needed to quickly follow up on claims and provide straightforward and easy-to-understand directions to consumers to ensure they can enact their rights.
- Require companies to immediately begin reporting to consumer agencies and housing authorities on discrimination testing outcomes.
- Ensure that individuals may exercise private rights of action (e.g. the ability for an individual to pursue litigation) when a company has violated their rights.
- Examine evidence of companies' claims about the predictive nature of their automated decision-making systems and ensure that those claims are true, justified, and not deceptive to consumers.
- Provide information to the public on the frequency and types of complaints by company, similar to what consumers can access for other sectors of the economy with an impact on an individual's safety or economic security (e.g. childcare facilities,[48] long-distance moving companies,[49] consumer complaint database[50]).

5. Conclusion

The world is digital. Everything we do is tracked, monitored, stored, and often monetized—for good, bad, and everything in between. An anti-discrimination framework will only get us so far. If a landlord or property manager is using video technology or online rental management software to surveil tenants, people in the building could be equally affected by monitoring algorithms—irrespective of race, gender, or other protected characteristics—but the collection, storage, use, and transmission of personal data pose ethical and humanitarian questions we must address.

We must ensure that our rights are not so narrowly considered that they can be dismissed by a model developer who hasn't considered the implications of the data they are using—or can only be enforced through lengthy, complex sectoral protections that have limited utility in the modern, intersectional, heavily digitized context in which we live our lives. We need an integrated, federal approach to privacy and technology, and it must extend to our most basic need: housing.

Appendix A - Privacy-Enhancing Technologies

Appendix A provides a more in-depth explanation and set of examples for each of the Privacy-Enhancing Technologies (PETs) outlined within the paper.

Homomorphic Encryption:

Homomorphic Encryption (HE) allows for computations on encrypted data—meaning that you can analyze data to make decisions without anyone actually seeing the individual data itself. HE makes it easier to process data because you do not have to trust the people interacting with the data in order to keep it secure. It is secure by the nature of its encryption.[51] With the decryption, the result of the calculations will be identical to results that would have been produced if the computations were performed on original plaintext data.

HE uses a public key-generation algorithm to generate a pair of private[52] and public[53] keys and an evaluation key which is needed to provide computation on the encrypted data when it is shared with a processor. Keys are like passwords that are native to that HE. Public and private keys work together to encrypt and decrypt data that occupies a network. The public key can be shared with anyone as needed while the private key should optimally solely be known to the owner.[54] The entity that retains access to the private key can decrypt the results of the computation performed by the processor.

HE's promise is that any entity that holds only the public and evaluation keys cannot gain access to or learn about the encrypted data.[55] HE allows researchers to ensure their data is secure while also keeping private data private. Additionally, it allows for leveraging of shared computing resources and fosters collaboration with third parties without revealing the results or the sensitive nature of the data. This is incredibly important, especially in the wake of recent regulations such as GDPR (General Data Protection Regulation) and CCPA (California Consumer Privacy Act), which hold strict penalties and fines for the misuse of handling, transferring, or collecting data.[56] Homomorphic Encryption protects privacy by allowing sensitive data to be encrypted before being sent for analysis or computation, thus preventing unauthorized access to data while also enabling data collection and analysis to detect algorithmic discrimination by allowing computations to be performed on encrypted data without decrypting it. This allows for the analysis of sensitive data while also maintaining the privacy and security of that data.

HE can help data providers ensure security and confidentiality as its encryption is at rest, in transit, and during computation, thus minimizing the risk of data breaches if they occur. In addition, as HE does not require data alteration prior to encryption; it provides a level of guarantee that the computation on homomorphically encrypted data would be the same as if it were performed on unencrypted data.

At present, HE is a promising tool that balances the need to process protected class information with best privacy practices but is not yet robust enough to be the only methodology for ensuring civil rights in privacy law.

In addition, there are important open questions about HE's underlying encryption strength as well as recent analysis suggesting that the method runs the risk of leaking privacy information and can be compromised. Furthermore, organizations cannot run ad-hoc or discovery-based queries with its methodology. Full homomorphic encryption is new and was only established as recently as 2009, thus it is a slower process in a climate where efficiency is optimal. The technique would need to be refined before its wide use and implementation to cater to a competitive market and improve user experience while also protecting user privacy.

