Privacy Accounting and Quality Control in the Sage Differentially Private ML Platform

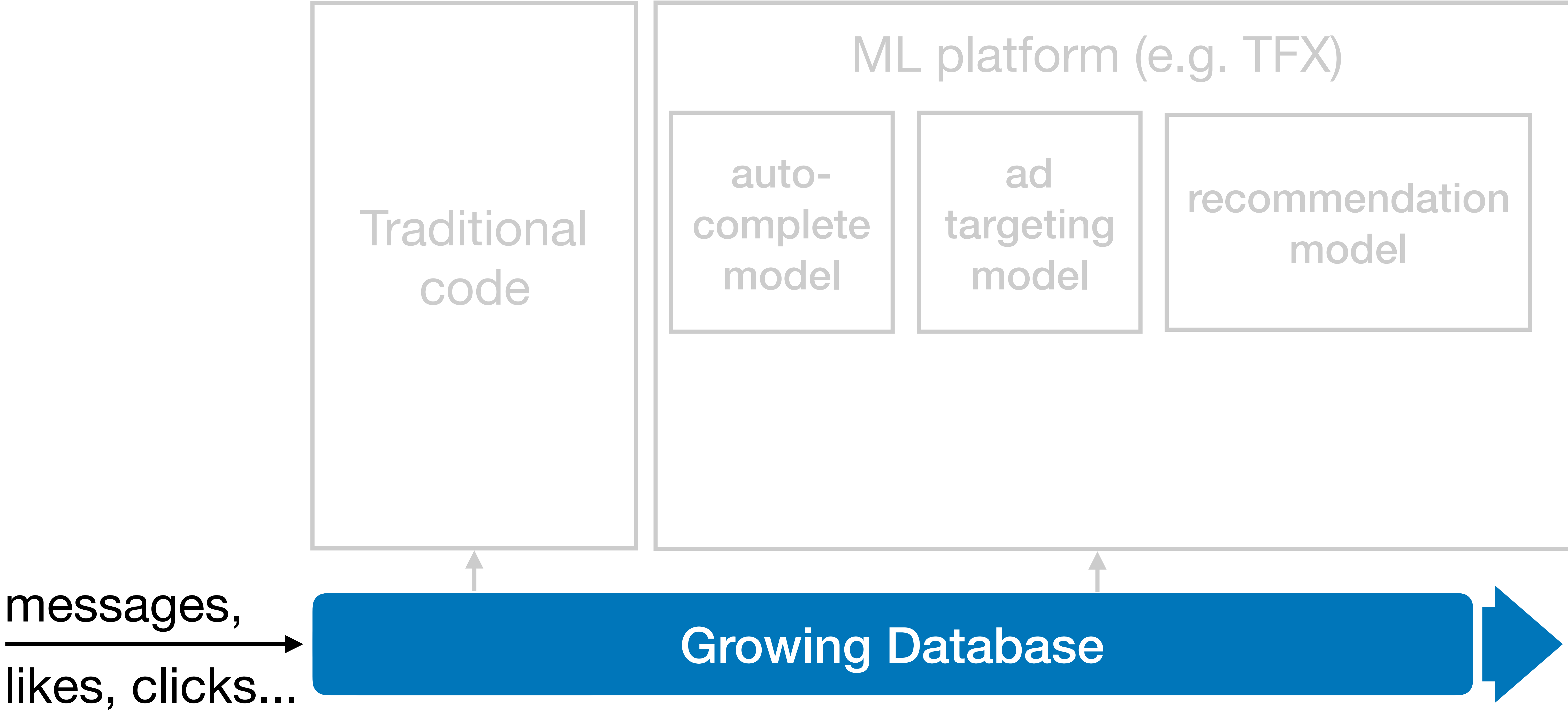
Mathias Lécuyer

With:
Riley Spahn, Kiran Vodrahalli, Roxana Geambasu, and Daniel Hsu

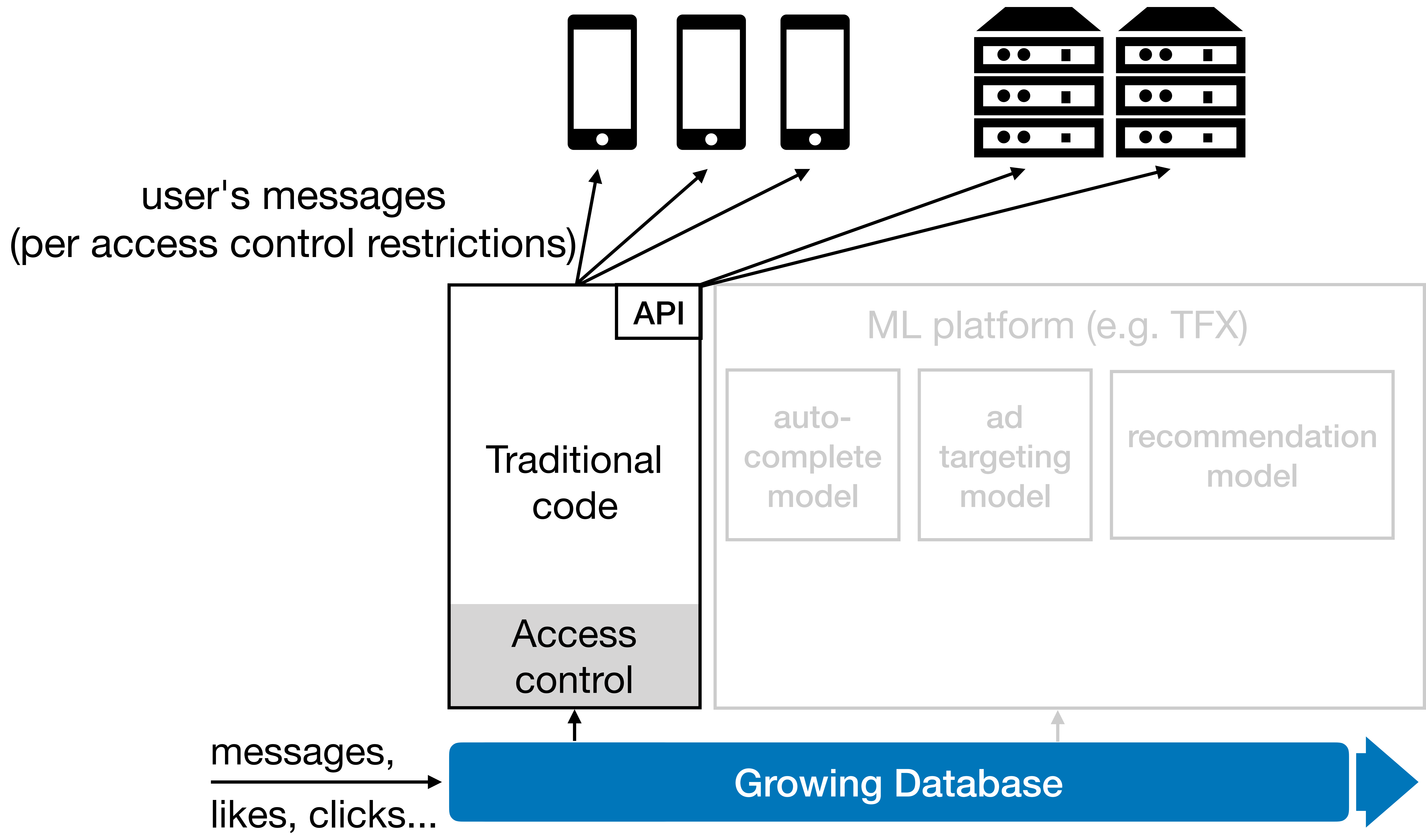
Machine Learning (ML) introduces a dangerous double standard for data protection

