DataGrail

Building Safer AI

Balancing Data Privacy with Innovation

Stephanie Kirmer, Sr MLE
DataGrail
www.stephaniekirmer.com
www.datagrail.io

This is not legal advice.

DataGrail

Key Concepts

Terminology

Data privacy

is the responsible, ethical, and regulation-compliant use of individuals' personal data.

People may consent to who gets to see/use their data and how

People have the ability to revoke consent

People have specific means to control their data (review, deletion, etc)

Data security

is the technical and physical protections implemented to secure a digital environment.

Data is protected from breach or exposure to unauthorized parties

Data is protected from corruption or alteration without authorization

https://www.datagrail.io/blog/data-privacy/data-privacy-vs-data-security-a-guide/

PII vs Personal Data

PII (Personally Identifiable Information)

- Could be used to identify a specific individual

Personal data (also known as Personal Information)

- Includes PII, but adds more
- Could be combined with other data to identify a specific individual
- Still risky and still sensitive to customers

Examples of Personal Information from CCPA:

- Unique identifiers (PII)
- Characteristics of protected classes
- Property or purchase/transaction records
- Biometric information
- Internet/app usage history (including search history)
- Geolocation data
- Multimedia recording (such as video/photos)
- Professional or employment-related information
- Education information not otherwise publicly available
- Any profile reflecting preferences, characteristics, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes

“

87% of the population in the USA had reported characteristics that likely made them unique based only on ZIP, gender and date of birth.

– Eugenia Politou, *Journal of Cybersecurity*

Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions | Journal of Cybersecurity | Oxford Academic

Data Localization

Where can personal data be stored?

The physical footprint of data centers and servers may need to be in a specific country (or NOT in a specific country) depending on the nationality/residence of the people whose data is involved.

Even if you have the right to access and use the data for machine learning, you need to be cautious about where the data is stored.

Example:

Are you keeping files in S3 using AWS-East-1? Is that data permitted to be stored in the United States?

You may need data warehousing solutions based in datacenters in different jurisdictions to be compliant.

https://medium.com/@s.kirmer

DataGrail

Legal Frameworks

Photo by Olga Dudareva on Unsplash

Common Expectations in Law

A baseline set of things you should consider having/doing

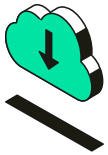
Informed consent to data usage

Public, transparent communication about data collection, storage, and usage

Method to opt out/refuse consent without discrimination

Method to revoke consent

Define covered data as more than strictly PII



Allow individuals to access, correct, and delete stored data about themselves

Specific Legal Obligations for ML/AI

These don't apply in all jurisdictions, but are increasingly popular in comprehensive privacy legislation.

Require minimization

Use the minimum amount of data necessary to do the job, and be able to prove that you were careful to minimize risk.

Prevent discrimination

Algorithms/ML models/AI can't be used to discriminate on protected characteristics (or do any other illegal things).

Conduct impact assessments

Complete required impact assessments on algorithms you want to deploy, identifying any risks and your steps to mitigate them.

Localities

In the US

Fragmented policies (no national law as yet) mean complicated legal obligations.

Products used by EU citizens are covered by many EU regulations, even if the provider is in US.

In the EU

Strict policies are in place with stringent penalties.

Regulations are relatively easier to follow because of consistency compared to US.

Elsewhere

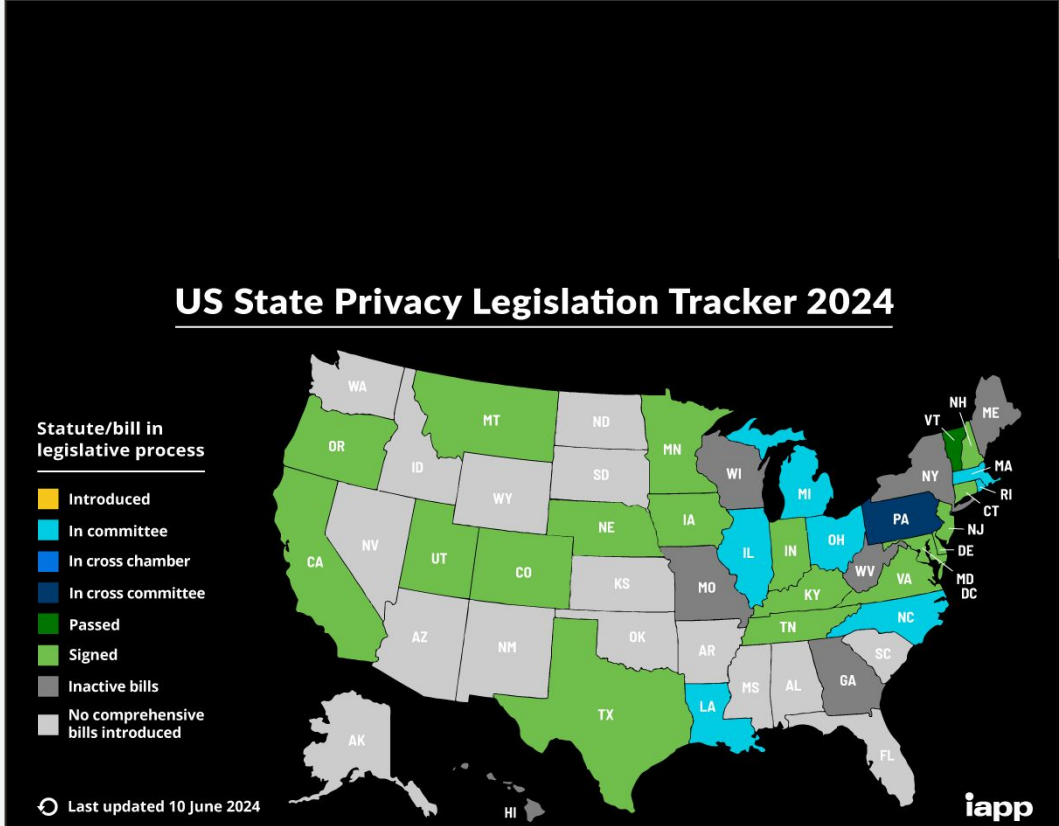
Strong nationwide data privacy laws are gaining popularity, many with requirements similar to GDPR (Japan, India, Vietnam, Indonesia, China, Sri Lanka, Brazil).