Zero-Knowledge Cryptography or Zero Knowledge Proofs (ZKPs):

Zero-knowledge cryptography or zero-knowledge proofs (ZKPs) is a protocol where one entity called a prover (usually an individual) can prove to another entity called a verifier that they are in the possession of a secret.[57] For example, a prover can use a ZKP protocol to prove to an honest verifier, another party, that they are over 18 without conveying any information like their birthday to the verifier apart from the fact that the statement is true. A ZKP protocol relies on completeness (the verifier can verify if the prover is telling the truth), soundness (if the information is false, then the verifier must be able to refute the prover), and zero knowledge (the verifier does not receive more information than those provided by the prover). ZKPs do not require complete encryption; the user's private information is not revealed, and it is a common process already widely implemented in the finance sector, blockchain, online voting, authentication, and machine learning. Though widely liked and used in the market, its widespread implementation has uncovered some major issues.[58] ZKPs have high control over data handling: so one's files will not only be encrypted but also stored in a cloud.[59]

ZKPs can help to achieve data protection compliance with data minimization principles as they only share required information with a verifier and the security principle as confidential data does not have to be shared with other parties, thus protecting sensitive information that could increase one's likelihood of discrimination. This protection of sensitive information can allow one to test for algorithmic discrimination by only testing required information, thereafter using ZKPs to authenticate individuals' identities without compromising their privacy.

However, when a ZKP is applied to the design and implementation of a computation process, there is a need to assess whether the uncertainty associated with the protocol is sufficiently low enough that its benefits outweigh the potential risks such as a data breach it may pose to consumers. ZKPs are also commonly used in the crypto-asset community which holds high climate costs. Published approximations of global electricity usage for crypto assets are anywhere from 120 to 240 billion kilowatt-hours/year,

an amount that overtakes the total annual electricity usage of many countries such as Australia and Argentina; therefore, such environmental costs must be taken into consideration.[60]

Zero Knowledge Cryptography can allow for secure communications and transactions without revealing any sensitive information to the recipient so that sensitive data such as personal information, financial data, or health records can be shared securely and confidentially between parties without the risk of unauthorized access or exposure. It can also be used to enable data collection for testing algorithmic discrimination while maintaining the privacy of the individuals whose data is being collected. For example, if a researcher wanted to attempt to identify patterns of algorithmic bias, ZKPs could be used as a way to verify individuals' identities in the dataset to test for algorithmic discrimination without compromising any of their personal information.

Secure Multi-Party Computation (SMPC):

Secure multi-party computations (SMPCs) distribute computation across multiple parties so no individual party has access to the data of other parties and provide a mechanism that enables the computation of encrypted data without the decryption of underlying values. SMPCs offer a method that aims to maintain data privacy and data utility. SMPCs aim to eliminate the tradeoff between data privacy and data utility since complete data does not need to be shared with third parties or model owners to be utilized. It also eliminates the risks of data breaches and misuse stemming from data collection.[61] Implementation of a SMPC does not reveal intermediate information during computation thereby providing a higher security level. Lastly, SMPC can help demonstrate the security principle of a privacy policy; because the complete data is not known to all the parties involved in the computation process, it can demonstrate the data minimization principle as no party learns beyond what is shared with them. This makes SMPC stand out as a privacy implementation that can help minimize the risk of personal data breaches when performing computations with multiple parties.

Additionally, secure multi-party computation can simultaneously protect privacy and test for algorithmic discrimination because it enables collaboration between many different parties without revealing their sensitive data to each other. For traditional data analysis, the data is collected and analyzed centrally by a single entity, which can create a potential privacy risk since the entity can have access to sensitive personal data. However, with SMPCs, many parties can collaborate and perform analysis on their own data without revealing underlying data to others. Thus, for instance, SMPCs can allow for multiple banks to analyze lending data for potential discrimination by utilizing multi-party computation to analyze the data collectively while ensuring each bank's data remains private. SMPCs enable multiple parties to collaborate and share data while protecting the privacy of individuals and their sensitive information. Through SMPCs, organizations can collaborate to test for algorithmic discrimination and improve decision-making processes without violating individual rights to privacy.

However, an SMPC protocol can be compromised if an attacker's capabilities and goals are not considered as part of the threat models in the design of its protocol.[62] This compromise can not only result in the reverse engineering of the computation based on SMPC's secret shares, but it can also result in the reconstruction of the complete input data, thereby violating the security and data minimization requirements for a strong privacy policy. Additionally, SMPCs only reveal the output of a computation and this limits the transparency of the system; if the output is personal data, a separate privacy module will be required to prevent access to the personal data.