Example: messaging app

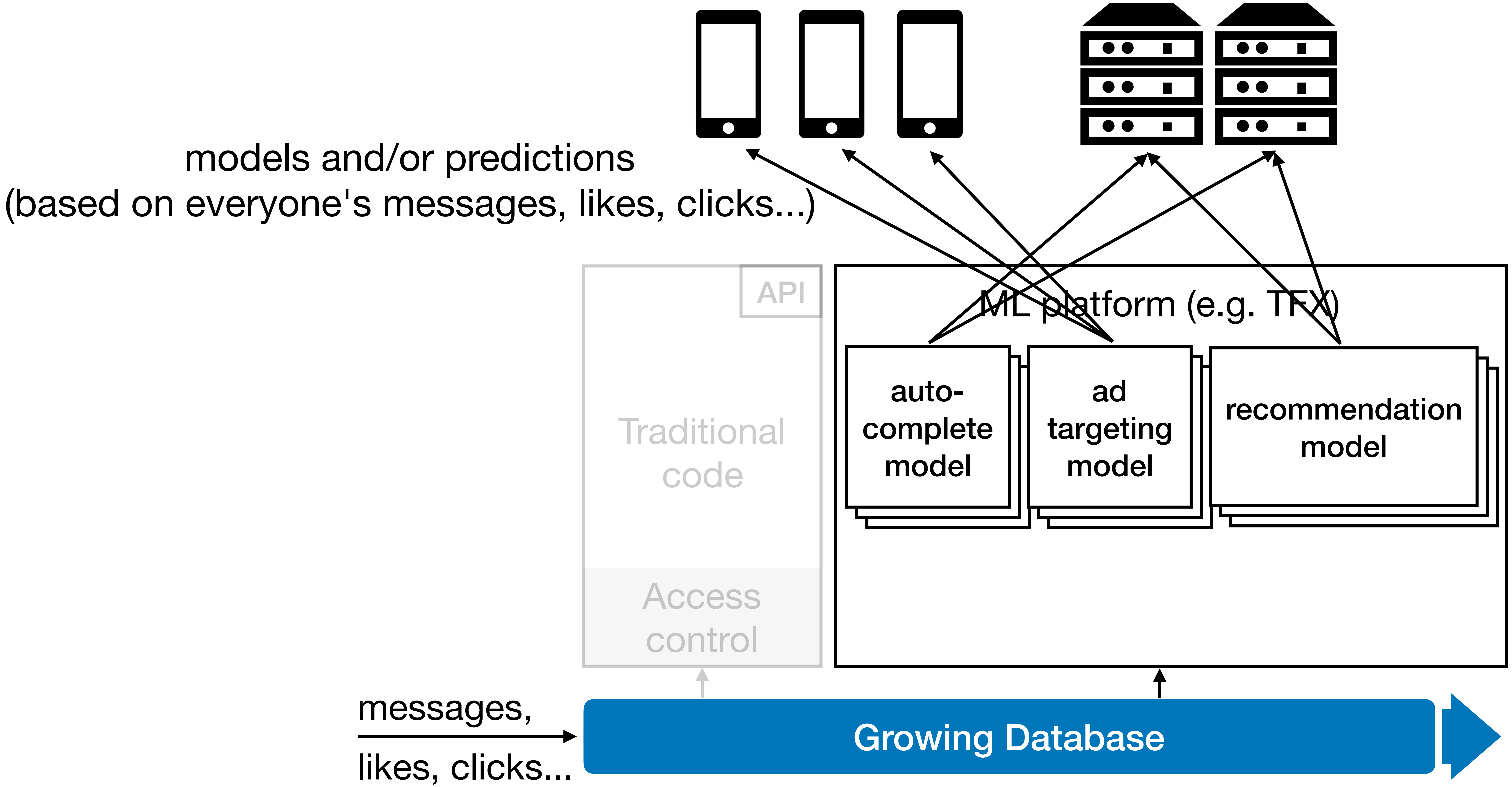
Example: messaging app



Example: messaging app

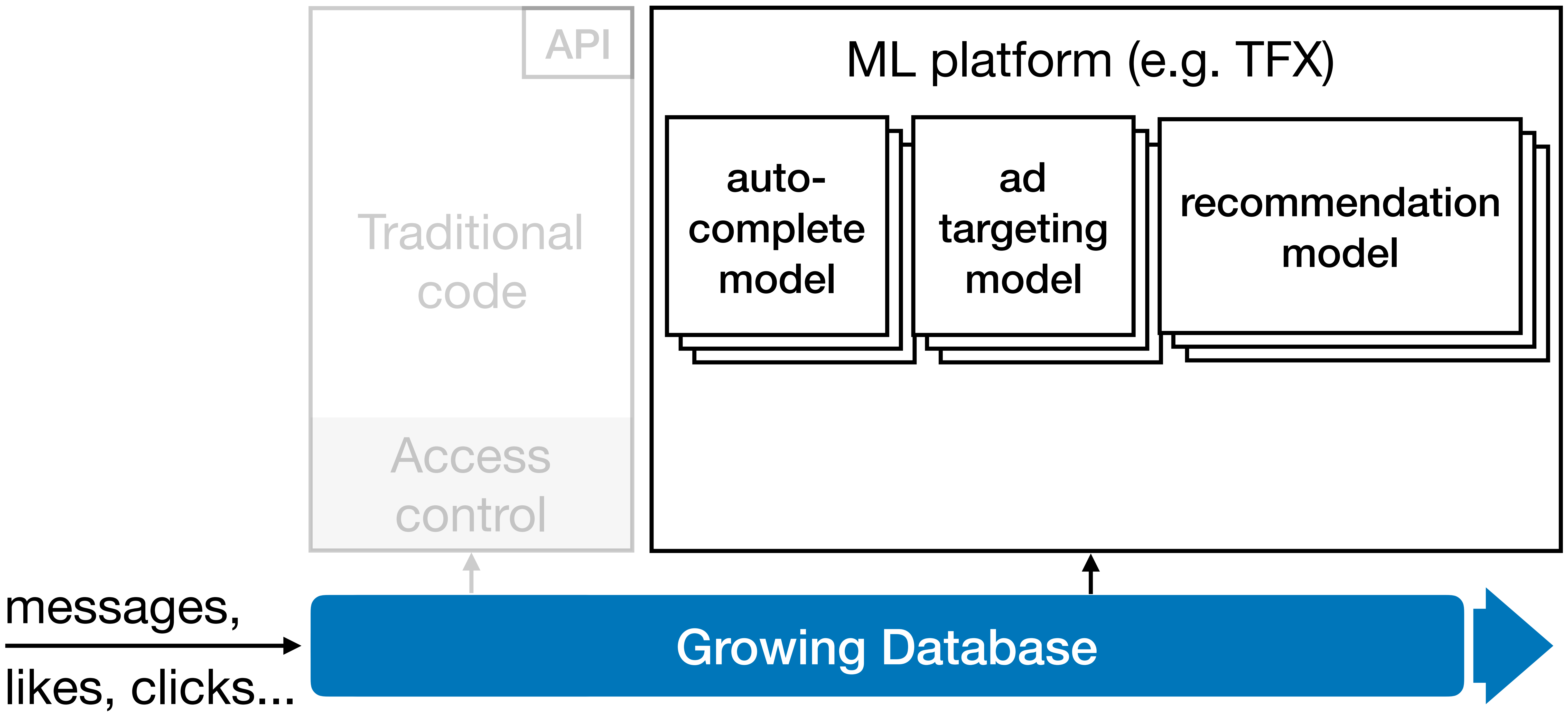


Example: messaging app

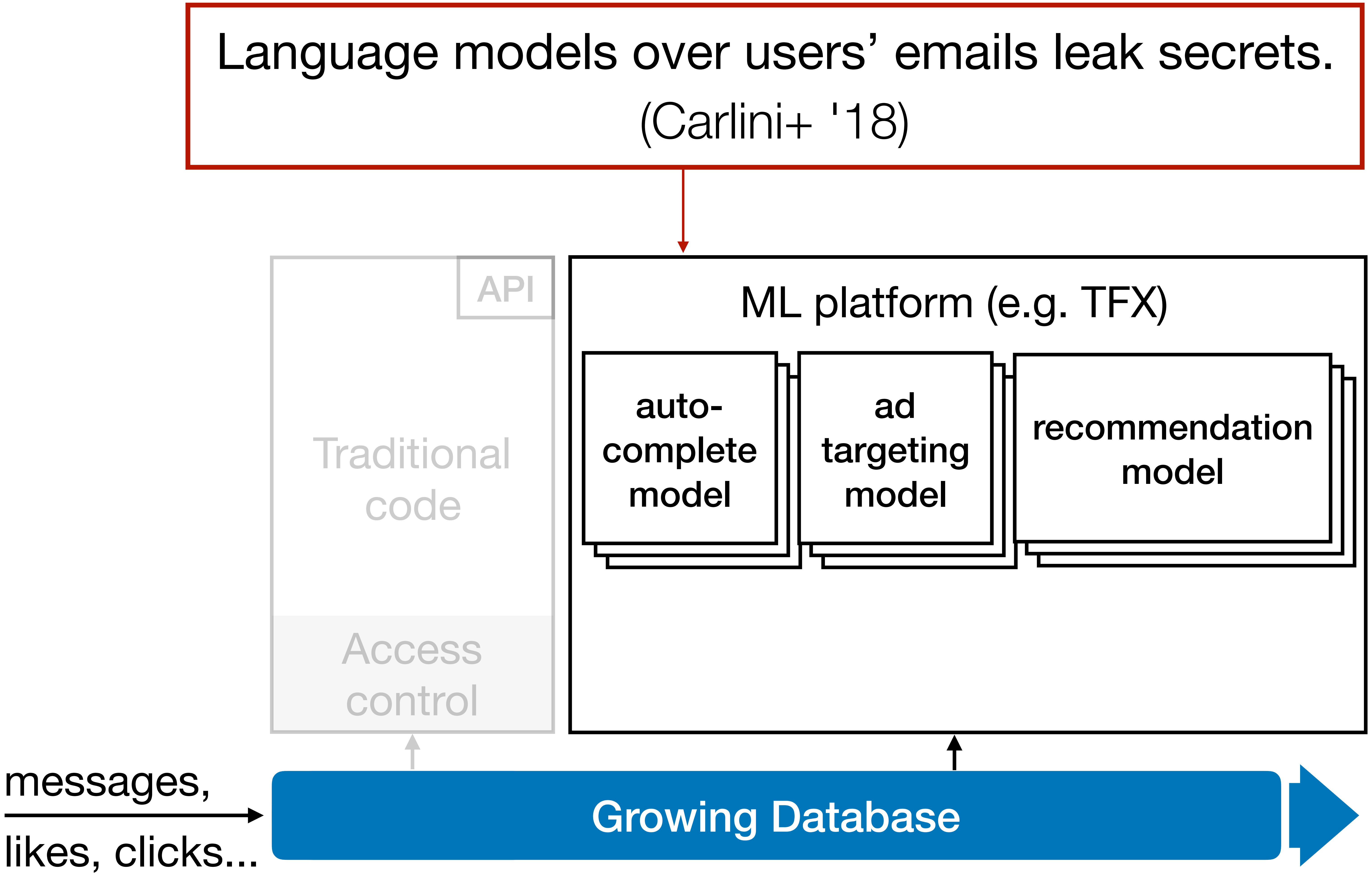


Example: messaging app

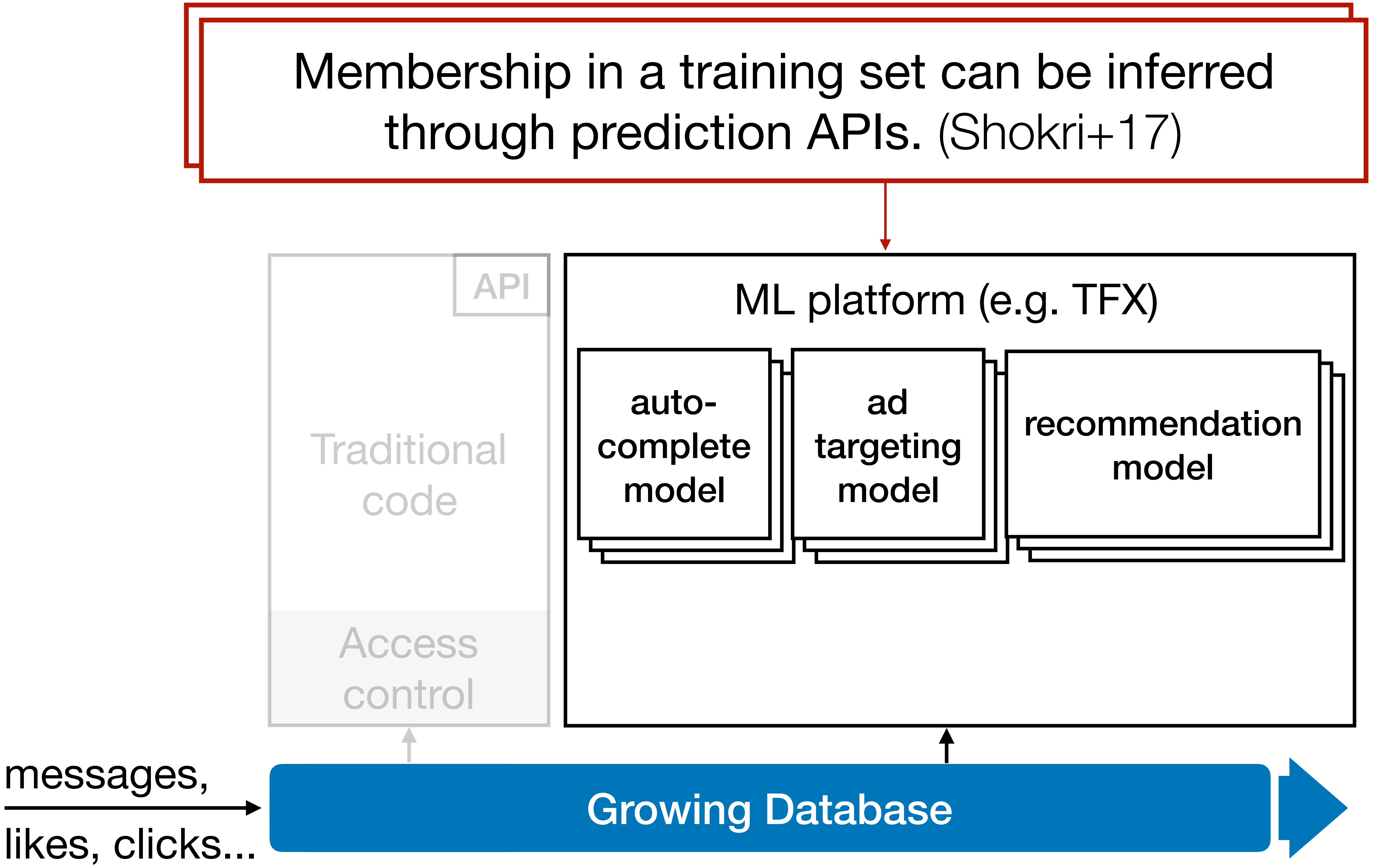
ML should only captures general trends from the data, but often captures specific information about individual entries in the dataset.



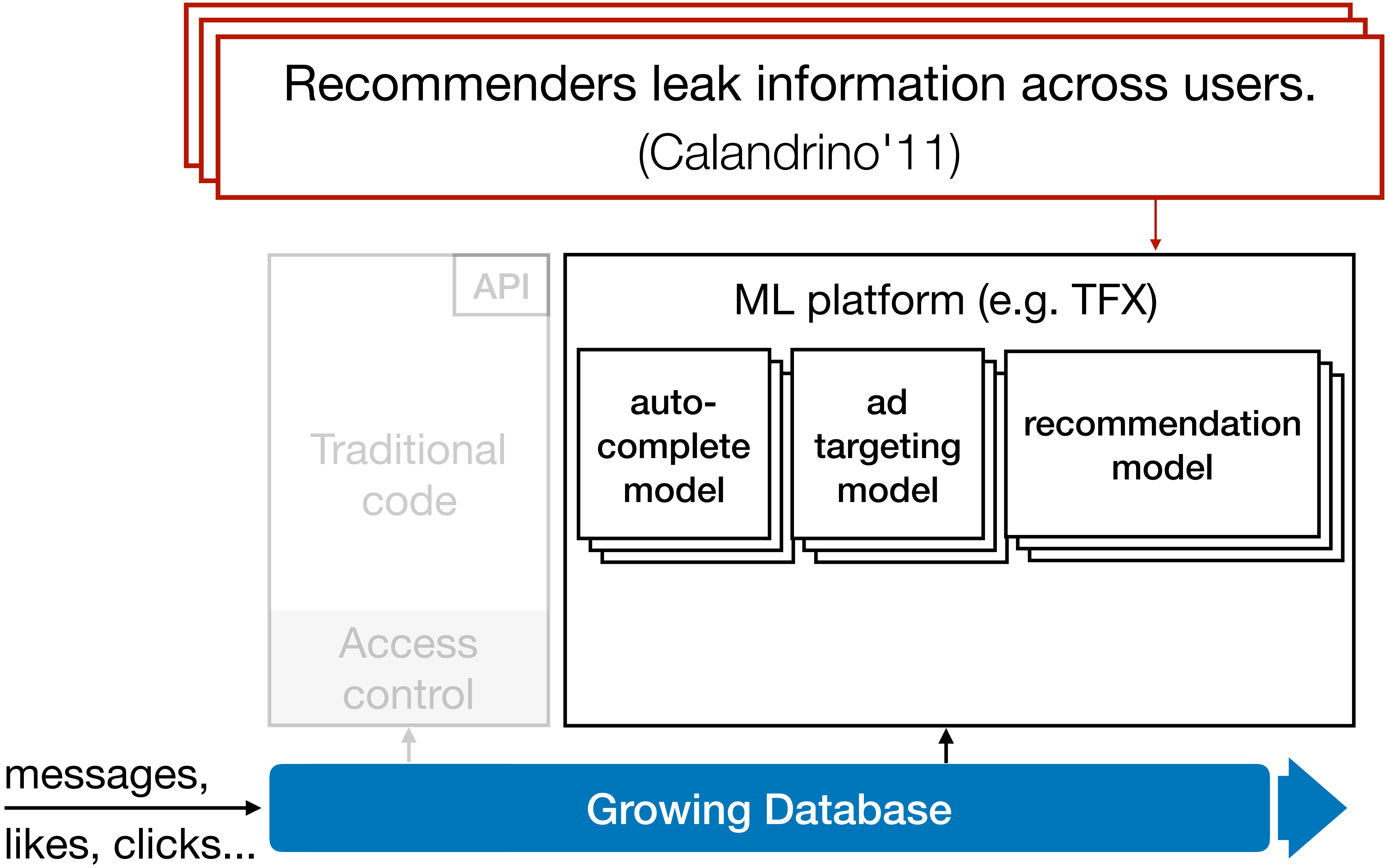
Example: messaging app



Example: messaging app

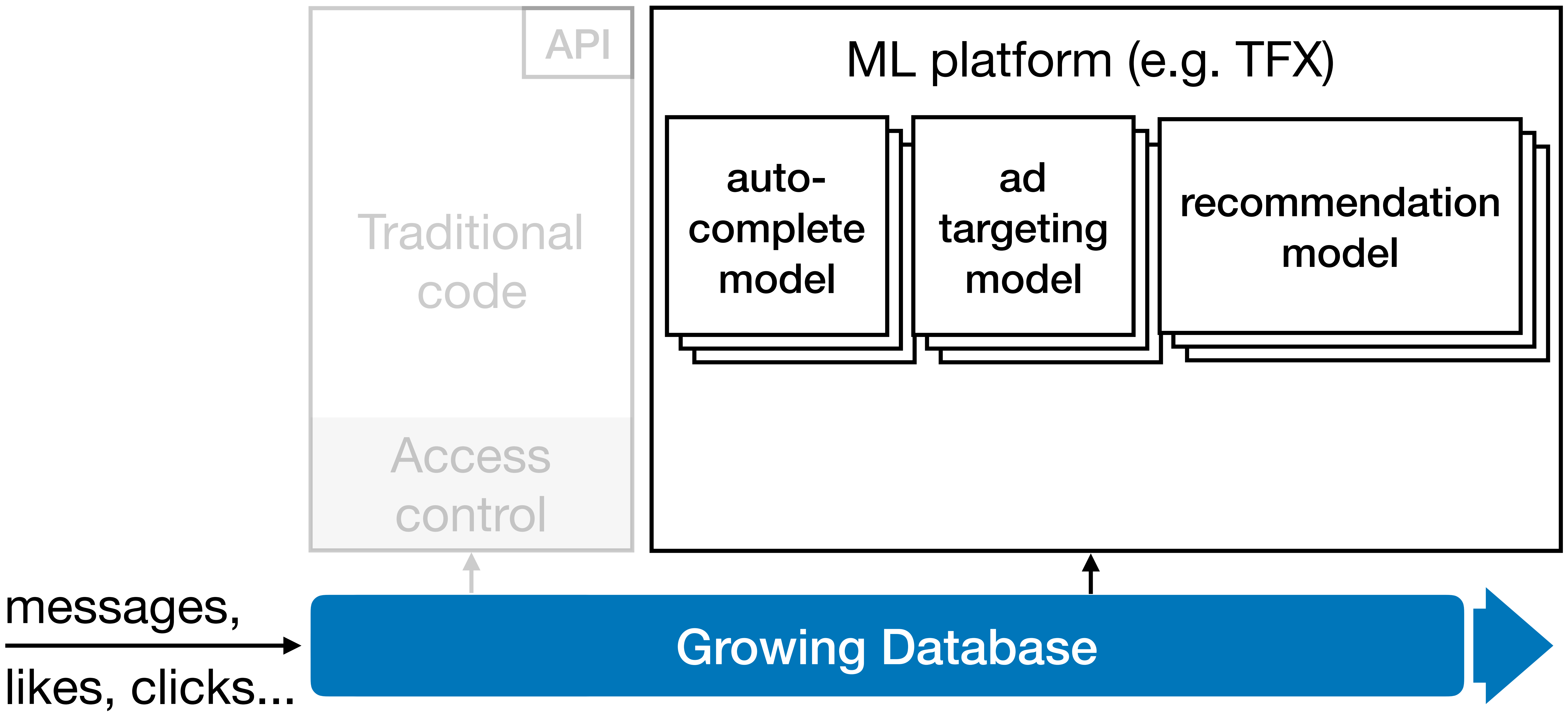


Example: messaging app



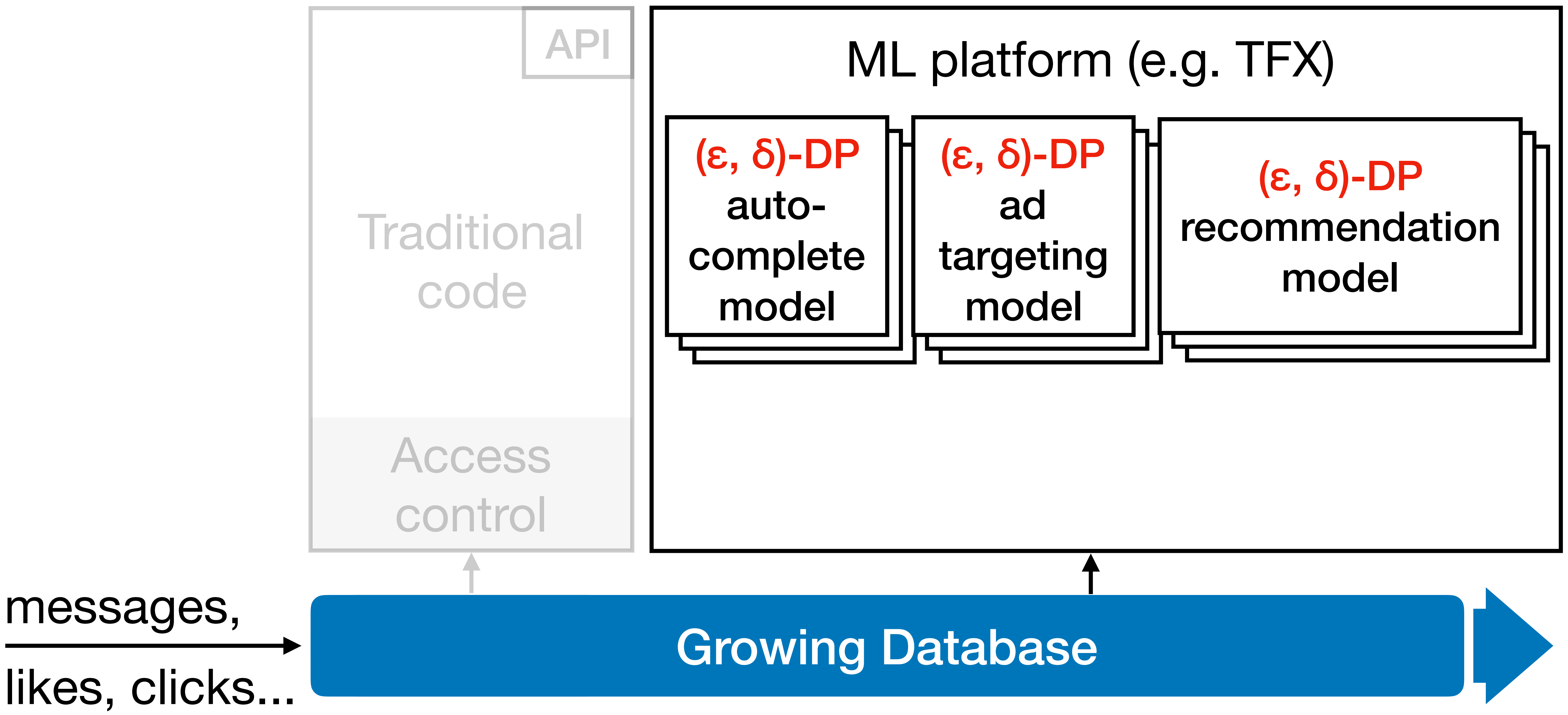
Example: messaging app

- Making individual training algorithms Differentially Privacy (DP) is good but insufficient, because old data is reused many times.
- No system exists for managing multiple DP training algorithms to enforce a global DP guarantee.



