Privacy & Bias in Edge Computing

Chapter 8

Outline

- ❑ Privacy concerns & forms
  - New edge privacy risks
  - Data, location, identity privacy
- ❑ Privacy techniques
  - Anonymization & k-anonymity
  - Differential privacy, HE, FL
- ❑ Bias types & causes
  - 13 forms of bias
  - Where bias comes from
- ❑ Bias mitigation
  - Pre / in / post-processing
  - Open research problems
- ❑ Summary & Practice

Privacy in Edge Computing

Privacy concerns at the edge

- Outsourcing data to third-party edge services splits data ownership from control — raising seven distinct concerns across the cloud/edge/device layers.

Distributed storage

Data split across nodes complicates integrity & invites leakage.

Retained identifiers

Untrusted operators may keep user-identifiable data after service ends.

Honest-but-curious

Authorized entities exploit their access to harvest sensitive data.

Predictable locations

Bringing compute closer makes user location easier to infer.

Decentralized admin

Dispersed devices resist central control — easier to attack.

Transmission risk

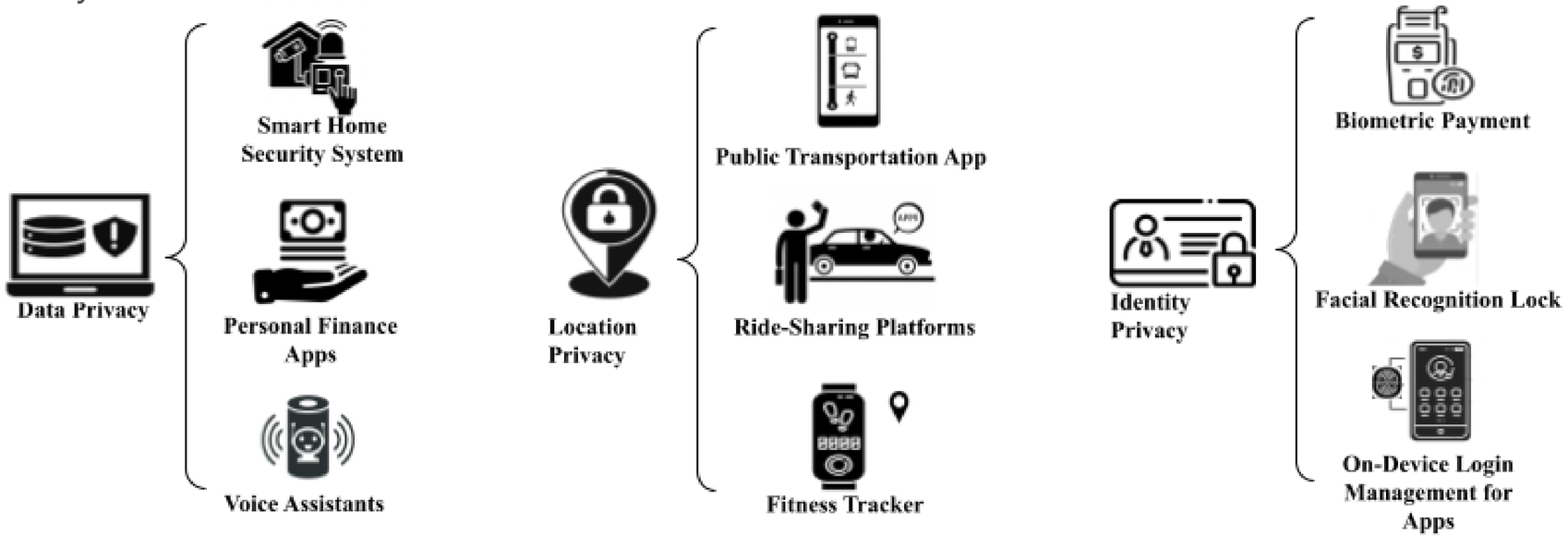
Raw texts & images can leak while moving between layers.

Uncontrollable sharing

Once data leaves, the owner loses say over how it's shared.

Privacy in Edge Computing

Three forms of privacy



Privacy forms: data, location and identity privacy.