Trusted Execution Environments (TEEs):

Trusted Execution Environments are isolated areas on the central processing unit (CPU) of a computer device.[63] It ensures that data is stored, processed, and protected in a secure environment. TEEs provide protection for any connected "thing" by enabling end-to-end security, protected execution of authenticated code, confidentiality, authenticity, privacy, system integrity, and data access rights. The isolation of a TEE from the rest of the operating system ensures that the operating system or hypervisor (a process that partitions a computer's hardware from its OS and applications) can neither read the code nor the data in the TEE. This design allows a TEE to provide secure communication with applications external to it.[64]

Applications that sit within the TEE are known as trusted applications. The data and code stored on and executed by trusted applications are protected and interactions made (whether between applications or the device and end user) are securely executed. Some benefits of TEEs include how they can secure peripheral access, secure communication with remote entities, and allow for trusted device identity and authentication. Using a TEE provides a higher level of trust in data and code stored in the environment relative to working directly from the main OS. Additionally, TEEs do not suffer from encryption overhead as the actual computation is performed on unencrypted data and there is no need to add noise to the data.

TEEs assist with data protection compliance by limiting data processing to a specific part of the CPU with zero access to external code or external computer nodes. This provides a level of data integrity, data confidentiality, code integrity, and code confidentiality while protecting both data and code from exposure to attacks from bad actors. In addition, TEEs can help with compliance with accountability principles by providing evidence of the steps taken to mitigate privacy risks.

TEEs provide a way to protect privacy while allowing for data collection and testing of algorithmic discrimination by using TEEs to execute algorithms and models within a secure environment that ensures the code and data are protected from data tampering or malicious attacks, all without compromising the privacy of the individuals or entities involved. Thus if a financial institution used TEEs to securely sort and process sensitive data such as credit scores or income information, the institution could then use that data to develop and test lending models for discrimination within the TEE to ensure the models are protected from unauthorized access or modification from external unwarranted parties and individual privacy is protected.

However, scalability can be an issue for big data processing due to limited memories and poor processing power, although combining TEEs with other PETS may help overcome these limitations.[65] Additionally, the security of a trusted execution environment assumes that the environment is isolated but trusted execution environments are not always isolated in practice, and as a result, it is possible to release information from the environment.

Federated Learning:

Federated Learning (FL), sometimes referred to as collaborative learning, is a privacy-preserving methodology that can help protect the privacy of training data by having the data scientist train models locally and then upload the updated parameters to a central model. For federated learning, the model moves to data rather than the data moving to the model, meaning training is occurring through consumer interaction with end devices.[66]

This approach minimizes risks of unfairness or algorithmic discrimination, and it can prevent issues with single points of failure.[67] Federated learning can minimize risks of unfairness or discrimination because first data stays on user devices or individuals, which protects the privacy of individuals and minimizes risks of unfairness or bias that can arise from centralizing data. Additionally, instead of raw data being shared, only model data is shared. Federated learning enables individual devices to train the model locally and only share model updates with the central server instead of sharing raw data, which also helps minimize the risk of unfairness or bias that can arise from centralizing data.

There are two types of Federated Learning, centralized FL (cFL) and decentralized FL (dFL). For centralized federated learning, a central server is utilized to arrange the different steps of the algorithm and coordinate the partaking nodes during the learning process. Node selection at the beginning of training and for aggregation of the received model is the server's responsibility. Due to how all selected nodes must send updates to a single entity, the server can become a bottleneck of the system.[68] For decentralized federated learning, the nodes can coordinate themselves to reach the global model. Such a setup hinders single-point failures when updates on the model are exchanged between interconnected nodes without the need of a central server.[69] FL can prove to be beneficial because computation is moved to the devices of end users, increased access to data can help increase the model accuracy and fairness, model developers can learn multiple models simultaneously at a reduced cost, model developers can train models using private and sensitive information without handling the data, and it helps data stewards remain compliant with data protection regulations such as GDPR.

In Federated Learning, model inversion and membership inference attacks are some of the immediate threats to inferred features[70] and patterns learned in a federated environment. For example, an attacker may observe the patterns identified in a federated learning system and then use that knowledge to extract personal information, compromising the privacy of the individuals represented in the training data.

Lastly, as the training process is exposed to multiple parties in a federated environment, an attacker can exploit the changes in model updates over time, observe a specific model update to inject malicious intents into the global model, or manipulate the model to the advantage of a certain demographic represented in the training data.

Synthetic Data Generation (SDG);

Synthetic Data Generation (SDG) is the use of data synthesis algorithms to produce artificial data which replicate patterns and statistical properties of real data. For clarity, real data means data that represent actual humans or objects in nature while artificial data or synthetic data only represent the patterns and statistical properties of their underlying data. A popular example of SDG is Generative Adversarial Networks (GANs), a privacy-preserving methodology that utilizes two neural networks—a generator and a discriminator—trained to reproduce characteristics and structure of the underlying real data to generate synthetic data.[71] The generator generates the data while the discriminator cross-checks the input with the real output. GANs can produce high-quality realistic results, objects are generated fairly quickly once the model is trained through GANs, and GANs can consistently spit out realistic outputs that look similar to the original real data.[72] Therefore, synthetic data protects privacy because one can replace sensitive information with synthetic data points that do not contain identifiable information. For example, if a dataset included protected class information such as data about one's race, ethnicity, gender, etc., synthetic data can be generated that mimics the statistical patterns of the original data while replacing the sensitive data with synthetic data points. This allows the original data to not be traced back to an individual and protect their privacy while still allowing the algorithm to be tested for unfairness or discrimination. By analyzing the statistical patterns in the new dataset with synthetic data, one can determine if there are biased or discriminatory patterns in the original "real" dataset.