Example: messaging app

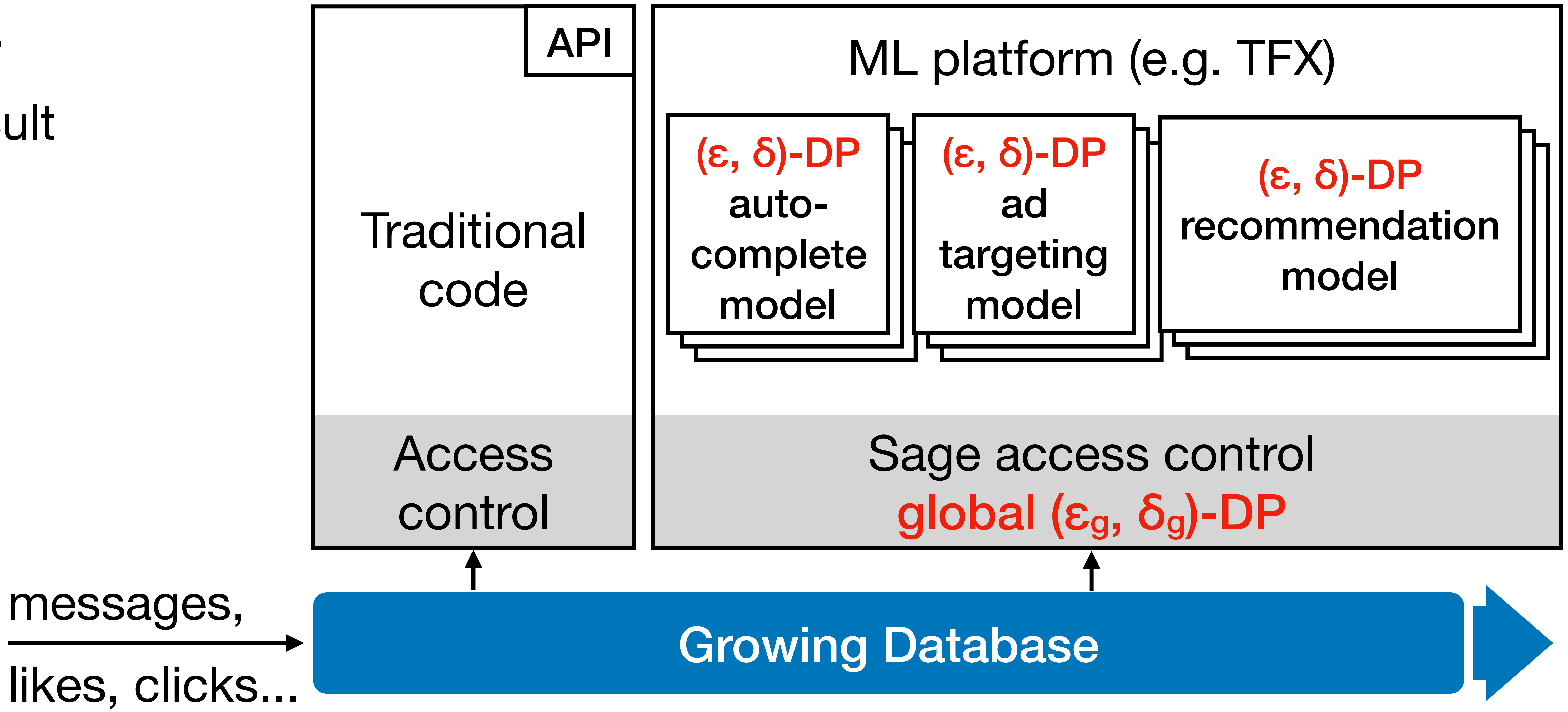
- Making individual training algorithms Differentially Privacy (DP) is good but insufficient, because old data is reused many times.
- No system exists for managing multiple DP training algorithms to enforce a global DP guarantee.



Can we make Differential Privacy practical for ML applications?

Sage

- Enforces a global (ε_g,δ_g)-DP guarantee across all models ever released from a growing database.
- Tackles in practical ways two difficult DP challenges:
  1. "Running out of budget"
  2. "Privacy-utility tradeoff."



Outline

Motivation

Differential Privacy

Two practical challenges

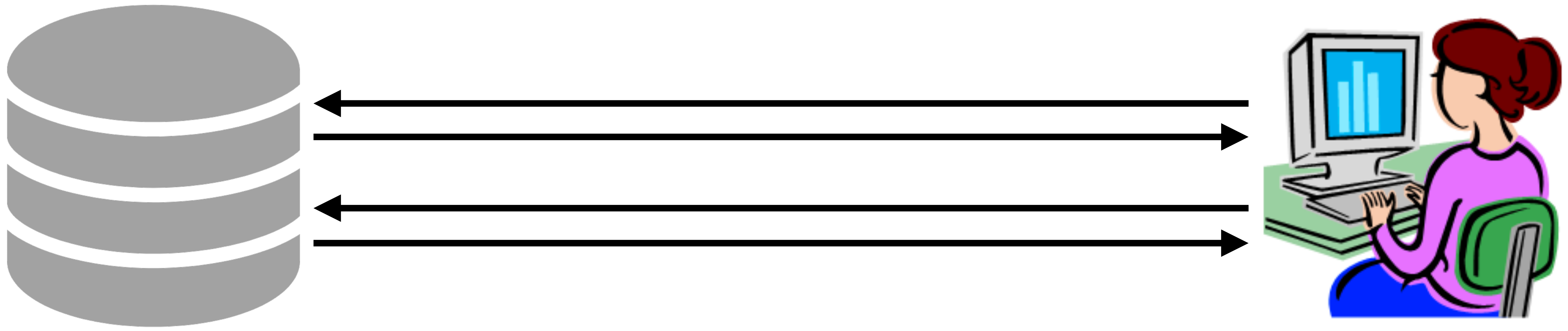
Sage design

Evaluation

Differential Privacy (DP)

(Dwork+ '06)

- Developed to allow privacy-preserving statistical analyses on sensitive datasets (e.g., census, drug purchases, …).
- First (and only) rigorous definition of privacy suitable for this use case.



Definition

- DP is a stability constraint on computations running on datasets: it requires that no single data point in an input dataset has a significant influence on the output.
- To achieve stability, randomness is added into the computation.

Definition

- DP is a stability constraint on computations running on datasets: it requires that no single data point in an input dataset has a significant influence on the output.
- To achieve stability, randomness is added into the computation.

- A randomized computation f: $D \rightarrow O$, is (ε, δ)-DP if for any pair of datasets D and D' differing in one entry, and for any output set $S \subset O$:

$$P(f(D) \in S) \leq e^{\varepsilon} P(f(D') \in S) + \delta$$

DP in ML

- Approach: make training algorithms DP.
- It prevents membership query and reconstruction attacks (Steinke-Ullman '14; Dwork+ '15; Carlini+ '18).
- DP versions exist for most ML training algorithms:
  - Stochastic gradient descent (SGD) (Abadi+16, Yu+19).
  - Various regressions (Chaudhuri+08, Kifer+12, Nikolaenko+13, Talwar+15).
  - Collaborative filtering (McSherry+09).
  - Language models (McMahan+18).
  - Feature and model selection (Chaudhuri+13, Smith+13).
  - Model evaluation (Boyd+15).
  - Tensorflow/privacy implements several of these algorithms (McMahan+19).

Outline

Motivation

Differential Privacy

Two practical challenges

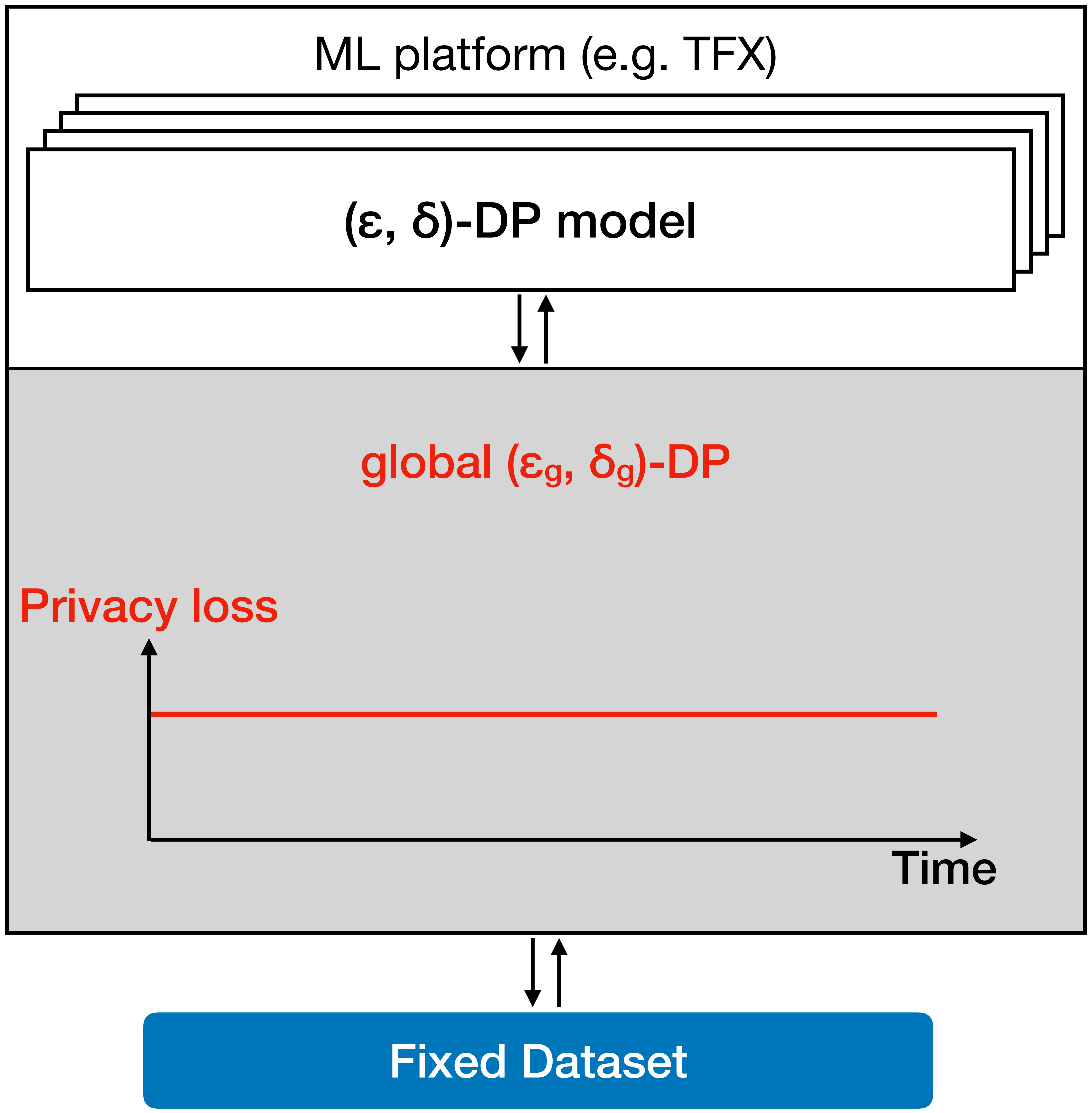
Sage design

Evaluation

Challenge 1 - Running out of privacy budget

Most DP work focuses on a fixed database model:

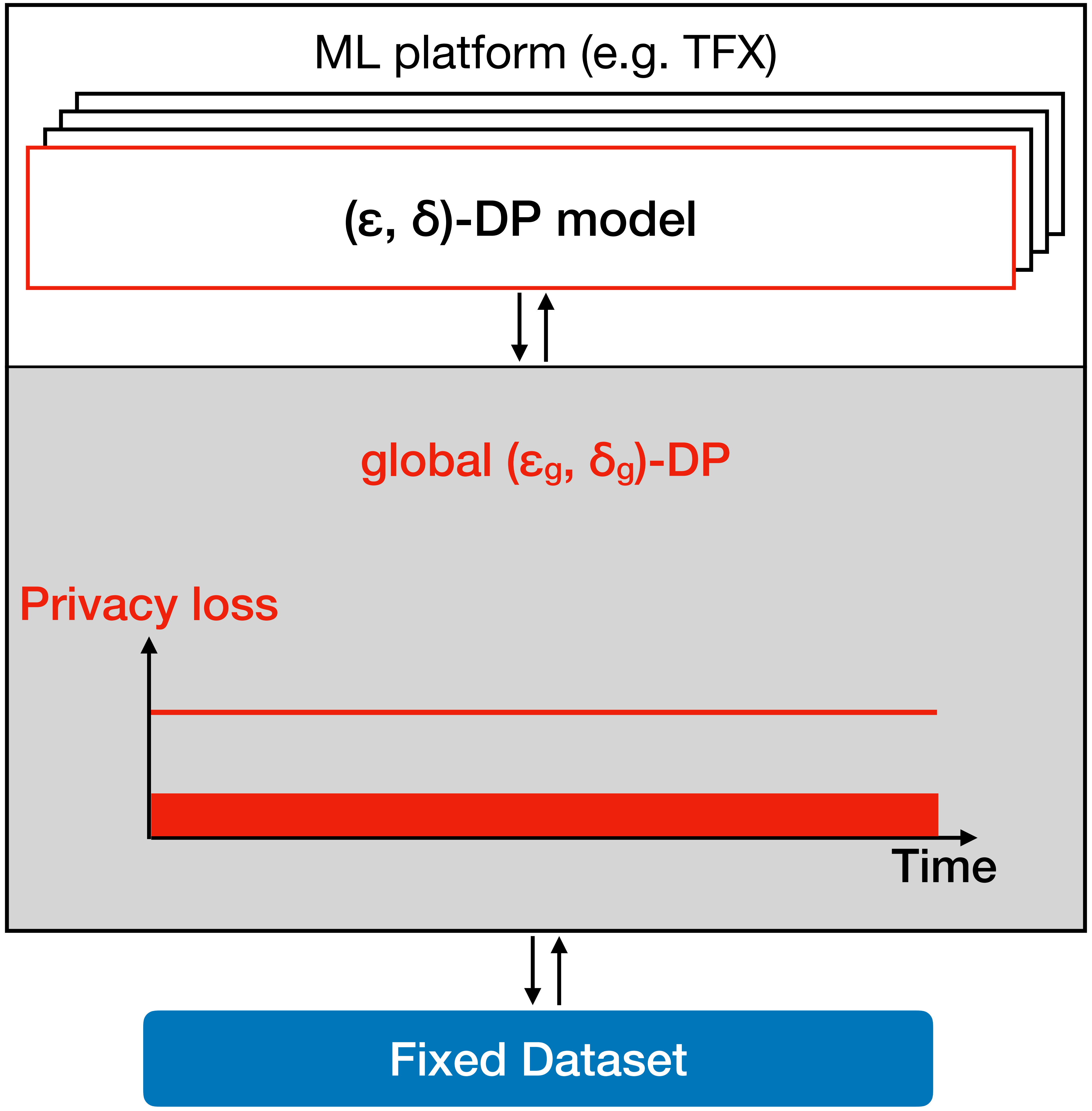
- Each model consumes some privacy budget.
- When the budget is exhausted, the data cannot be used anymore: the system can "run out of budget".