Data privacy

Protect sensitive data stored, processed or moved across edge servers & devices — smart homes, finance apps, voice assistants.

Location privacy

Curious servers infer location from connection points; movement patterns leak — transit, ride-share, fitness apps.

Identity privacy

Protect identity in verification (knowledge, possession, biometric, PUF); a cloned identity exposes all linked data.

Privacy-Preserving Techniques

Data anonymization

➢ Remove or obscure any information that can directly or indirectly identify someone — before data is published. Three core operations:

Data masking

Change characters in an attribute via shuffling, substitution or encryption.

e.g. John.Doe123@example.com → Zzzz.Zzz111@zxxxxzz.zzz

Generalization

Replace specific values with consistent ranges to reduce identifiability.

e.g. John Smith, 29, $50,000 → Person 1, 20–30, $50–60k

Suppression

Remove an entire column or tuple when it can’t otherwise be anonymized.

e.g. Sensitive value → ****

Privacy-Preserving Techniques

K-ANONYMITY: Making each record one-of-k

- k-anonymity ensures each record is indistinguishable from at least k–1 others, so no one is identifiable with probability > 1/k. It combines suppression + generalization; the challenge is balancing utility vs. privacy.

- **Original dataset**

Name	Age	Gender	Zip	Disease
Dana	24	F	45011	Diabetes
John	30	M	45012	None
Jack	28	M	45013	Asthma
Alex	26	M	45014	Hypertension
Chloe	27	F	45015	None
Fiona	29	F	45016	Diabetes

2-anonymity

- **Anonymized (2-anonymity)**

Name	Age	Gender	Zip	Disease
Suppressed	20–25	F	450**	Diabetes
Suppressed	26–31	M	450**	None
Suppressed	26–31	M	450**	Asthma
Suppressed	26–31	M	450**	Hypertension
Suppressed	26–31	F	450**	None
Suppressed	26–31	F	450**	Diabetes
Suppressed	20–24	F	450**	Diabetes

Privacy-Preserving Techniques

Differential privacy: Privacy by adding noise

ε-differential privacy. A provably secure method: the output barely changes whether or not any single record is present, so it resists membership-inference attacks — without relying on assumptions about the attacker's knowledge.

Privacy budget ε

Controls how much noise is added — smaller ε means stronger privacy, less utility.

Global sensitivity

The largest change in a query's output between two neighboring datasets.

Noise addition

Random noise proportional to sensitivity (not dataset size) hides individuals.

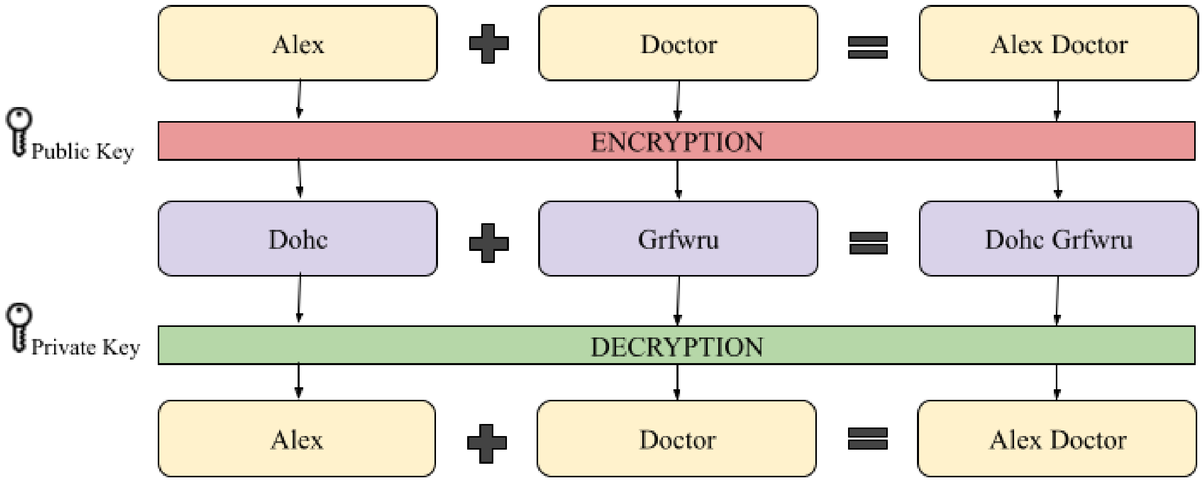
Laplacian mechanism

Adds Laplace-distributed noise; a Gaussian variant gives (ε,δ)-DP.