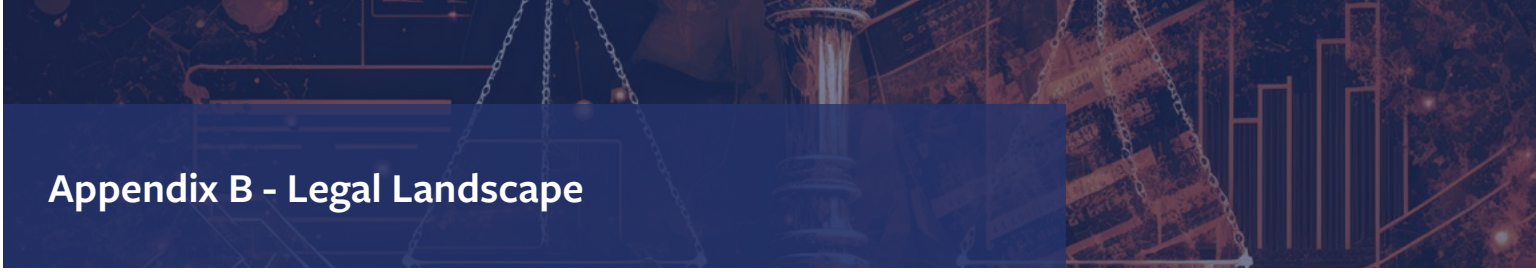
Difficulties for synthetic data includes how data synthesis algorithms may be unstable during training which would increase training time, computational costs, and the cost of reproducing or replicating the data. Lastly, synthetic data do not represent real individuals, unless the model trained on synthetic data is used to make business decisions that cause adverse impacts on consumers, it may be difficult to enforce any privacy laws or other civil rights laws on the basis of the trained model, as it is more difficult to confirm that the trained model is causing harms to the people represented in the training data. This can make it extremely difficult and unrealistic to assess the model for fairness and other ethical principles. This is arguably the biggest limitation of artificial data.

Differential Privacy:

Differential privacy can be used to measure how much information the output of a computation reveals about an individual. Differential privacy may be described as standard to manage and quantify risks for which a plethora of technological tools are conceived.[73] Differential privacy works in a way so that minimum distraction in the information from the database is introduced and the distraction is big enough to protect privacy but also limited enough so information provided to analysts is useful.[74] More simply, differential privacy introduces noise into the dataset to form data anonymization; this allows data experts to carry out statistical analysis without identifying any kind of personal information. Such datasets may contain information on hundreds of thousands of individuals to help solve public issues, but will still protect the information about the individuals. Differential privacy applications range from recommendation systems to location-based services and social networks. For example, the U.S. Census Bureau used differential privacy when collecting personal data from individuals for the 2020 US Census to prevent matching between an individual's identity, their data, and a specific data release.

Differential privacy has many benefits over more traditional privacy techniques such as protecting access to perfect data from attackers. Through differential privacy, a differentially private computation for each query can be applied, leading to separate answers for the same query by different researchers, thereby making it harder to identify the identity or personal information of the records in the query's result set.[75] The addition of random noise guarantees 'plausible deniability'[76] of a particular individual's personal data being in the dataset. Lastly, differential privacy provides a solid quantifiable measure of privacy guarantee by using the concept of "epsilon" or ε, which determines the level of added noise. Epsilon is also known as the "privacy budget" or "privacy parameter". By adjusting the privacy budget, data aggregators can control the level of privacy relative to how sensitive the dataset may be.[77] While preserving the privacy of individuals in the dataset through adding noise, differential privacy can also be used to test for discrimination by analyzing the differential privacy bounds on different subgroups within the dataset. So for example, if the differential privacy bounds for a particular subgroup (i.e. Black individuals) are wider than for the overall dataset, it can indicate a higher risk of discrimination for that particular group.