Challenge 1 - Running out of privacy budget

Most DP work focuses on a fixed database model:

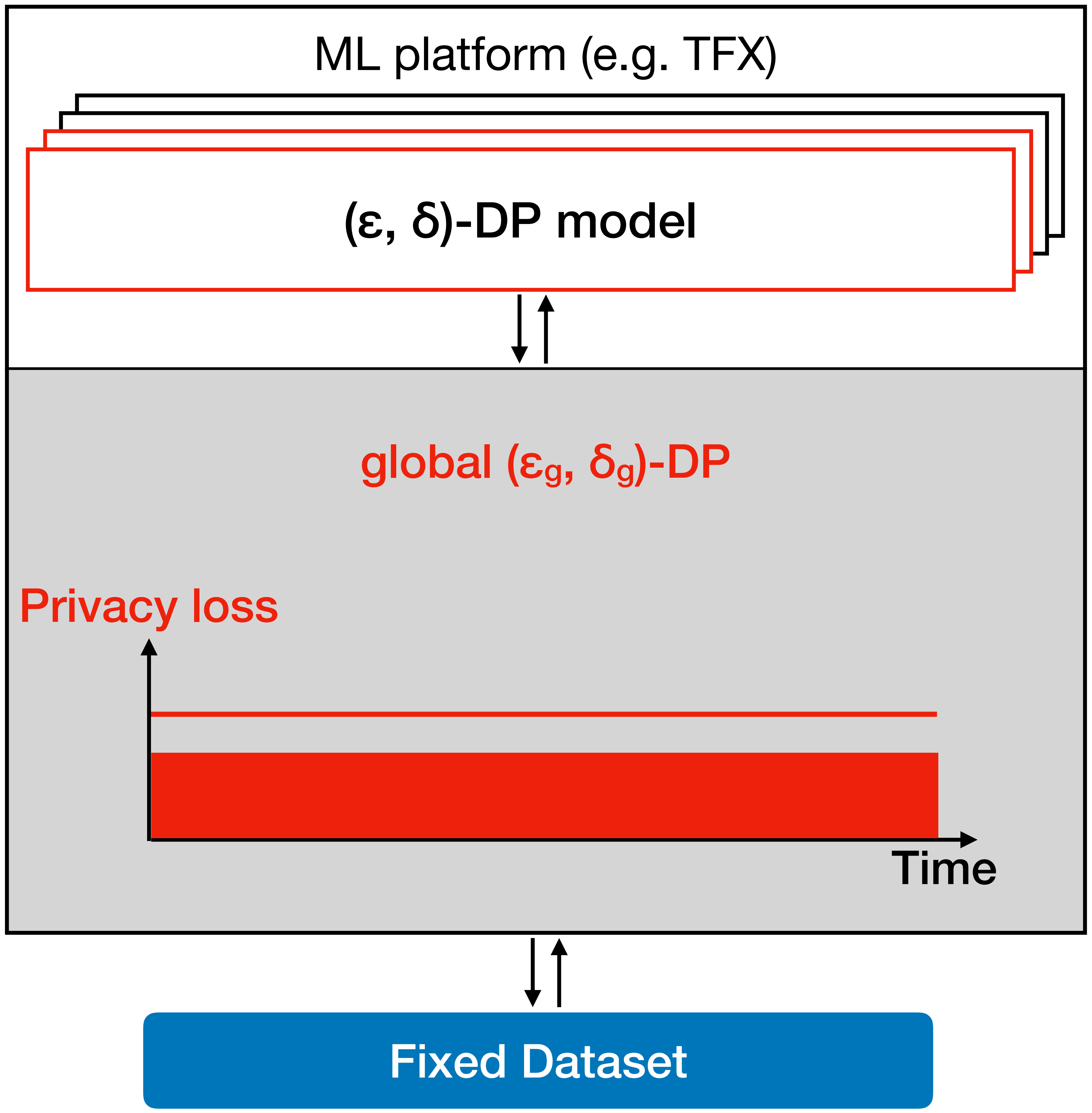
- Each model consumes some privacy budget.
- When the budget is exhausted, the data cannot be used anymore: the system can "run out of budget".



Challenge 1 - Running out of privacy budget

Most DP work focuses on a fixed database model:

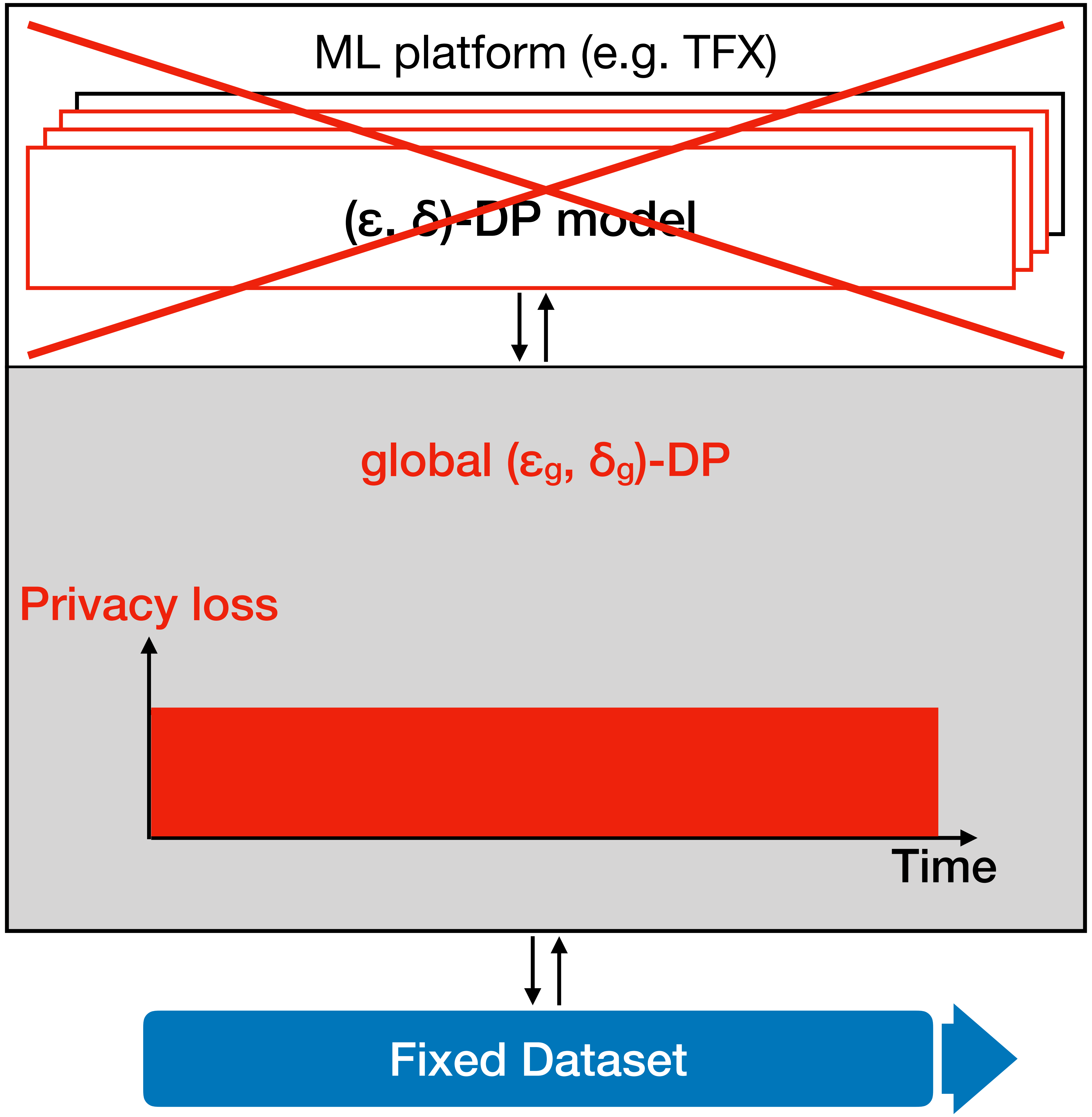
- Each model consumes some privacy budget.
- When the budget is exhausted, the data cannot be used anymore: the system can "run out of budget".



Challenge 1 - Running out of privacy budget

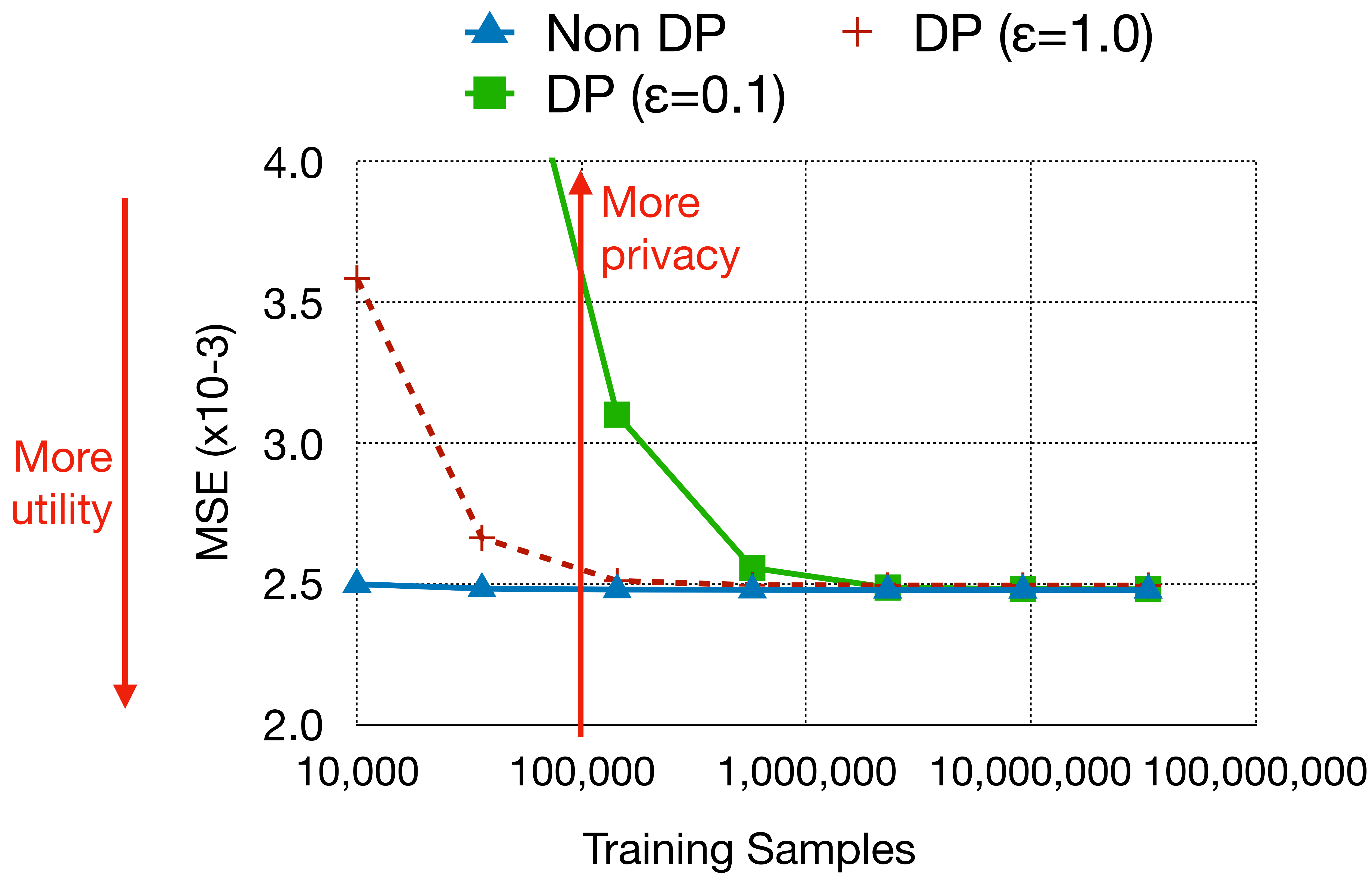
Most DP work focuses on a fixed database model:

- Each model consumes some privacy budget.
- When the budget is exhausted, the data cannot be used anymore: the system can "run out of budget".