Edge use: end devices perturb location data (Voronoi cells + local DP) before sending to the nearest edge node — obscuring exact positions while keeping statistical utility.

Privacy-Preserving Techniques

Homomorphic Encryption: Computing on encrypted data



Operations on ciphertext (Alex+Doctor) decrypt to the correct plaintext result.

The idea.

Additions & multiplications on ciphertext reflect directly on the underlying plaintext — so a third party can process data it can never read. $E(m_1)\cdot E(m_2) = E(m_1\cdot m_2)$.

Edge use case

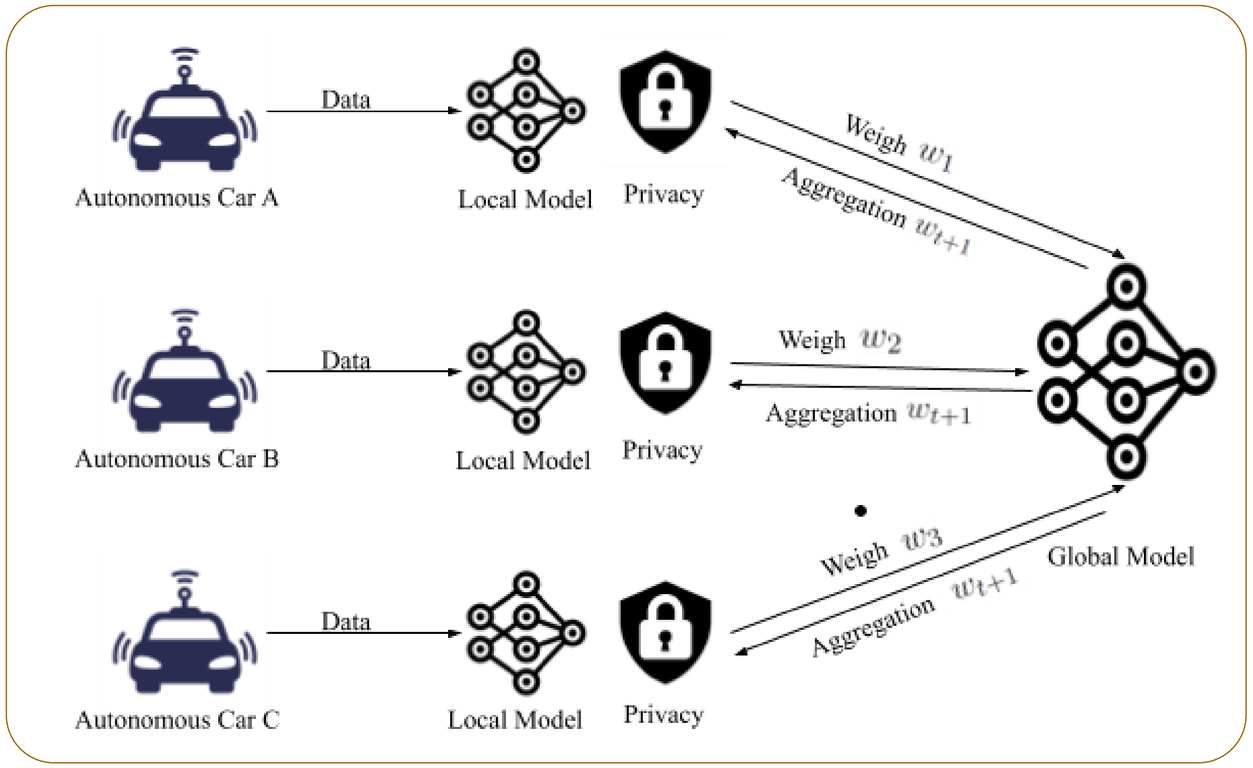
Devices encrypt sensitive data locally → the edge server runs heavy computation on the ciphertext → results return to the device, which decrypts them. The server's power is used without ever exposing plaintext.