A major drawback of differential privacy is that it is inapplicable to small data because the inaccuracy that is introduced in differential privacy is capable of being overlooked in large datasets but not for small ones. If the dataset is small, noise added by differential privacy may affect any analysis founded on it. In addition, there is currently no guidance on how to select the privacy parameter ε due to a number of reasons, including lack of agreement over the optimal level of ε that will make the data to be concurrently useful and private.[78] For example though ε= 0 may be the perfect privacy case, it also changes the original data and makes it useless. In the case where differential privacy applications become further popular, more guidelines to optimally reach levels of data distortion for privateness and usefulness for different use cases may be settled in the future.



Appendix B - Legal Landscape

Appendix B provides a brief description of laws that are referenced within the policy table and overall paper.

General Data Protection Regulation (GDPR)

Category: Privacy regulation

The GDPR deals first and foremost with the data collection, processing, and security of data subjects in the European Union. It does this by tackling the internal structures and safety of the company's processes and employee behavior. These protections are systemic and extend to ensure that data subjects explicitly give consent (and define what consent means under the law) and to increase the rights of data subjects to retrieve and block the use of their personal data.

California Consumer Privacy Act (CCPA)

Category: Privacy regulation

The CCPA, and the subsequent California Privacy Rights Act, build upon the lessons and structures outlined in the GDPR and are the strongest laws in the United States on the privacy and digital rights of consumers. The CCPA is comprised of four main components: the right to know what information a business collects about you and what it is used for, the right to delete personal information, the right to opt out of the sale of your personal information, and the right to non-discrimination for exercising rights under CCPA.[79]

California Privacy Rights Act (CPRA)

Category: Privacy regulation

CPRA amended the CCPA by additionally giving consumers the right to correct information, opt out of automated decision-making systems, restrict sensitive personal information, and more. Together, CCPA and CPRA make up the most robust privacy laws that currently exist in the United States.

Federal Trade Commission Act, Section 5

Category: Privacy regulation and consumer rights

Section 5 of the Federal Trade Commission Act prohibits unfair or deceptive practices in the marketplace. The FTC can bring enforcement actions to protect consumers' privacy and personal data. The FTC also has the authority to enforce sector-specific laws and often takes action when issues intersect with other laws on this list like the Equal Credit Opportunity Act, the Fair Credit Reporting Act, and others. The Commission is a primary enforcer of privacy and consumer rights at a federal level; their authority allows them to uniquely address harms that develop within emerging technologies and business models.[80] Federal financial regulators also have the authority to supervise and enforce compliance with UDAP with respect to regulated entities.

Unfair, Deceptive, or Abusive Acts or Practices (UDAAP)

Category: Consumer rights

The Consumer Financial Protection Bureau (CFPB) was granted authority to enforce a provision within the Dodd-Frank Act that makes it unlawful for any provider of consumer financial products or services to engage in any unfair, deceptive or abusive act or practice. It also provides CFPB with the authority to detect and assess risks to consumers and markets for consumer financial products and services.

Fair Housing Act (FHA)

Category: Civil rights regulation - housing

The FHA protects people from discrimination when they are renting or buying a home, getting a mortgage, seeking housing assistance, or engaging in other housing-related activities on the basis of race, color, religion, sex, national origin, familial status, and disability.[81]

Equal Credit Opportunity Act

Category: Civil rights regulation - financial & credit protections

The ECOA aims to ensure that creditors do not discriminate on the basis of protected class —race, color, religion, national origin, sex, marital status, age, receipt of public assistance, or good faith exercise of any rights under the Consumer Credit Protection Act.[82] [83]

Fair Credit Reporting Act

Category: Consumer rights

The FCRA protects the information and data that is collected by consumer reporting agencies like credit bureaus and tenant background screening services. The Act outlines when information can be shared (and with whom) and outlines when consumers must be notified when an adverse action is taken because of one of these consumer reports. It allows consumers to correct erroneous data and requires consent before information can be shared with employers.[84]

Endnotes

[1] See, Housing Discrimination Under the Fair Housing Act. (n.d.). HUD.Gov / U.S. Department of Housing and Urban Development (HUD). Retrieved December 16, 2022, from https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_act_overview

[2] See, What protections do I have against credit discrimination? Consumer Finance Protection Bureau (CFPB). Retrieved April 25, 2023, from https://www.consumerfinance.gov/fair-lending/#:~:text=What%20is%20credit%20discrimination%3F,practices%20in%20home%20financing%20illegal.