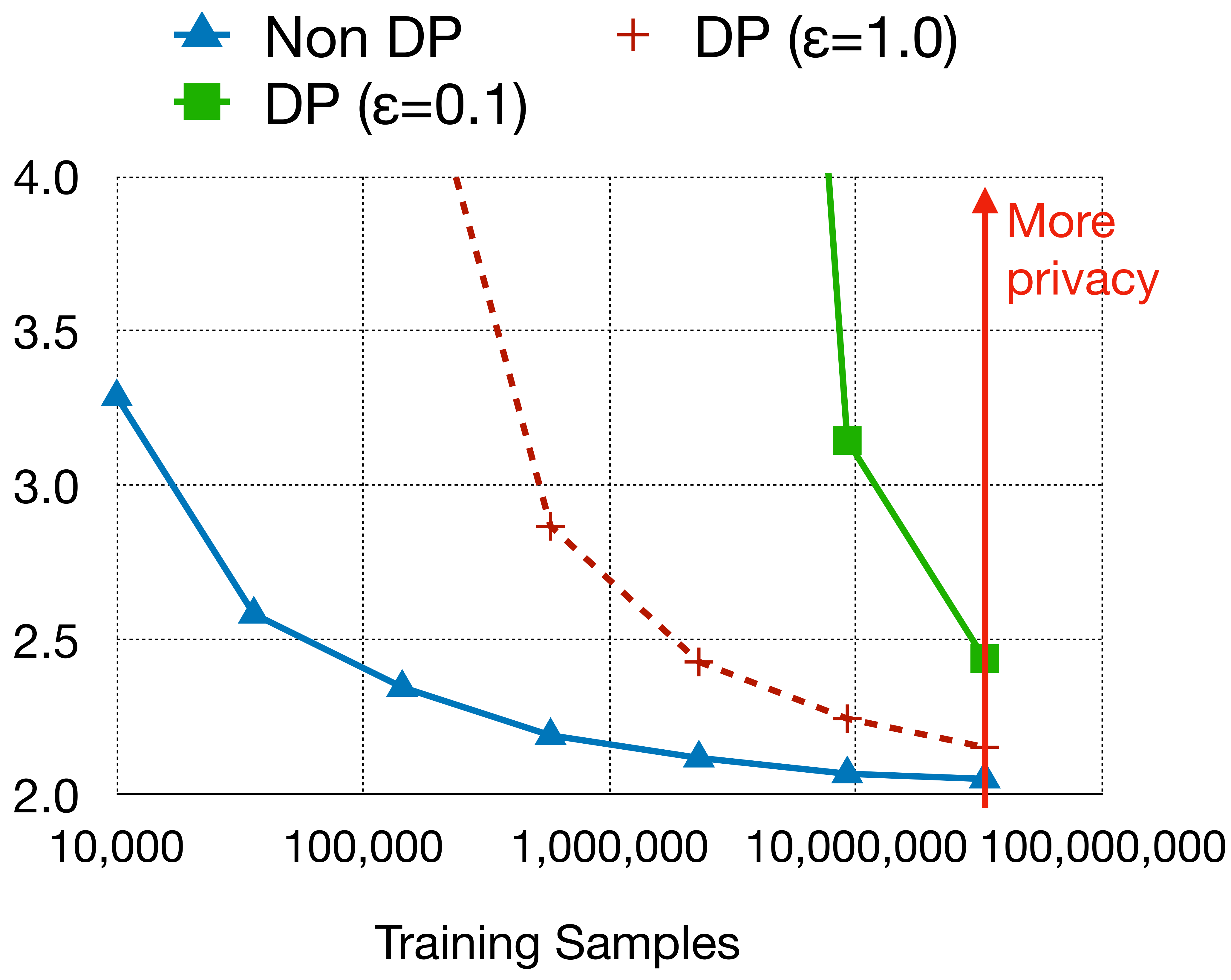


Challenge 2 - Privacy/utility trade-off

Challenge 2 - Privacy/utility trade-off



Linear Regression



Deep Neural Network

Outline

Motivation

Differential Privacy

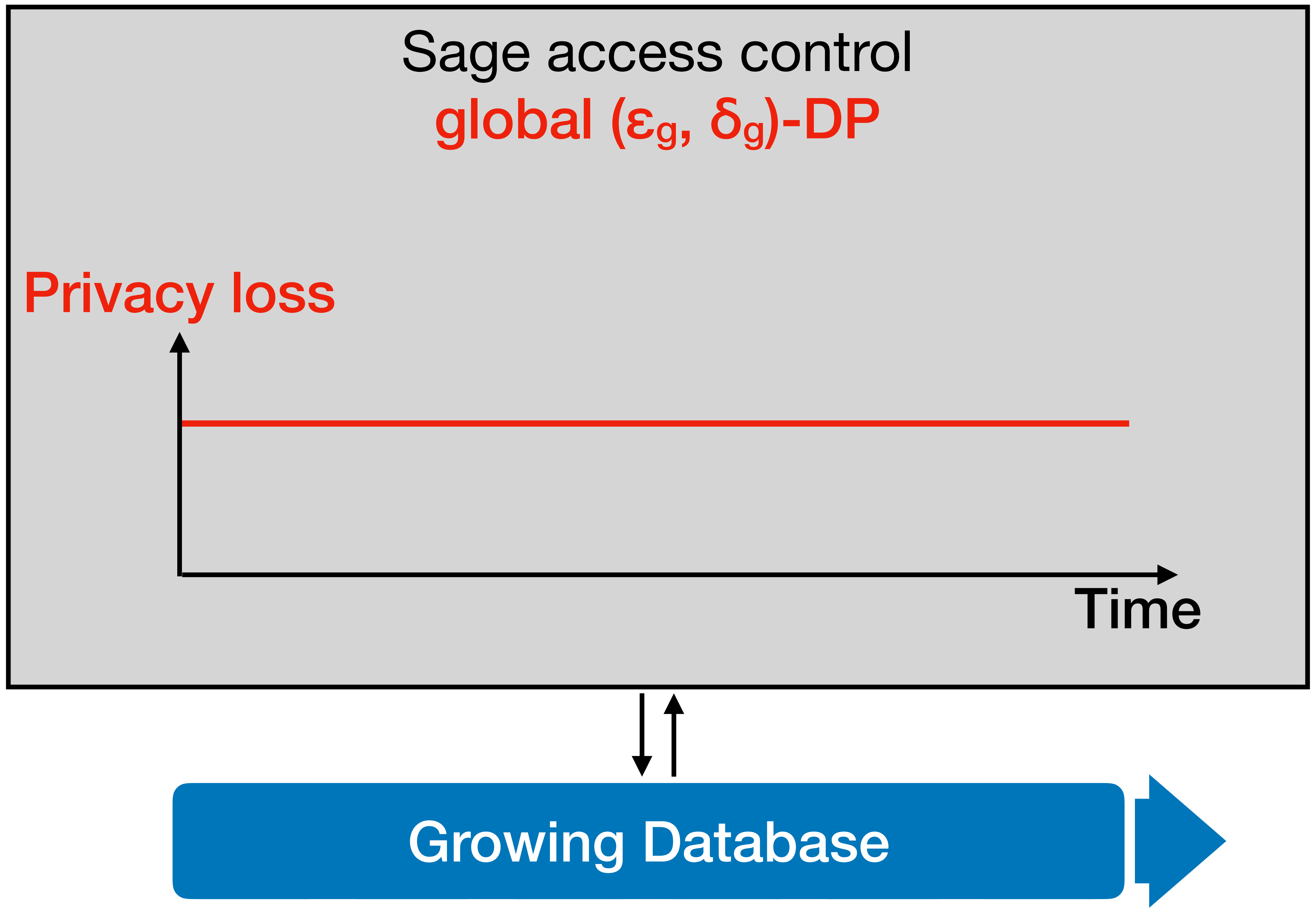
Two practical challenges

Sage design

Evaluation

Sage block composition (challenge 1)

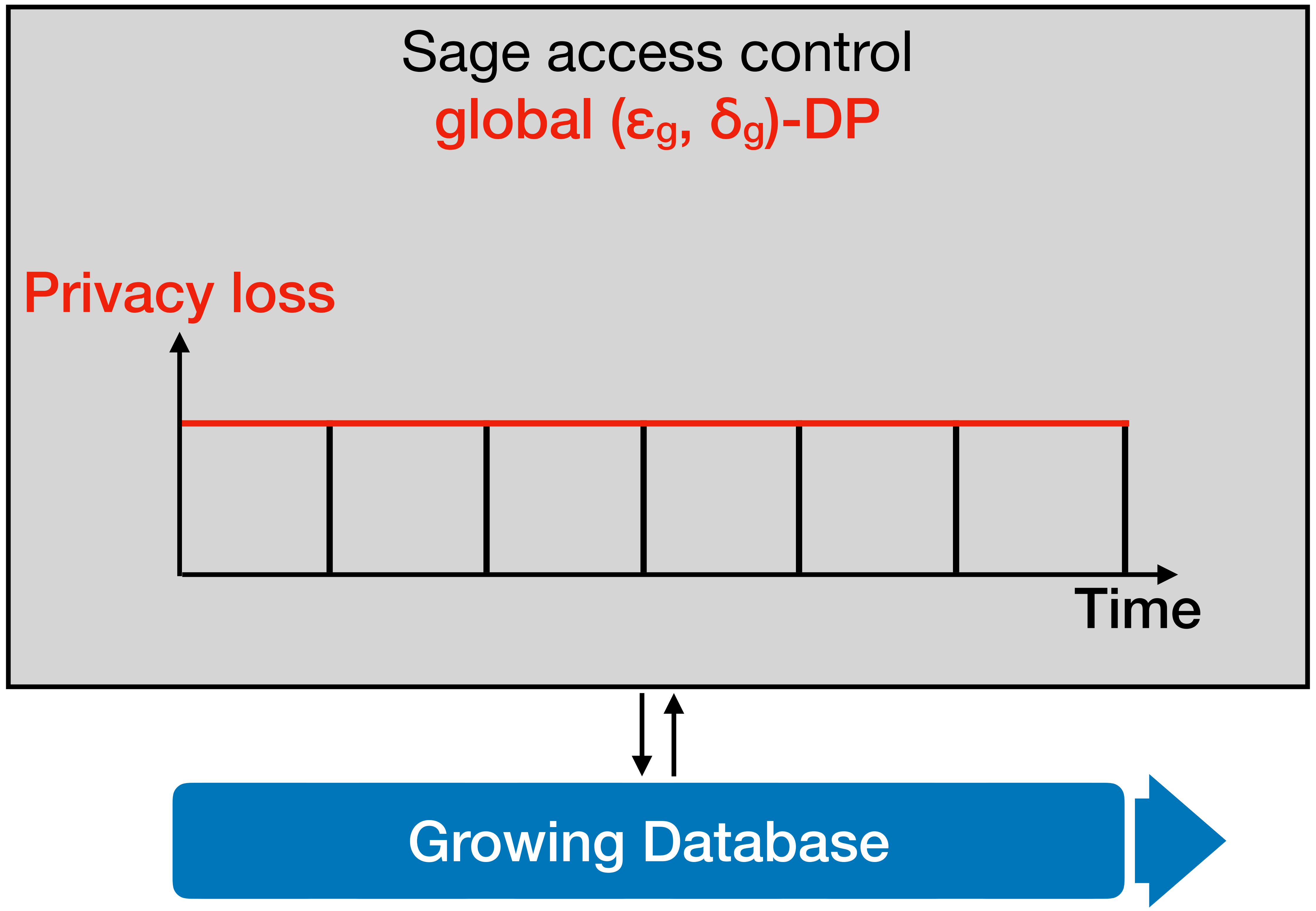
Key realization: ML platforms operate on a growing database.



Sage block composition (challenge 1)

Interaction model:

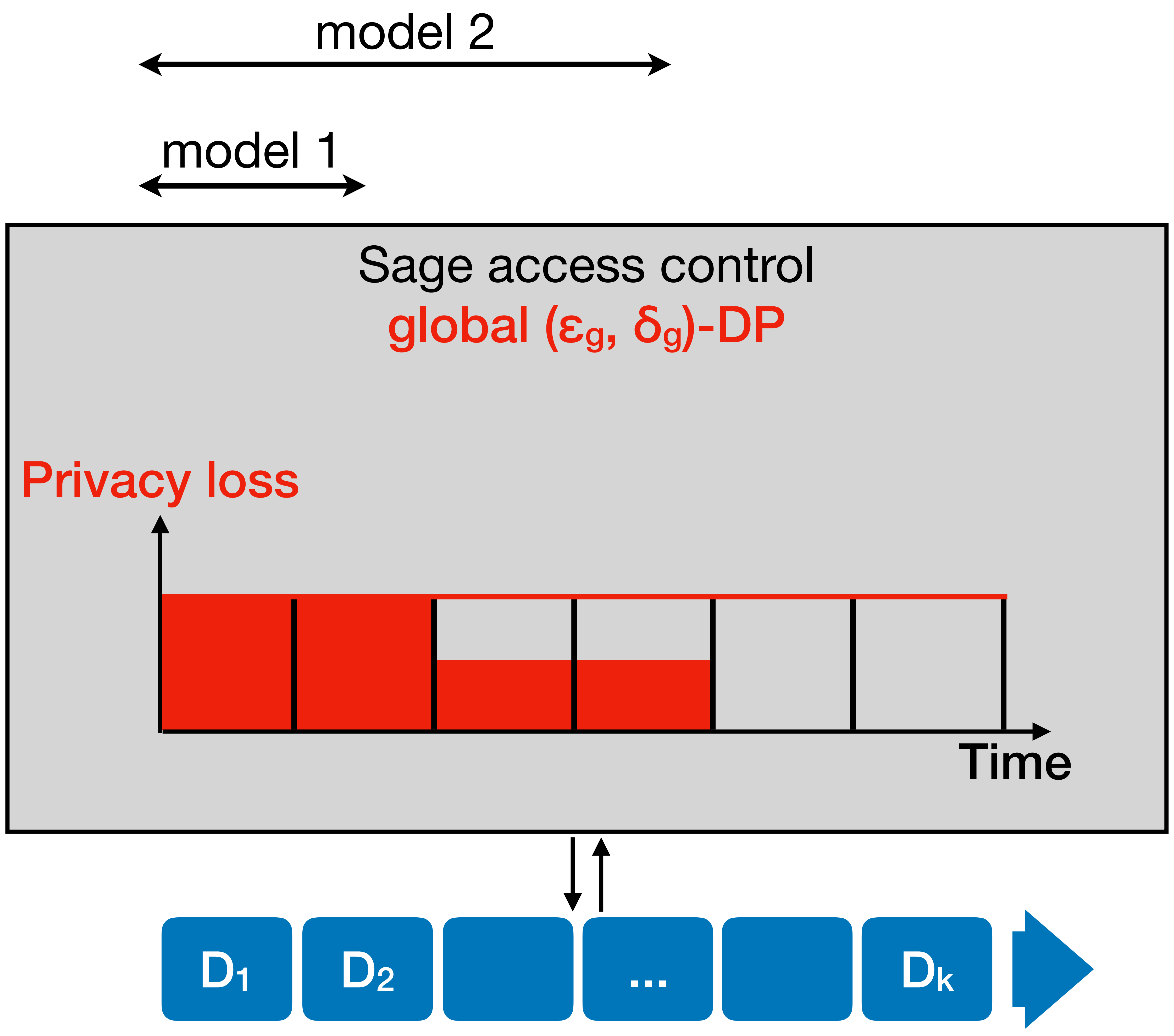
- Split the growing database into time based blocks.
- Models can adaptively combine blocks to form larger datasets.
- Account for privacy loss only against blocks used by each models.
- Models can influence future data and privacy budgets.



Sage block composition (challenge 1)

Interaction model:

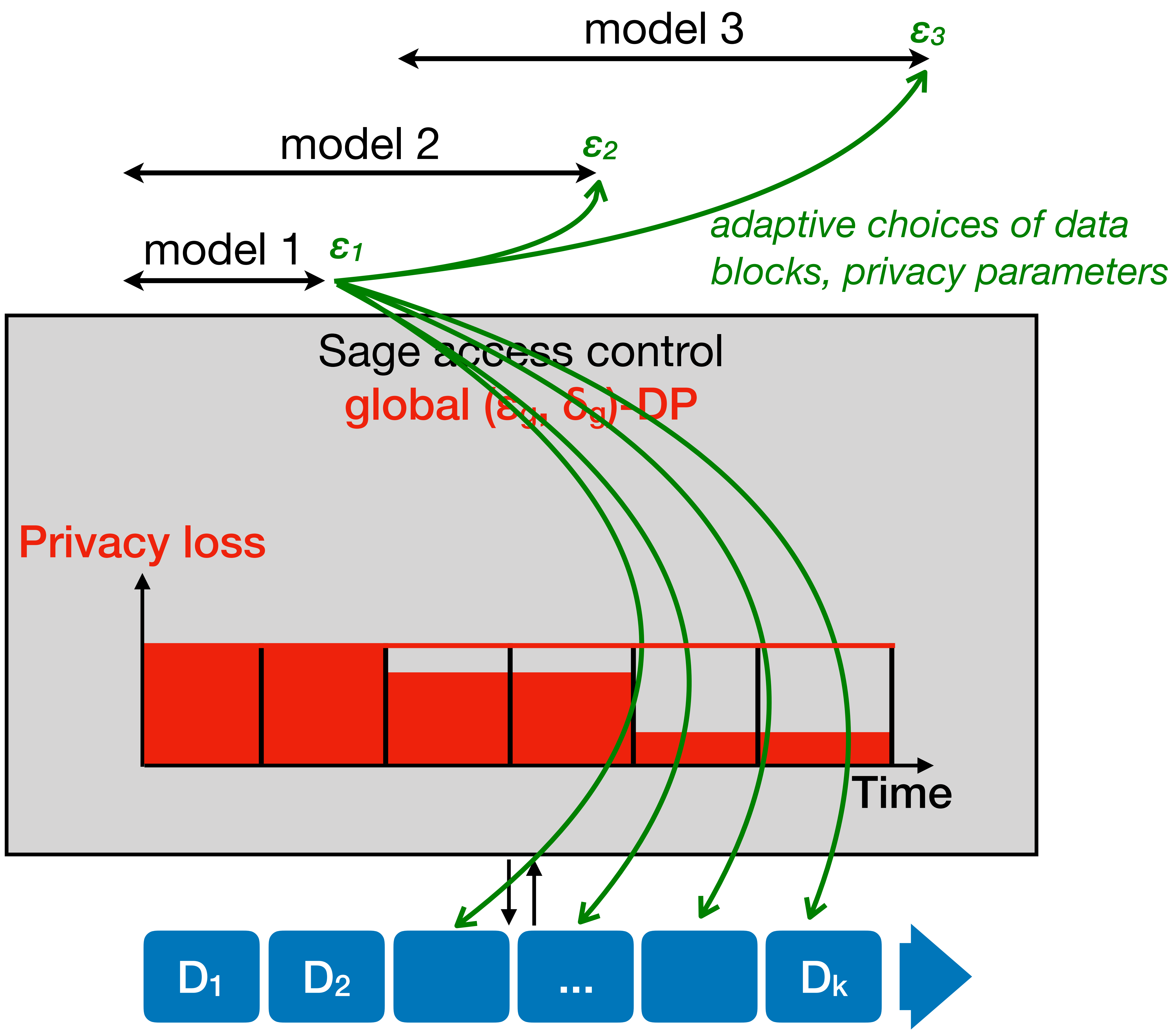
- Split the growing database into time based blocks.
- Models can adaptively combine blocks to form larger datasets.
- Account for privacy loss only against blocks used by each models.
- Models can influence future data and privacy budgets.