The caveat

Homomorphic encryption is computationally heavy and incurs significant latency — often impractical for real-time edge applications.

Privacy-Preserving Techniques

Learn together, keep data local



Federated learning: local models train on-device; only weights are aggregated into a global model.

How it works.

A global model lives on an aggregator (e.g. an MEC server). Each node trains locally and sends weight updates the server aggregates them and returns the new global model — repeating until convergence.

Privacy preserved Only parameters are exchanged — raw source data never leaves the device.

AV example Autonomous vehicles fuse GPS, LiDAR & camera data into a shared global model.

Benefits & cost Cuts communication cost, boosts scalability — but needs capable on-device compute.

Swapnil Sadashiv Shinde et al. “On the design of federated learning in latency and energy constrained computation offloading operations in vehicular edge computing systems”. In: IEEE Transactions on Vehicular Technology 71. 2 (2021), pp. 2041–2057.

Privacy-Preserving Techniques

Privacy: Open research problems

FL efficiency

Balance computational cost against privacy for resource-limited edge nodes.

DP for data streams

Apply differential privacy to continuous, real-time edge data flows.

Lightweight HE

Develop homomorphic encryption light enough for constrained devices.

AI-driven anonymization

Use transformers & LLMs to anonymize data without losing utility.

Security–privacy–trust

Balance the three across multiple untrusted or semi-trusted parties.

Legal & ethical

Build legal frameworks that ensure compliance and protect user rights.

Bias in Edge Computing

Bias & the machine learning lifecycle

A deviation from a normative standard

Bias covers behaviors like gender or racial bias that can harm groups defined by sensitive attributes even with no malicious intent.

At the edge, heterogeneous sensing and conditions break the assumption that local data is independent & identically distributed (IID) — producing biased global models, unfair decisions, and unreliable performance.

➢ Bias can enter at every stage

- **Data collection** Skewed or low-quality sources.
- **Data preparation** Labeling, cleaning & sampling choices.
- **Modeling** Objective favors the majority group.
- **Evaluation** Tested on unrepresentative data.
- **Deployment** Used in a different context than trained.

Bias in Edge Computing

Types of bias

Data bias

Variability & prejudice from uneven source quality.

Algorithmic bias

Algorithm adds or amplifies bias not in the data.

Statistical bias

Output deviates from the true estimate.

Sampling bias

Nonrandom sampling of subgroups misgeneralizes.

Social bias

External behavior skews human judgment.

Historical bias

Past societal bias baked into training data.

Representation bias

Training data under-represents some segments.

Bias in Edge Computing

Types of bias

Group bias

FL performance differs across sensitive-attribute groups (urban vs. rural).

Attribute bias

Attribute distributions diverge after network propagation.

Structural bias

Network structure amplifies group differences (job ads by race).

Measurement bias

Feature selection/measurement skews (self-reported exercise).

Evaluation bias

Model judged on unrepresentative evaluation data.

Population bias

User population differs from the intended target population.

Bias in Edge Computing

Causes of biases

Biased datasets

Biased device measurements, historical human decisions, erroneous reports.

Error-minimizing algorithms

Minimizing aggregate error favors majority over minority groups.

Homophily of attributes

Race, gender, nationality directly influence predictions.

Data quality gaps

Missing values, small samples, misclassification & distribution bias.

On-device constraints

Limited memory, compute & energy propagate bias through design choices.

Edge bias (FL)

Each edge overfits its data; closed-world assumption skews the global model.

Bias in Edge Computing

Impact of biases: How bias hits edge algorithms

- Edge-cloud computing and FL scale up tasks like object detection — but they are not immune to bias, which degrades their performance in specific ways.