Many policies govern data localization as well.

US State Level

US States with comprehensive data privacy laws:

- California
- Oregon
- Montana
- Utah
- Colorado
- Texas
- Iowa
- Nebraska
- Indiana
- Tennessee
- Kentucky
- Virginia
- New Jersey
- Connecticut
- Delaware
- Maryland
- New Hampshire



DataGrail

How to Cope

Getting your work done while complying with regulations

Before Modeling

Tasks to do **before** you start writing the code

Be Informed About Responsibilities

- Know **whose data** is going in to this model, and who will be using it.
- Know the **jurisdictions** that apply to your work.
- Consult your **legal department** if you have any questions or uncertainties about how to be compliant.

Collect Data Carefully

- Get involved in the **consent** process - make sure you get the permissions you will need.
- Collect and use the **minimum amount of data** you'll need - not less, but not more either.
- Set up processes to **monitor consent** (so you know if it's revoked or changes).
- Make sure you meet **data localization** requirements.

When in doubt, don't use personal data in your model!

Revoking Consent and the Right to Be Forgotten

A customer consents to their data being used for ML, and you train a model using that data. ***What if a customer later revokes that consent?***

1. You **can** still use the model you built originally for inference
2. You **cannot** train any MORE models with that customer's data

Consider Alternatives to Individual Data

- Synthetic data
- Aggregated data
- Open source data
- De-identification

Of course, sometimes you just need to use individual data, but it should be your last resort.

Photo by Joseph Chan on Unsplash

What to do during/after modeling

Protect the data **during and after** you use it

Know where your data is going

If you transmit data outside your network to tune a third party base model, know what that provider's data privacy/IP policies are.

Is that data being protected?

Consider model applications

When you deploy, know where the results of the model inference will be used and who will see them.

Will outputs be applied for any bad purposes?

Redteam/Test your model

Test and build guardrails to prevent unwanted data exposure in inference.

Think from the perspective of bad actors and preempt attacks.

Closing Thoughts

- Think about it like it’s your personal data being used - how would you want it protected?
- Regulation can be a pain, but it’s there to advocate for the rights of people like you and me
- Everything is a risk-reward tradeoff - there’s no such thing as zero risk in data. Negotiate with your legal/security teams to find the balance that works
- Being aware of the limitations ahead of time makes it easier - you can plan ahead instead of having your work disrupted midway through

References

Understanding the Space

- https://www.datagrail.io/blog/data-privacy/data-privacy-vs-data-security-a-guide/
- https://iapp.org

Legal Frameworks

- https://www.oreilly.com/radar/how-will-the-gdpr-impact-machine-learning/
- https://academic.oup.com/cybersecurity/article/4/1/tyy001/4954056
- https://oag.ca.gov/privacy/ccpa
- https://energycommerce.house.gov/posts/committee-chairs-rodgers-cantwell-unveil-historic-draft-comprehensive-data-privacy-legislation
- https://www.mineos.ai/articles/the-state-of-data-protection-in-asia
  https://www.mineos.ai/articles/a-guide-to-indias-data-protection-law-what-to-expect-from-dpdp

Ethical AI Policy Help

- https://www.datagrail.io/resources/reports/responsible-ai-use-principles-policy-guide/
- https://www.datagrail.io/ai-and-data-privacy/

Thank you!

www.stephaniekirmer.com
www.datagrail.io

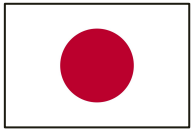
Appendix

Informed Consent

Photo by SEO Galaxy on Unsplash

- Subject is competent to consent (not a child, for example)
- Subject has been given all necessary information (about what will happen, risk potential)
- Subject understands the meaning of the information provided
- Subject consented voluntarily, without coercion or undue influence

Highlights in APAC



Japan – APPI

Protection of Personal Information (APPI) law passed in 2003, amended in 2015 and 2022.

China – PIPL

Personal Information Protection Law passed 2021. Different from GDPR. Includes data localization.

India – DPDP

Passed 2023. Covers nonresidents in India. Less regulation of ML/AI based decision making than GDPR.

Vietnam

Decree on Protection of Personal Data passed 2023. Includes data localization.

All in effect as of writing.

Key Legislation in EU/US

GDPR (In effect)

General Data Protection Regulation protects data privacy rights for EU citizens and residents

CCPA (In effect)

California Consumer Privacy Act protects data privacy rights for California citizens and residents

EU AI Act (passed)

Regulates AI development and usage affecting EU citizens and residents

APRA (bill in committee)

American Privacy Rights Act would protect data privacy rights for citizens & residents of the United States and regulate decision-making using AI/ML