Sage block composition (challenge 1)

Interaction model:

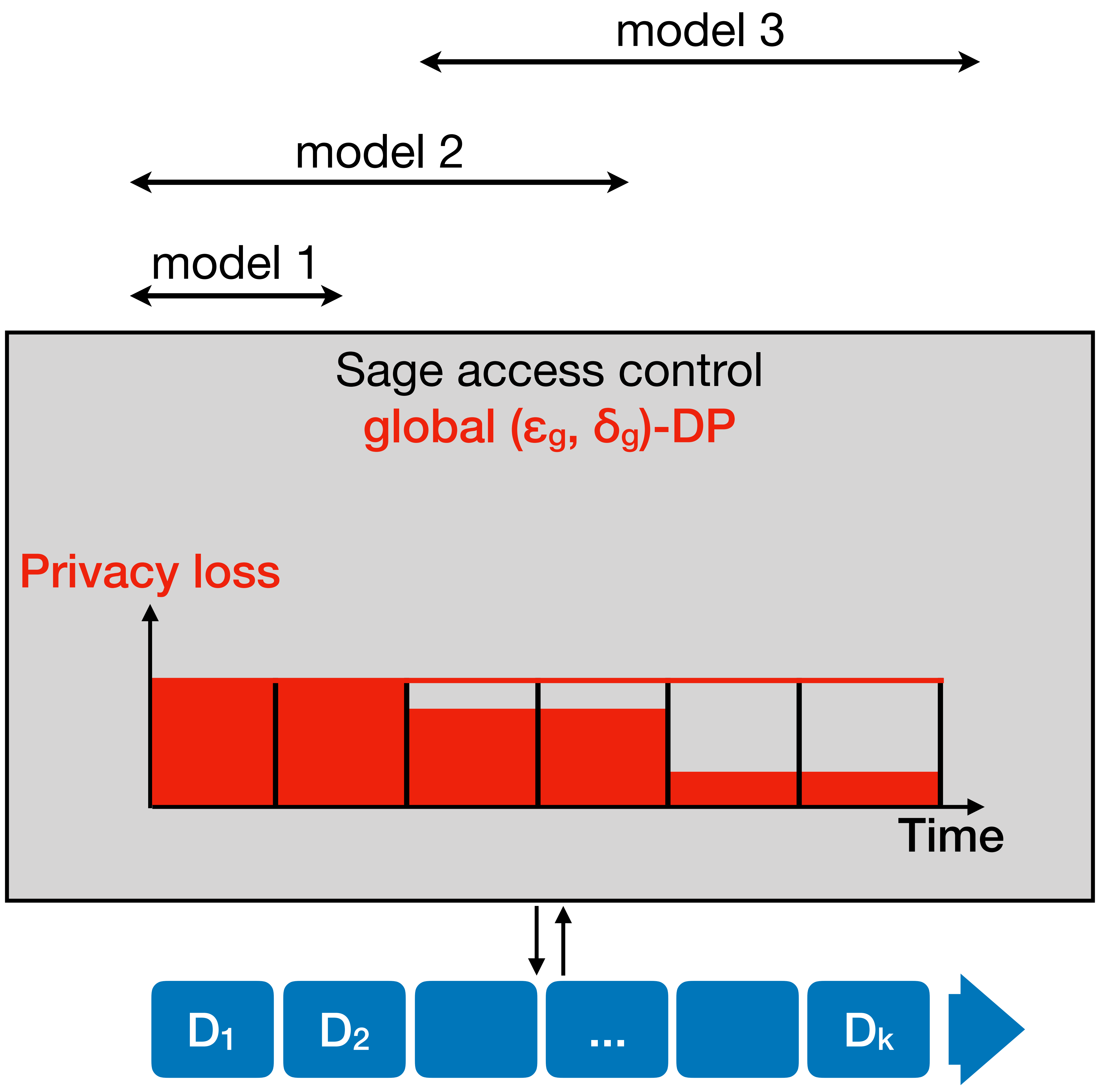
- Split the growing database into time based blocks.
- Models can adaptively combine blocks to form larger datasets.
- Account for privacy loss only against blocks used by each models.
- Models can influence future data and privacy budgets.



Sage block composition (challenge 1)

Theorem:

$|\ \text{PrivacyLoss(stream)}\ | \leq \max_k |\ \text{PrivacyLoss}(D_k)\ |$



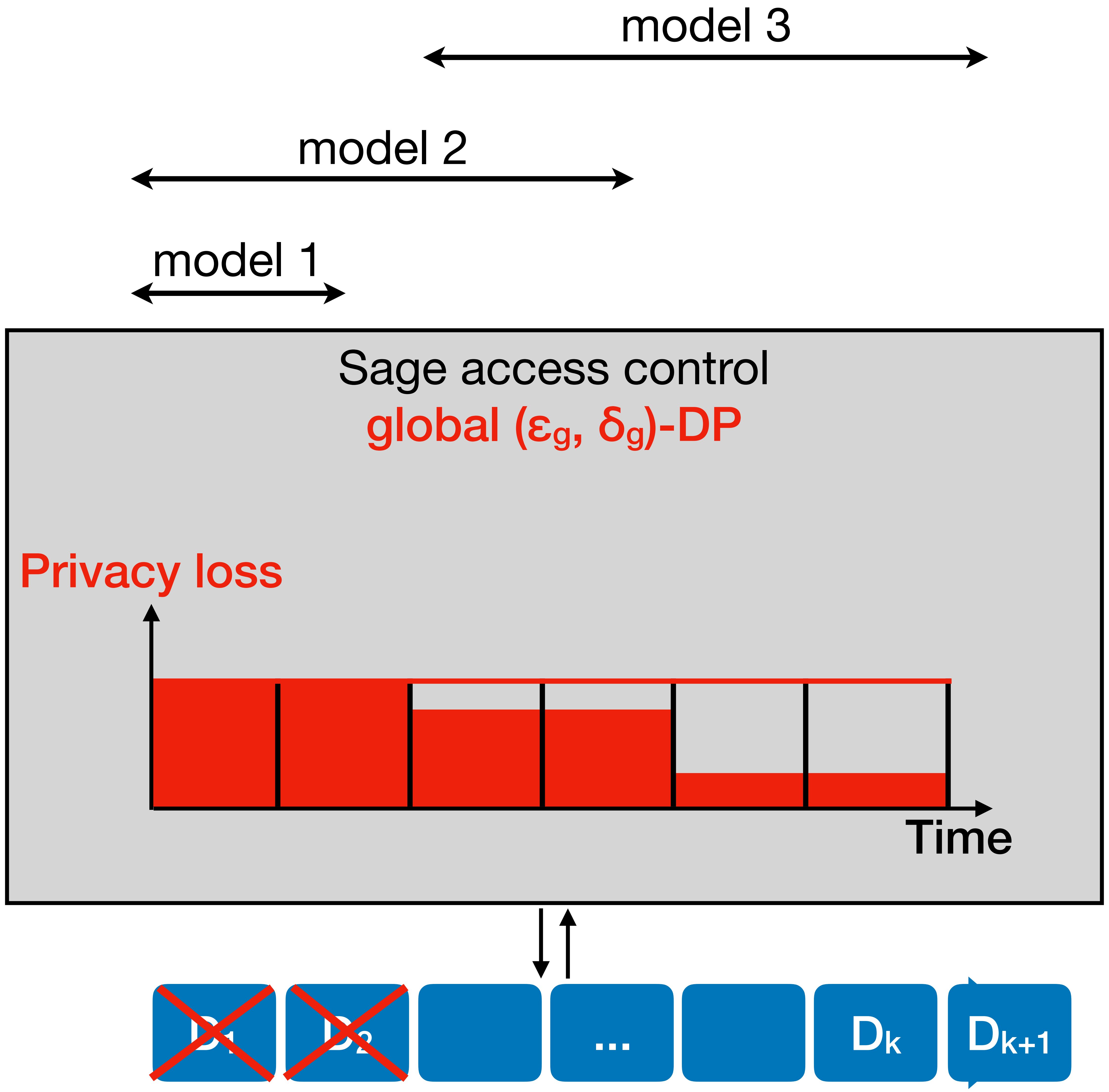
Sage block composition (challenge 1)

Theorem:

$| \text{PrivacyLoss}(\text{stream}) | \leq \max_k | \text{PrivacyLoss}(D_k) |$

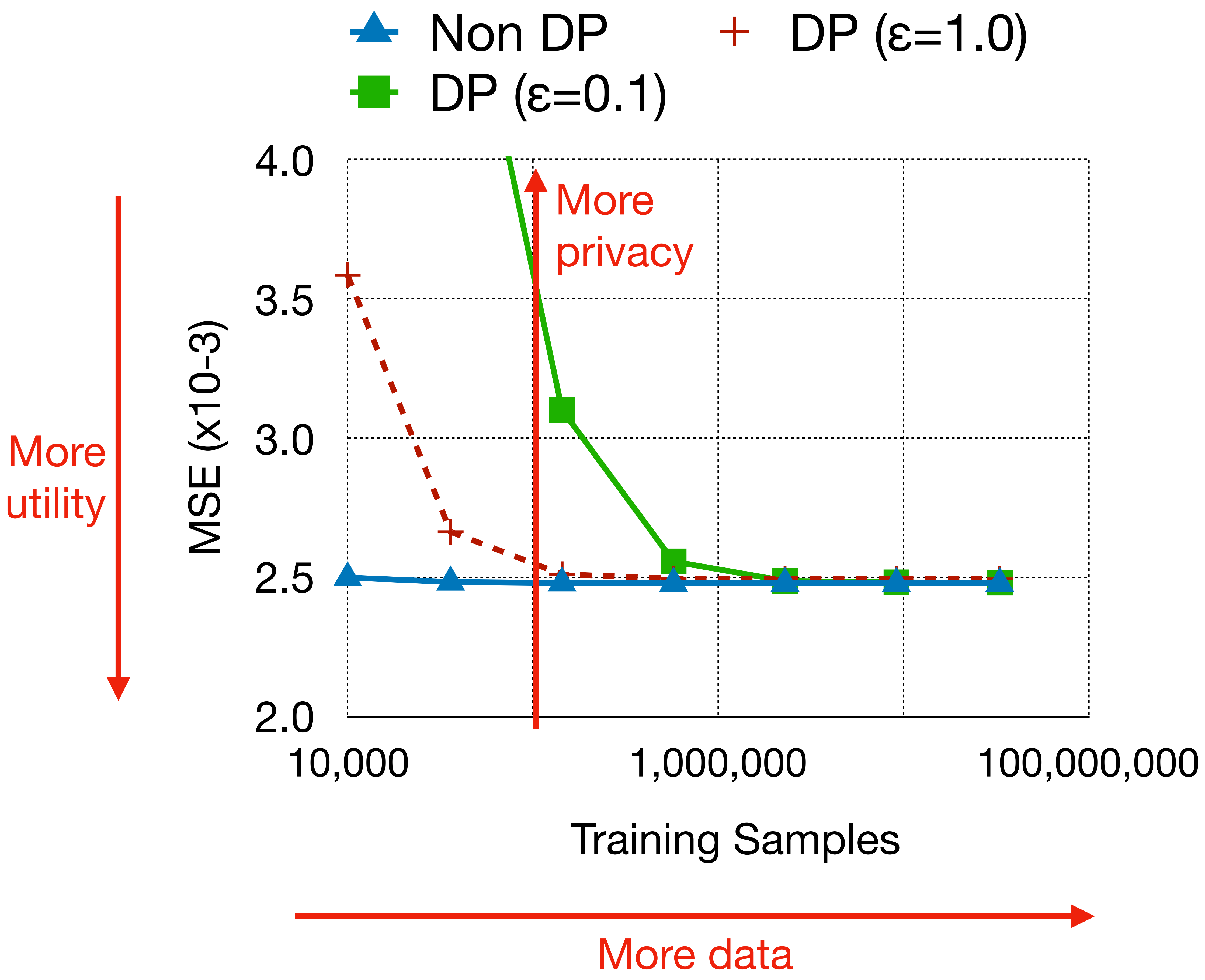
Why is this important?

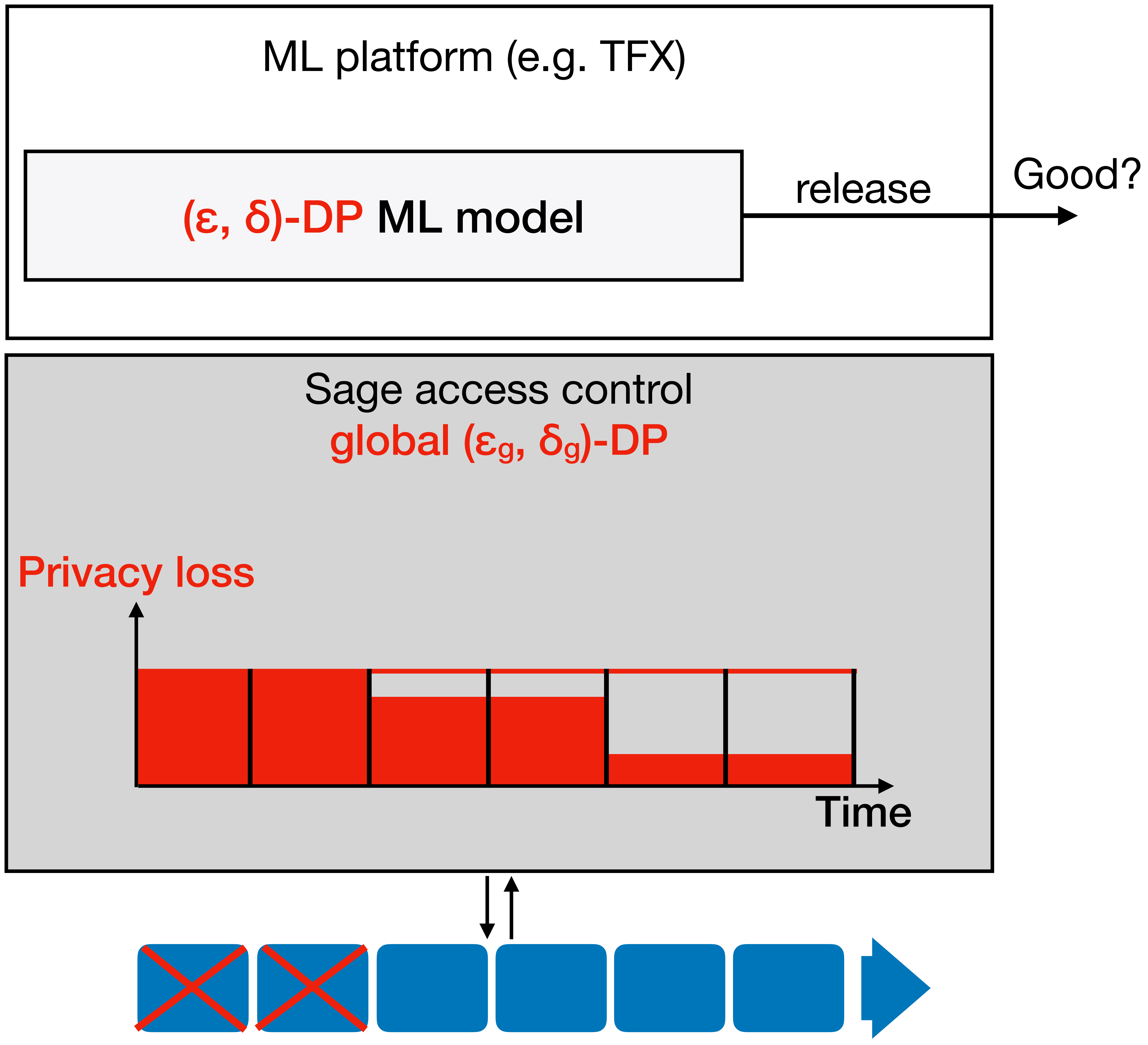
- Controlling each block's privacy loss controls the global privacy loss.
- New blocks arrive with zero loss and constantly renew the budget.