Data class disparity

Imbalanced labels & biased ground truth cause inaccurate object detection across cooperating devices.

Heterogeneous generation

Vehicles, sensors & apps generate data differently; environment adds inconsistency.

Accumulated bias

Distribution shifts during dynamic collection/scheduling — critical in disaster response.

FL global bias

Unreliable wireless biases the model toward well-connected devices, drifting from optimal.

Bias Mitigation

Pre-, In-, and Post-processing

Preprocessing

Fix the data before training

- Reweighing — weight groups/outcomes
- Suppression — drop sensitive attributes
- Massaging — relabel to reduce bias
- Multiple imputations — fill missing values

In-processing

Enforce fairness during training

- Fairness constraints in the objective
- Regularization / fairness regularizer
- Bias-aware algorithm choice
- Adversarial debiasing; fair representation

Postprocessing

Adjust the outputs after training

- Modify predictions for fairness
- Balance group & individual fairness
- Mind the accuracy–fairness trade-off
- Apply fairness performance metrics

Bias Mitigation

Preprocessing

Reweighing

Assign higher or lower importance to group–outcome combinations to balance representation.

Suppression

Remove sensitive attributes or selected fields so the model cannot directly use them.

Massaging

Modify selected training labels to reduce discrimination before fitting the model.

Multiple Imputation

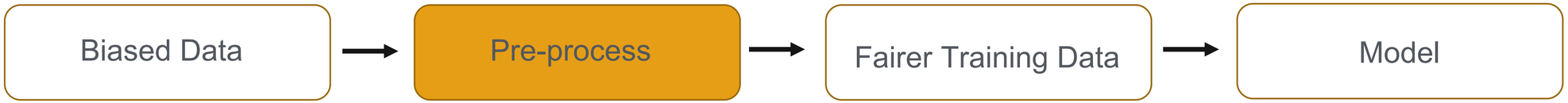
Estimate missing values to reduce skew caused by incomplete data.

Representative reweighing formulation

$$w(a, y) = \frac{P(A = a)\,P(Y = y)}{P(A = a, Y = y)}$$

A = sensitive attribute, Y = class label. Underrepresented (A,Y) combinations receive larger weights.

Key idea: Improve fairness before training by changing the data distribution—not the learning algorithm.



Bias Mitigation

In-processing

Fairness Constraints

Add explicit fairness requirements to the optimization objective.

Regularization

Penalize discriminatory behavior while optimizing predictive performance.

Algorithm Choice

Prefer learning algorithms that are less likely to amplify existing bias.

Adversarial Debiasing

Train an adversary to infer the sensitive attribute, then make inference harder.

Fair Representation Learning

Learn internal representations that preserve task-relevant information while reducing dependence on sensitive attributes.

Fairness-aware objective

$$\min_{\theta} \mathcal{L}_{task}(\theta) + \lambda \mathcal{L}_{fair}(\theta)$$

λ controls the trade-off between predictive accuracy and fairness.

Example fairness penalty: demographic parity gap

$$\mathcal{L}_{fair} = \left| P(\hat{Y} = 1 \mid A = 0) - P(\hat{Y} = 1 \mid A = 1) \right|$$

Smaller gap ⇒ more similar positive prediction rates across protected groups.

Bias Mitigation

Post-processing

Output Adjustment

Change decisions after the model produces scores or predictions.

Group Thresholds

Apply group-specific thresholds or decision rules to reduce disparities.

Fairness Targets

Optimize for group fairness, individual fairness, or a combination of both.

Accuracy Trade-off

Greater fairness may reduce predictive accuracy.

Group-specific decision threshold

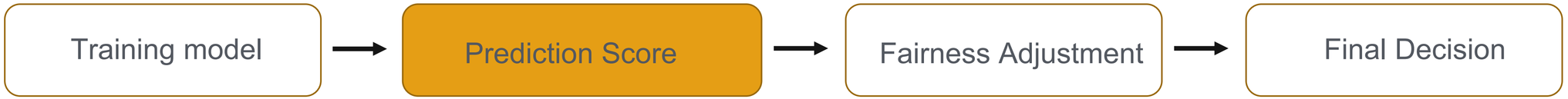
$$\hat{Y}' = 1 \text{ if } P(Y = 1 \mid X) \geq t_A$$

t_A is the decision threshold selected for protected group A.