[3] Automating Inequality. (2017, August 19). Virginia Eubanks. https://virginia-eubanks.com/automating-inequality/

[4] O'Neil, C. (2017). Weapons of Math Destruction by Cathy O'Neil: 9780553418835 | PenguinRandomHouse.com: Books. PenguinRandomhouse.com. https://www.penguinrandomhouse.com/books/241363/weapons-of-math-destruction-by-cathy-oneil

[5] Martinez, E., & Kirchner, L. (2021, August 25). The Secret Bias Hidden in Mortgage-Approval Algorithms – The Markup. Themarkup.org. https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms

[6] TechEquity. (2022, February 23). Tech, Bias, and Housing Initiative: Tenant Screening. TechEquity Collaborative. https://techequitycollaborative.org/2022/02/23/tech-bias-and-housing-initiative-tenant-screening/

[7] Little, H. V. C. (n.d.). Rent Going Up? One Company's Algorithm Could Be Why. ProPublica. https://www.propublica.org/article/yieldstar-rent-increase-realpage-rent

[8] See, Blueprint for an AI Bill of Rights, Office of Science and Technology Policy (OSTP). Retrieved April 25, 2023, https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[9] European Data Protection Supervisor, Glossary. Website. Retrieved 2.2.2023, https://edps.europa.eu/data-protection/data-protection/glossary/d_en

[10] White Housing Office of Science and Technology Policy, AI Bill of Rights, Data Privacy. Website. Accessed on 2.2.2023 https://www.whitehouse.gov/ostp/ai-bill-of-rights/algorithmic-discrimination-protections-2/

[11]HUD.gov / U.S. Department of Housing and Urban Development (HUD). (2018). Hud.gov. https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_and_related_law

[12]The White House. (2022). Blueprint for an AI Bill of Rights. The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[13]Cate, F. H., & Viktor Mayer-Schönberger. (2013). Notice and Consent in a World of Big Data. Digital Repository @ Maurer Law. https://www.repository.law.indiana.edu/facpub/2662/

[14]Ibid, The White House. (2022). Blueprint for an AI Bill of Rights.

[15]ProPublica. (2017, November 21). Facebook (Still) Letting Housing Advertisers Exclude Users by Race. ProPublica; ProPublica. https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

[16]U.S. charges Facebook with racial discrimination in targeted housing ads. (2019, March 28). Reuters. https://www.reuters.com/article/us-facebook-advertisers-idUSKCN1R91E8

[17]AirBnB. "Introducing Project Lighthouse to Uncover, Measure, and Overcome Discrimination." A New Way We're Fighting Discrimination on Airbnb, https://www.airbnb.com/resources/hosting-homes/a/a-new-way-were-fighting-discrimination-on-airbnb-201. Gilheany, John, et al.

[18]"The Model Minority? Not on Airbnb.Com: A Hedonic Pricing Model to Quantify Racial Bias against Asian Americans." Technology Science. techscience.org, https://techscience.org/a/2015090104/. Accessed 27 June 2023.

[19]Measuring discrepancies in Airbnb guest acceptance rates using anonymized demographic data The Airbnb anti-discrimination team. (n.d.). https://news.airbnb.com/wp-content/uploads/sites/4/2020/06/Project-Lighthouse-Airbnb-2020-06-12.pdf

[20]Consumer Financial Protection Bureau, Data Research, Home Mortgage Disclosure Act. Website. Accessed 2.2.2023 https://www.consumerfinance.gov/data-research/hmda/

[21]CFPB Finalizes Rule to Create a New Data Set on Small Business Lending in America. (n.d.). Consumer Financial Protection Bureau. https://www.consumerfinance.gov/about-us/newsroom/cfpb-finalizes-rule-to-create-a-new-data-set-on-small-business-lending-in-america/

[22]Richardson, J. (2022, August 25). The Critical Need to Address Missing Data in HMDA» NCRC. https://ncrc.org/the-critical-need-to-address-missing-data-in-hmda/

[23] NCRC. (2018, January 23). Rebuttal to personal privacy of HMDA in a world of big data» NCRC. https://ncrc.org/rebuttal-personal-privacy-hmda-world-big-data/

[24] The Home Mortgage Disclosure Act | Consumer Financial Protection Bureau. (2019, March 29). Consumer Financial Protection Bureau. https://www.consumerfinance.gov/data-research/hmda/

[25]Gilman, M. E. (2022, November 2). Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4266250

[26]Opinion | When an Algorithm Helps Send You to Prison. (2017, October 26). The New York Times. https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.htm

[27]EFF. (2019, March 7). Face Recognition. Electronic Frontier Foundation. https://www.eff.org/pages/face-recognition

[28]Wessler, F. W.-J., Nathan Freed. (2023, January 18). How the Arizona Attorney General Created a Secretive, Illegal Surveillance Program to Sweep up Millions of Our Financial Records | News & Commentary. American Civil Liberties Union. https://www.aclu.org/news/privacy-technology/how-the-arizona-attorney-general-created-a-secretive-illegal-surveillance-program

[29]Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[30]"Automating Inequality": Algorithms In Public Services Often Fail The Most Vulnerable. (n.d.). NPR.org. https://www.npr.org/sections/alltechconsidered/2018/02/19/586387119/automating-inequality-algorithms-in-public-services-often-fail-the-most-vulnerab

[31]Du, W., & Atallah, M. J. (2001, September). Secure multi-party computation problems and their applications: a review and open problems. In Proceedings of the 2001 workshop on New security paradigms (pp. 13-22).