Iterative training (challenge 2)

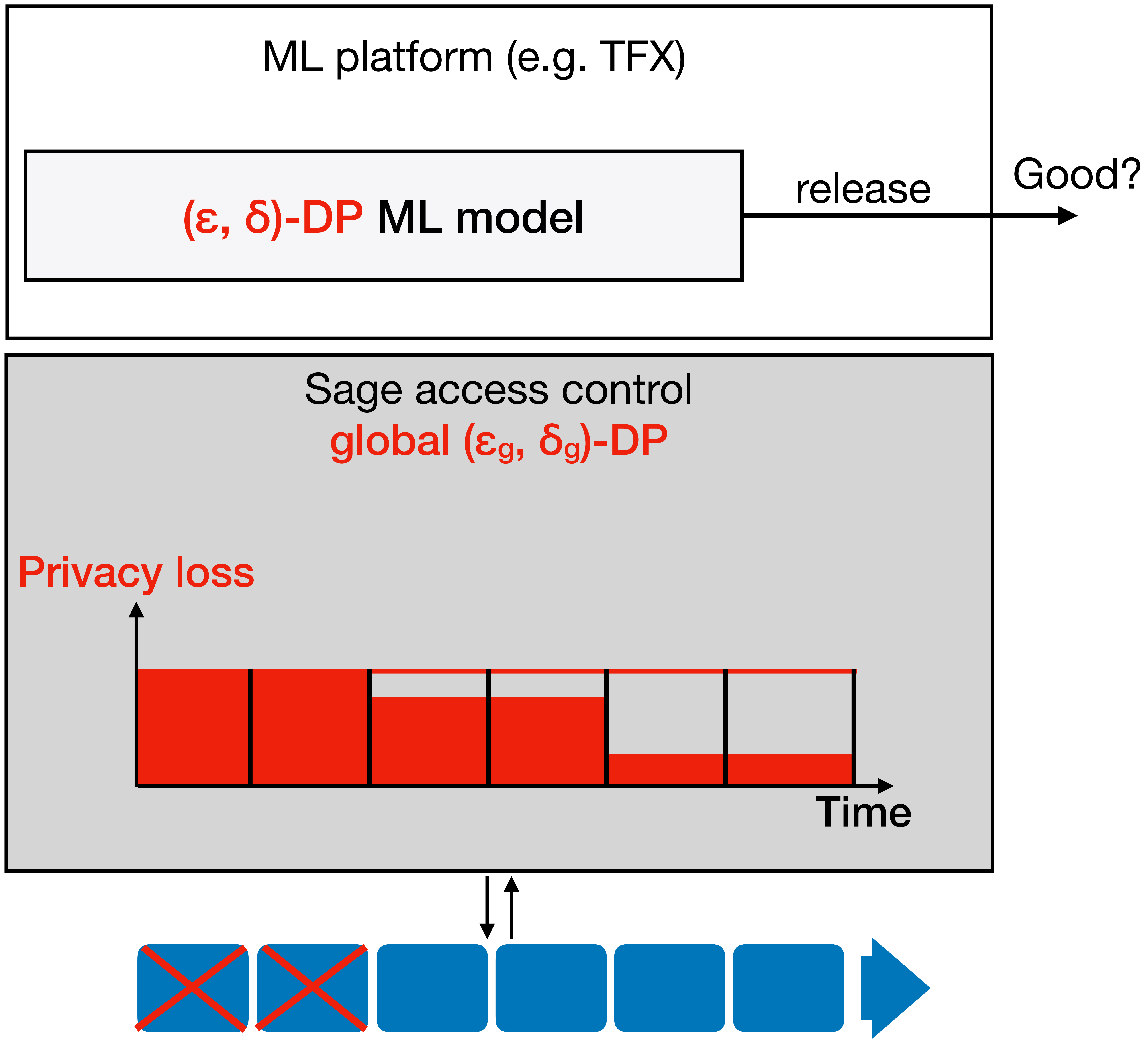
Iterative training (challenge 2)





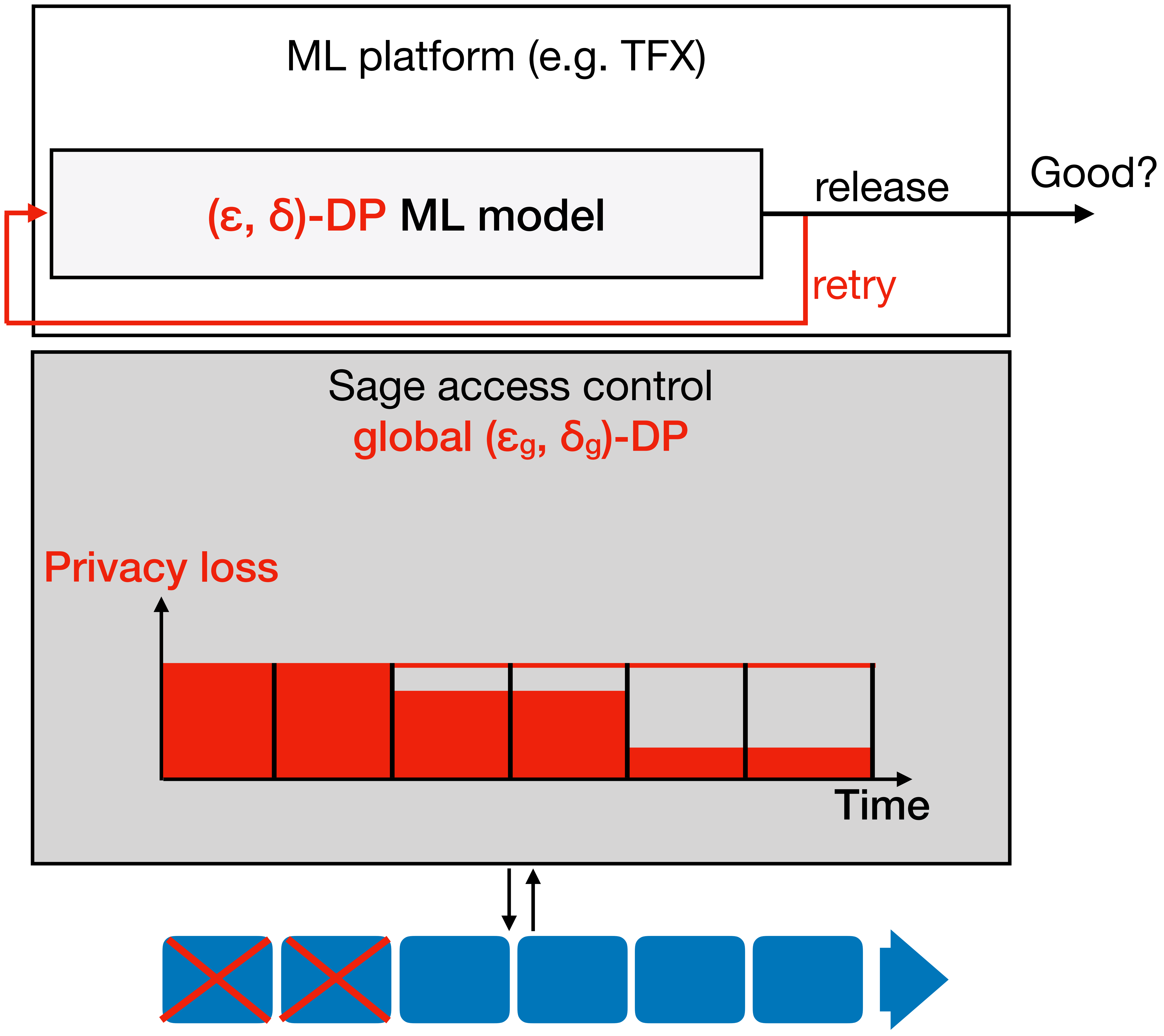
Iterative training (challenge 2)

- Adaptively trains on growing data and/or privacy budgets.
- Release when w.h.p. model accuracy surpasses a target.
- Accounts for the impact of DP noise in TFX-evaluate to give high-probability assessment of model accuracy.



Iterative training (challenge 2)

- Adaptively trains on growing data and/or privacy budgets.
- Release when w.h.p. model accuracy surpasses a target.
- Accounts for the impact of DP noise in TFX-evaluate to give high-probability assessment of model accuracy.

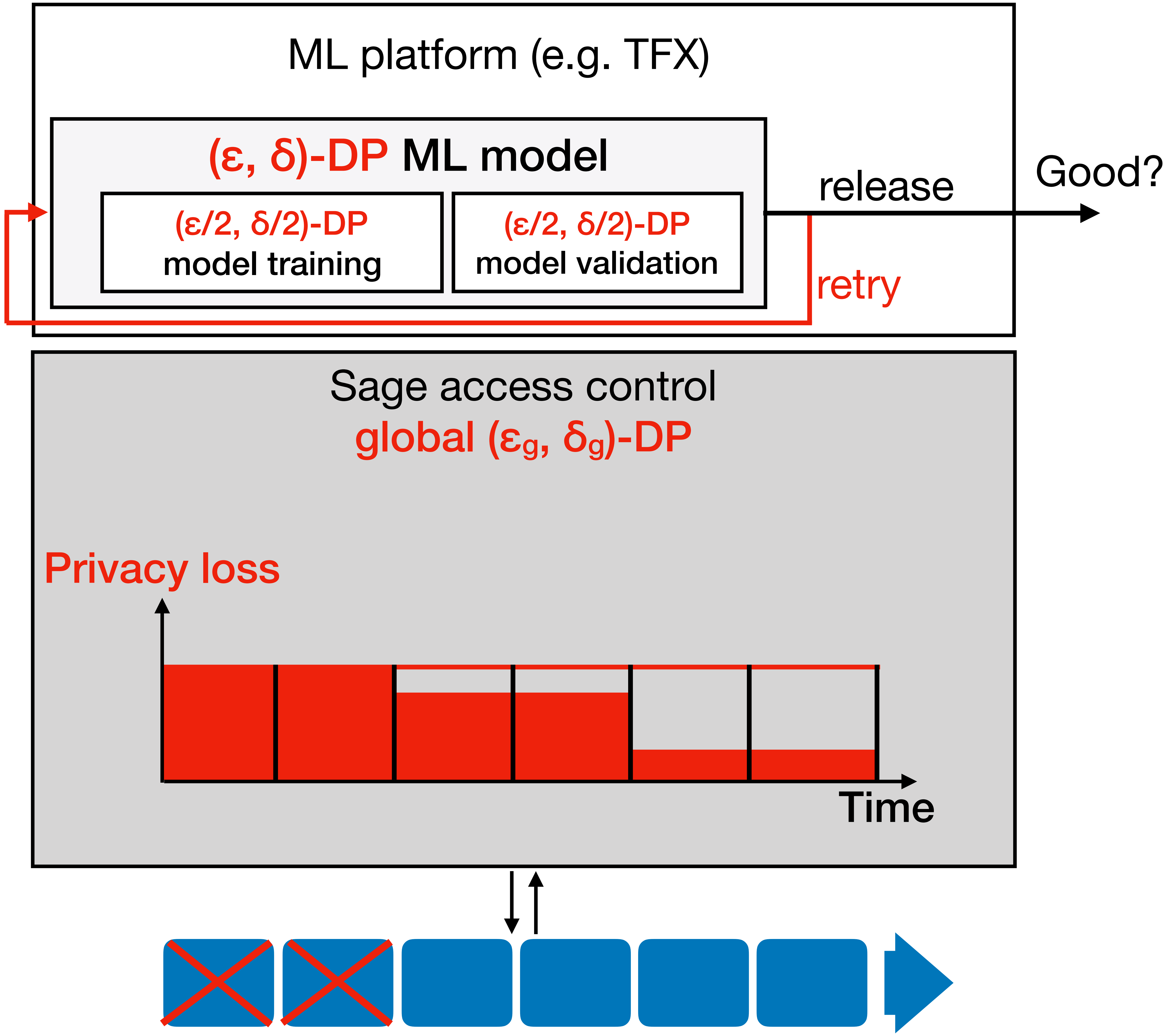


Iterative training (challenge 2)

- Adaptively trains on growing data and/or privacy budgets.
- Release when w.h.p. model accuracy surpasses a target.
- Accounts for the impact of DP noise in TFX-evaluate to give high-probability assessment of model accuracy.

Statistical test for evaluation:

P(acc < τ) ≤ η over sampling of test set.



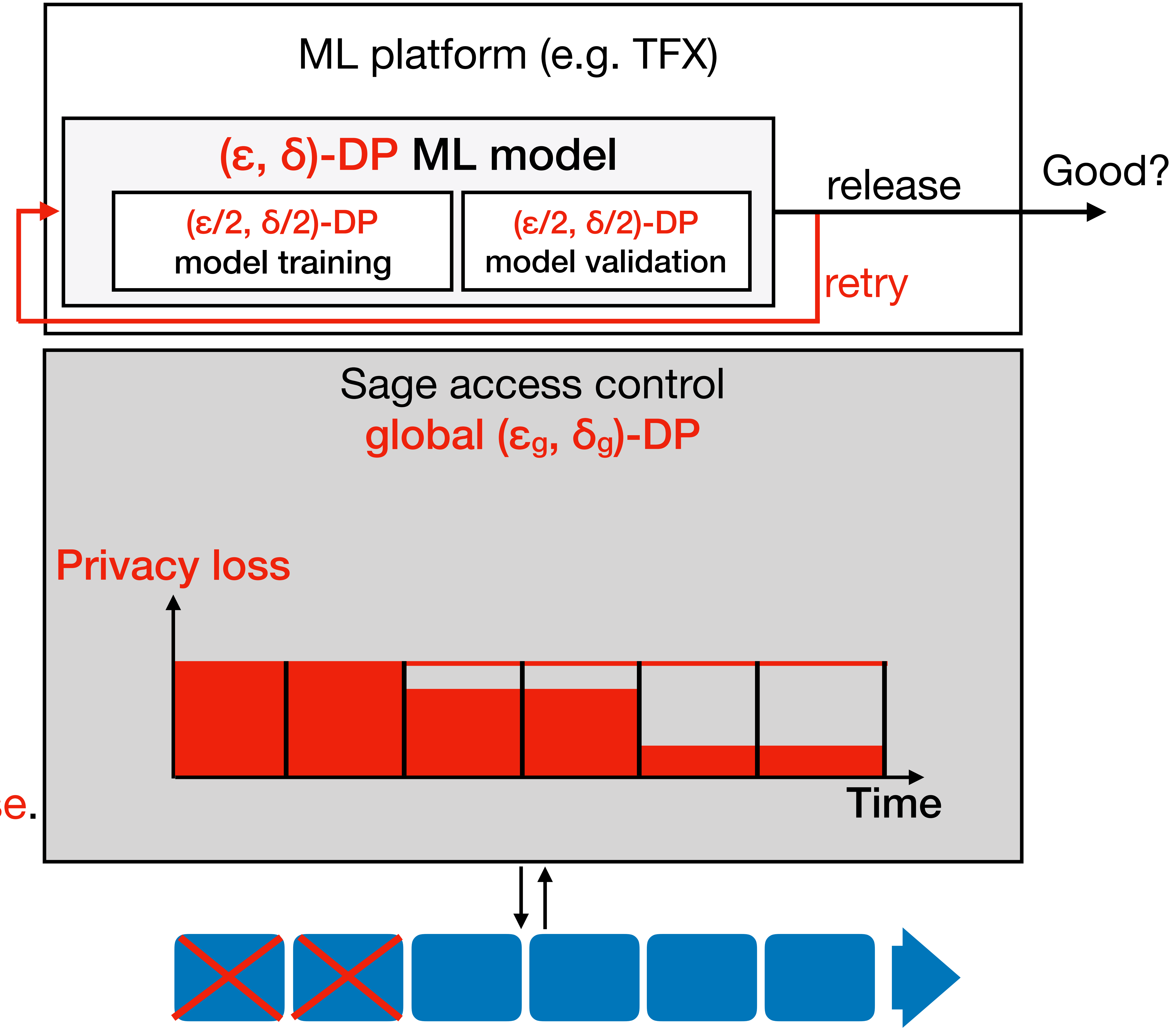
Iterative training (challenge 2)

- Adaptively trains on growing data and/or privacy budgets.
- Release when w.h.p. model accuracy surpasses a target.
- Accounts for the impact of DP noise in TFX-evaluate to give high-probability assessment of model accuracy.