Example target: equal opportunity

$$|TPR_{A=0} - TPR_{A=1}| \to 0$$

Reduce the true-positive-rate gap between protected groups.



Bias: Open Research Problems

Multimodal data

Mitigating bias across combined text, image & sensor modalities.

Novel metrics

New evaluation metrics that actually capture bias.

Generative models

Detecting & reducing bias in generative AI.

Interpretability

Making models' bias understandable & explainable.

Dynamic environments

Achieving fairness as data & conditions shift over time.

Summary

- Privacy at the edge remains important despite local processing because distributed systems introduce new risks.
- Data, location, and identity privacy are the key forms of privacy that require protection.
- Privacy-preserving techniques include anonymization, k-anonymity, differential privacy, homomorphic encryption, and federated learning.
- Bias can occur throughout the ML lifecycle, affecting data, algorithms, evaluation, and populations.
- Edge algorithms are vulnerable to bias due to class imbalance, heterogeneous data, and unreliable connectivity in federated learning.
- Bias mitigation can be applied through preprocessing, in-processing, and postprocessing techniques.
- Open challenges include lightweight privacy methods, streaming differential privacy, multimodal bias, and fairness in dynamic environments.

Practice

1. Identify the key factors that contribute to biases in AI models deployed on edge devices and discuss how these biases might manifest in real-world applications.

2. Discuss the potential privacy concerns associated with collecting and processing data at the edge

3. Analyze how biases in edge computing algorithms can impact different user groups, and outline steps to mitigate these biases.

4. Examine the legal measures necessary to ensure privacy protection in edge computing.

5. Discuss the technical challenges of implementing privacy-preserving techniques such as differential privacy and federated learning on edge devices, and how do these challenges affect privacy and bias.

Project

Implement and evaluate a bias mitigation algorithm.

Goal. Take a dataset used by an edge application, measure the bias in a trained model, then implement a mitigation technique and evaluate how fairness and accuracy change.

➢ What you'll do

1. **Assess the dataset** Identify sensitive attributes & measure class/label imbalance.
2. **Train a baseline** Build a model and record accuracy plus fairness metrics per group.
3. **Apply mitigation** Add a pre-, in-, or post-processing method (e.g. reweighing, adversarial debiasing).
4. **Compare** Re-measure fairness vs. accuracy; quantify the trade-off.

➢ What to measure

- **Group fairness** Performance gap across sensitive-attribute groups.
- **Individual fairness** Similar individuals get similar outcomes.
- **Accuracy** Overall predictive performance after mitigation.
- **Trade-off** How much accuracy was traded for fairness.

Example: on a health dataset skewed by age, apply reweighing + adversarial debiasing; report the fairness gain and the accuracy cost. (Other options: privacy-preserving app, blockchain for privacy/bias, AI-audit framework.)

References

- Swapnil Sadashiv Shinde et al. “On the design of federated learning in latency and energy constrained computation offloading operations in vehicular edge computing systems”. In: IEEE Transactions on Vehicular Technology 71. 2 (2021), pp. 2041–2057.
- Abdulmalik Alwarafy et al. “A survey on security and privacy issues in edge-computingassisted Internet of Things”. In: IEEE Internet of Things Journal 8. 6 (2020), pp. 4004–4022.
- Priyank Jain, Manasi Gyanchandani, and Nilay Khare. “Big data privacy: A technological perspective and review”. In: Journal of Big Data 3 (2016), pp. 1–25.
- Frank Stinar, Zihan Xiong, and Nigel Bosch. “An approach to improve k-anonymization practices in educational data mining”. In: Journal of Educational Data Mining 16. 1 (2024), pp. 61–83.
- Le Wu et al. “Learning fair representations for recommendation: A graph-based perspective”. In: Proceedings of the Web Conference 2021. 2021, pp. 2198–2208.