[32] Koeberl, P., Phegade, V., Rajan, A., Schneider, T., Schulz, S., & Zhdanova, M. (2015, August). Time to rethink: Trust brokerage using trusted execution environments. In the International Conference on Trust and Trustworthy Computing (pp. 181-190). Springer, Cham.

[33]Garfinkel, S. (2022). Differential privacy and the 2020 U.S. census.

[34] Brody, David, and Sean Bickford. Discriminatory Denial of Service: Applying State Public Accommodations Laws to Online Commerce. Lawyers' Committee for Civil Rights Under Law, 2020, https://lawyerscommittee.org/wp-content/uploads/2019/12/Online-Public-Accommodations-Report.pdf.

[35] "Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc." Oyez, https://www.oyez.org/cases/2014/13-1371. Accessed 3 May 2023.

[36] "Data Protection Impact Assessment (DPIA)." GDPR.Eu, 9 Aug. 2018, https://gdpr.eu/data-protection-impact-assessment-template/.

[37] Selbst, Andrew D., An Institutional View Of Algorithmic Impact Assessments (June 15, 2021). 35 Harvard Journal of Law & Technology 117 (2021), UCLA School of Law, Public Law Research Paper No. 21-25, Available at SSRN: https://ssrn.com/abstract=3867634

[38] Gilman, M. E. (2022, November 2). Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4266250

[39] Ibid. Gilman.

[40] ManageEngine . (n.d.). Data visibility and security solution. DataSecurityPlus. Retrieved December 13, 2022, from https://www.manageengine.com/data-security/what-is/data-minimization.html

[41] "Blueprint for an AI Bill of Rights | OSTP." The White House, https://www.whitehouse.gov/ostp/ai-bill-of-rights/. Accessed 15 June 2023.

[42] What data can we process and under which conditions? (n.d.). Commission.europa.eu. https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/overview-principles/what-data-can-we-process-and-under-which-conditions_en

[43] Data Protection and Privacy: 12 ways to protect user data. Cloudian. (2022, June 20). Retrieved December 13, 2022, from https://cloudian.com/guides/data-protection/data-protection-and-privacy-7-ways-to-protect-user-data/

[44] DLP is a set of strategies and tools one may utilize to hinder data from being lost, stolen, or accidentally deleted. Backups, snapshots, and replication solutions should be implemented to mitigate data loss or modification. Ibid.

[45] New digital realities; new oversight solutions. (2020, August 20). Shorenstein Center. https://shorensteincenter.org/new-digital-realities-tom-wheeler-phil-verveer-gene-kimmelman/

[46] Cox, J. (2022, May 3). Data broker is selling location data of people who visit abortion clinics. Vice. https://www.vice.com/en/article/m7vzjb/location-data-abortion-clinics-safegraph-planned-parenthood

[47] Gilman, M. E. (2022, November 2). Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4266250

[48] https://www.ccld.dss.ca.gov/carefacilitysearch/Search/ChildCare

[49] Search by Company. https://ai.fmcsa.dot.gov/hhg/Search.asp. Accessed 27 June 2023.

[50] "Consumer Complaint Database." Consumer Financial Protection Bureau, https://www.consumerfinance.gov/data-research/consumer-complaints/. Accessed 27 June 2023.

[51] Yi, X., Paulet, R., & Bertino, E. (2014). Homomorphic encryption. In Homomorphic encryption and applications (pp. 27-46). Springer, Cham.

[52] The private key is uniquely associated with the owner and is not made public. The private key is used to compute a digital signature that may be verified using the corresponding public key.

[53] A public key is a large numerical value that is used to encrypt data. The key can be generated by a software program, but more often, it is provided by a trusted, designated authority and made available to everyone through a publicly accessible repository or directory

[54] What are public and private keys?: Public key: Private key. AppViewX. (n.d.). Retrieved December 13, 2022, from https://www.appviewx.com/education-center/what-are-public-and-private-keys/

[55] Albrecht, M., Chase, M., Chen, H., Ding, J., Goldwasser, S., Gorbunov, S., ... & Vaikuntanathan, V. (2021). Homomorphic encryption standard. In Protecting Privacy through Homomorphic Encryption (pp. 31-62). Springer, Cham.