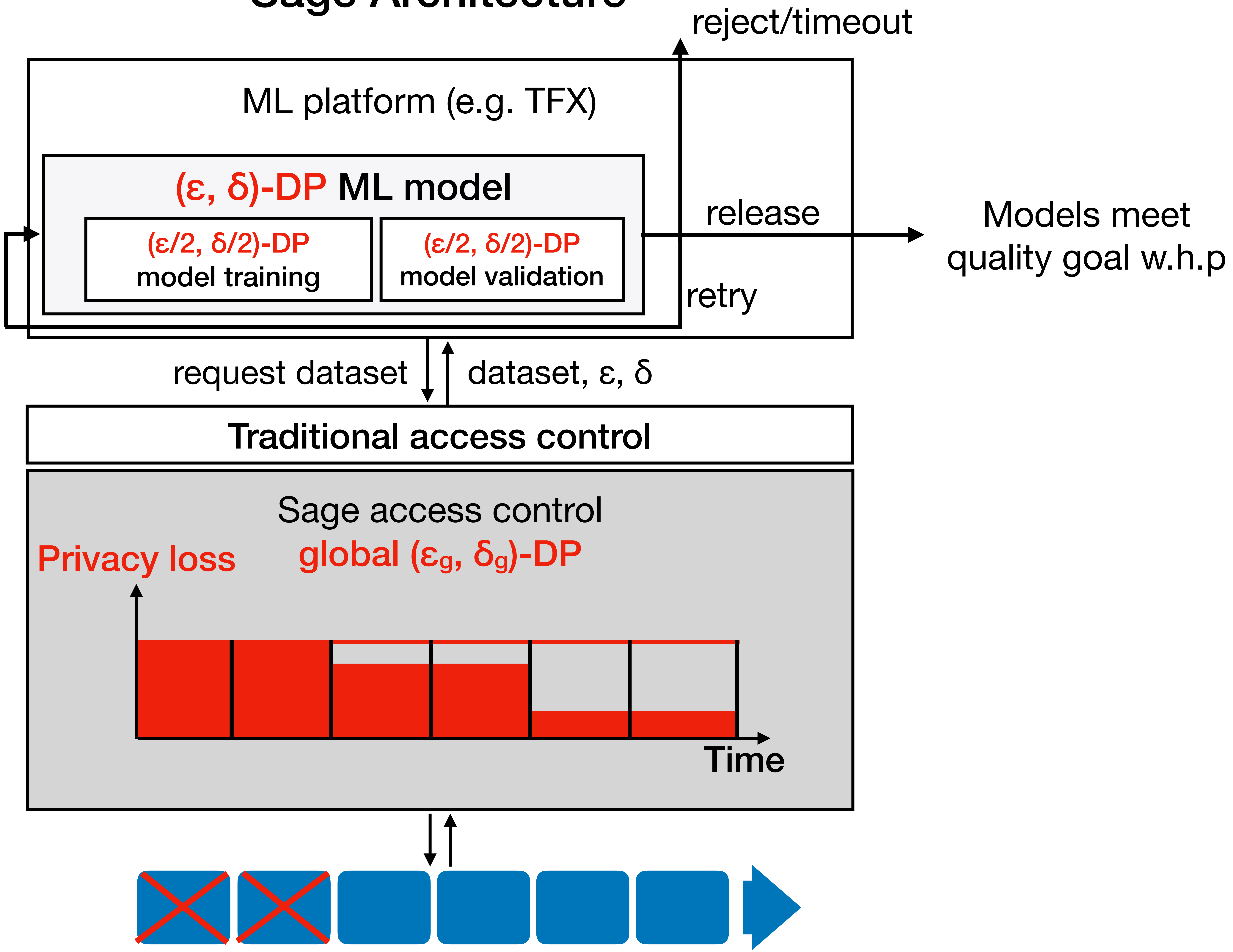
Statistical test for evaluation:

~~P(acc < τ) ≤~~ η over sampling of test set and DP noise.

$$\overline{\mathcal{L}}^{dp}_{te}(f^{\mathrm{dp}}) + \sqrt{\frac{2B\overline{\mathcal{L}}^{dp}_{te}(f^{\mathrm{dp}})\ln(3/\eta)}{\underline{n}^{\mathrm{dp}}_{\mathrm{te}}}} + \frac{4B\ln(3/\eta)}{\underline{n}^{\mathrm{dp}}_{\mathrm{te}}} \leq \tau_{loss}$$



Sage Architecture



Outline

Motivation

Differential Privacy

Two practical challenges

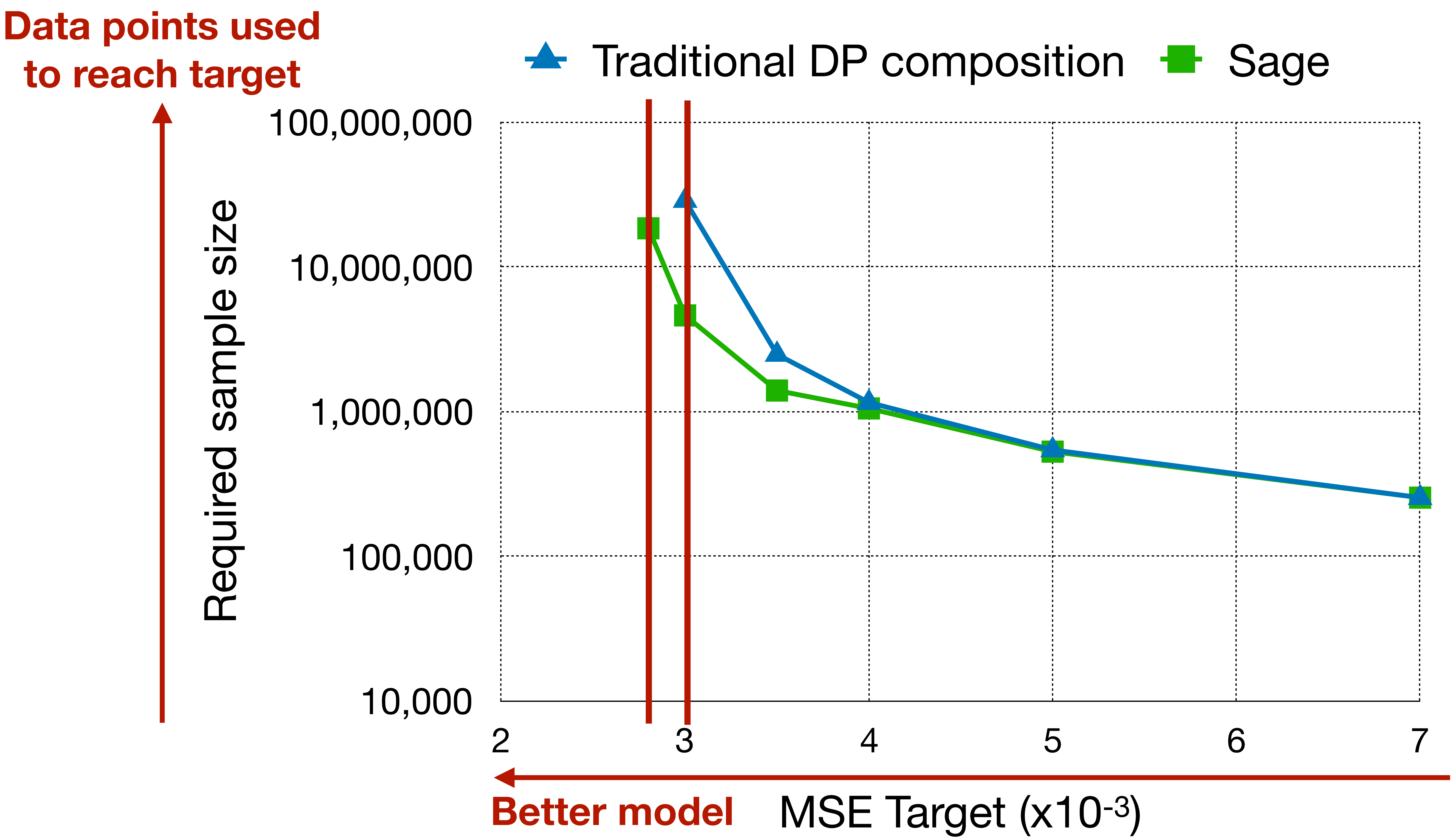
Sage design

Evaluation

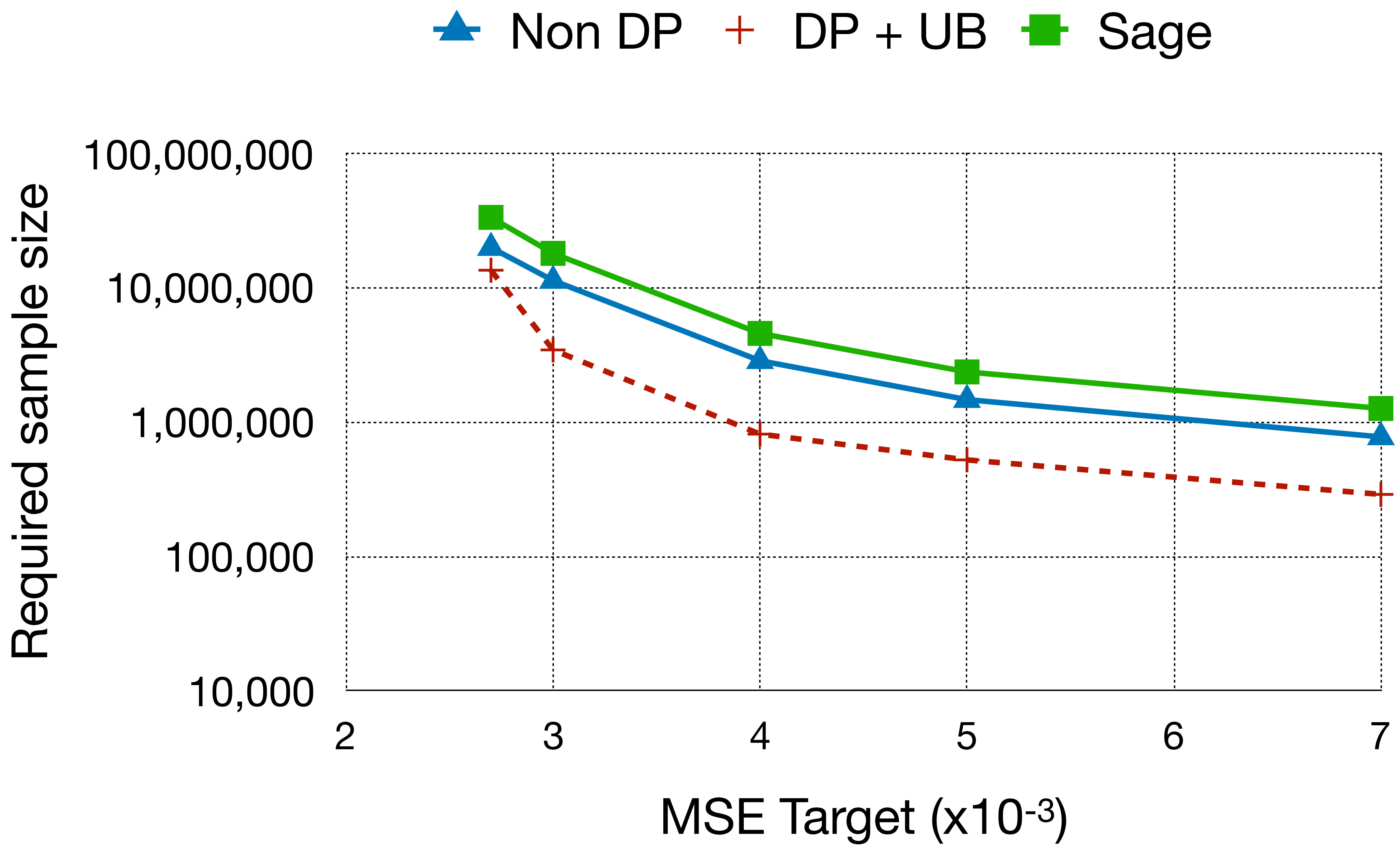
Evaluation:

1. Benefits of block composition versus traditional DP composition.
2. Importance of iterative training and DP aware performance tests.
3. Continuous operation on multiple models and growing database.

1. Benefits of block composition versus traditional DP composition

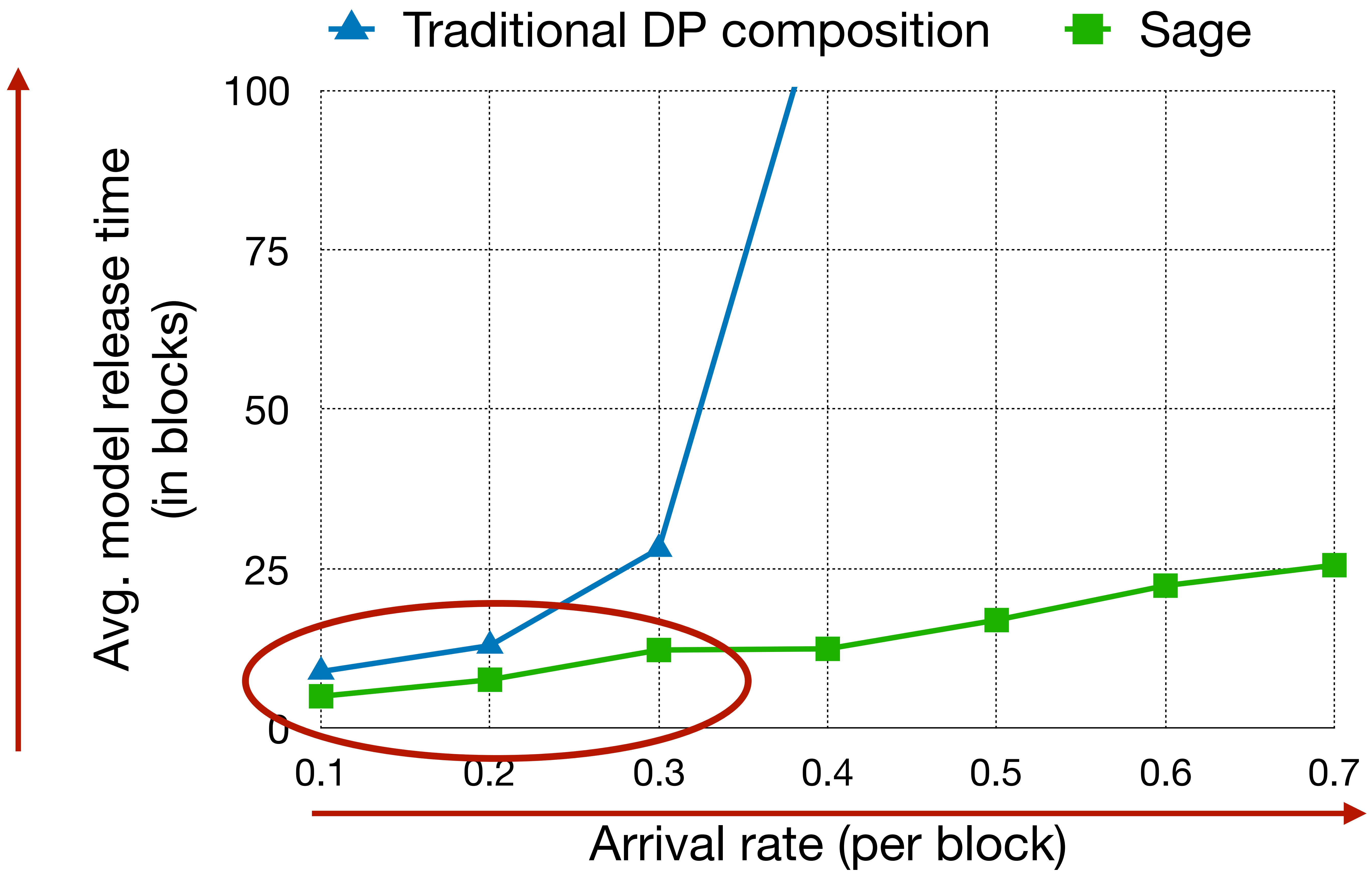


2. Importance of iterative training and DP aware performance tests



Test methodology	Non DP	DP + UB	Sage
Failure rate at 1% proba.	0.2%	1.7%	0.3%

3. Continuous operation on multiple models and growing database



Summary

- DP literature has mostly focused on individual ML algorithms running on static databases (which don't incorporate new data).
- ML workloads operate on growing databases: models incorporate new data and (adaptively) reuse old data.
- Sage is the first to adapt DP theory and practice to ML workloads on growing databases, for data protection.
  - Opens an exciting design space for efficient privacy resource allocation!