[56] Harold Byun, V. P. P. (2022, December 6). The advantages and disadvantages of homomorphic encryption. Baffle. Retrieved December 13, 2022, from https://baffle.io/blog/the-advantages-and-disadvantages-of-homomorphic-encryption/

[57] Feige, U., Fiat, A., & Shamir, A. (1988). Zero-knowledge proof of identity. Journal of cryptology, 1(2), 77-94.

[58] Enwood, D. (2021, October 5). Zero-knowledge proofs - a powerful addition to blockchain. Blockhead Technologies. Retrieved December 13, 2022, from https://blockheadtechnologies.com/zero-knowledge-proofs-a-powerful-addition-to-blockchain/

[59] Mahmood, Z., & Vacius, J. (2020, December). Privacy-Preserving Block-chain Framework Based on Ring Signatures (RSs) and Zero-Knowledge Proofs (ZKPs). In 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT) (pp. 1-6). IEEE.

[60] OSTP (2022). Climate and Energy Implications of Crypto-Assets in the United States. White House Office of Science and Technology Policy. Washington, D.C. September 8, 2022.

[61] Du, W., & Atallah, M. J. (2001, September). Secure multi-party computation problems and their applications: a review and open problems. In Proceedings of the 2001 workshop on New security paradigms (pp. 13-22).

[62] Du, W., & Atallah, M. J. (2001, September). Secure multi-party computation problems and their applications: a review and open problems. In Proceedings of the 2001 workshop on New security paradigms (pp. 13-22).

[63] Sabt, M., Achemlal, M., & Bouabdallah, A. (2015, August). Trusted execution environment: what it is, and what it is not. In 2015 IEEE Trustcom/BigDataSE/ISPA (Vol. 1, pp. 57-64). IEEE.

[64] Ibid.

[65] Koeberl, P., Phegade, V., Rajan, A., Schneider, T., Schulz, S., & Zhdanova, M. (2015, August). Time to rethink: Trust brokerage using trusted execution environments. In the International Conference on Trust and Trustworthy Computing (pp. 181-190). Springer, Cham.

[66] Editor. (2022, February 21). Federated learning: The shift from centralized to distributed on-device model training. AltexSoft. Retrieved December 13, 2022, from https://www.altexsoft.com/blog/federated-learning/

[67] Thursday, A. 06, & Learning Optimization, C. M. L. O.-device. (n.d.). Federated learning: Collaborative machine learning without centralized training data. – Google AI Blog. Retrieved December 13, 2022, from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

[68] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1-210.

[69] Ibid.

[70] "features" are the variables used in a model; when an attack like Model Inversion Attack or Membership Inference Attack successfully infers what features are used in the model, then the features become "inferred features".

[71] Anderson, J. W., Kennedy, K. E., Ngo, L. B., Luckow, A., & Apon, A. W. (2014, October). Synthetic data generation for the internet of things. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 171-176). IEEE.

[72] Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. Mathematics, 10(15), 2733.

[73] Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., ... & Vadhan, S. (2018). Differential privacy: A primer for a non-technical audience. Vand. J. Ent. & Tech. L., 21, 209.
[74] Tyagi, N. (n.d.). What is differential privacy and how does it work? Analytics Steps. Retrieved December 14, 2022, from https://www.analyticssteps.com/blogs/what-differential-privacy-and-how-does-it-work

[75] Dilmegani, C. (2022, October 15). Differential Privacy: How It Works, Benefits & Use Cases. AIMultiple. Retrieved December 14, 2022, from https://research.aimultiple.com/differential-privacy/

[76] Plausible Deniability means it is not possible to determine with a high degree of confidence that information relating to a specific individual is a present in the data

[77] Dilmegani, C. (2022, October 15). Differential Privacy: How It Works, Benefits & Use Cases. AIMultiple. Retrieved December 14, 2022, from https://research.aimultiple.com/differential-privacy/

[78] Ibid.

[79] California consumer privacy act(Ccpa). (2018, October 15). State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

[80] Federal Trade Commission. Privacy & Data Security. Update: 2018, https://www.ftc.gov/system/files/documents/reports/privacy-data-security-update-2018/2018-privacy-data-security-report-508.pdf.

[81] Housing discrimination under the fair housing act. (n.d.). HUD.Gov / U.S. Department of Housing and Urban Development (HUD). Retrieved December 15, 2022, from https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_act_overview

[82] Equal credit opportunity act. (2013, July 19). Federal Trade Commission. https://www.ftc.gov/legal-library/browse/statutes/equal-credit-opportunity-act

[83] Equal Credit Reporting act. (2013). CFPB Consumer Laws and Regulations. https://files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf

[84] Fair Credit Reporting Act. (2013, July 19). Federal Trade Commission. https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act

NFHA NATIONAL FAIR HOUSING ALLIANCE

TECHEQUITY COLLABORATIVE