Making Algorithms Forget





Data Externalities Tutorial - FAccT 2021

Mihaela Curmei, Andreas Haupt, Charles Cui, Yixin Wang, Rediet Abebe

FAccT 2021

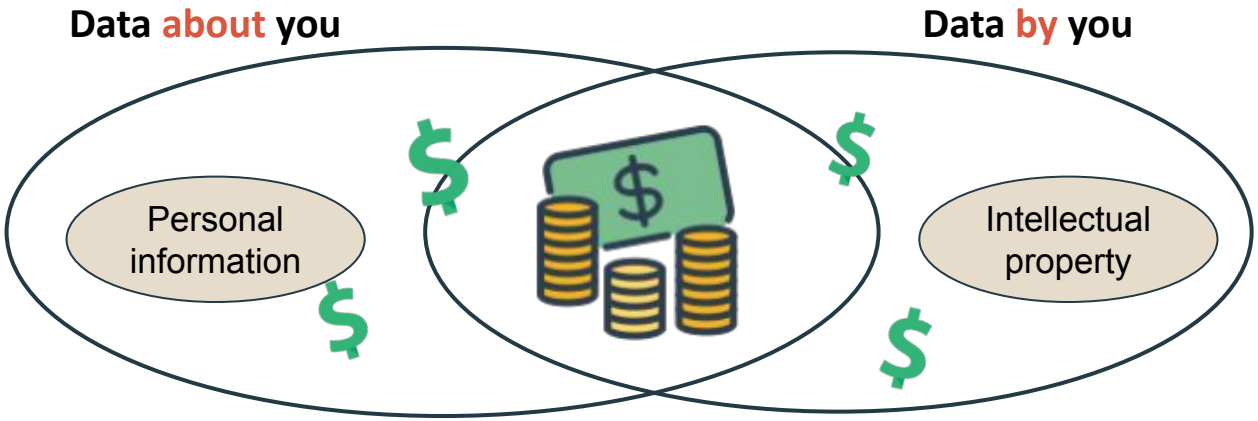
Session goals

- Expand definitions of user data.
- Motivate the need for better control of user data.
- Discuss technical and regulatory challenges exacerbated by machine learning applications.
- Enable cross-disciplinary conversation about potential solutions.

Introductions

1. Location (city/country)
2. Positions (grad student, undergraduate, professor, industry, etc)
3. Research interest/ Area of expertise

What is **your** data?



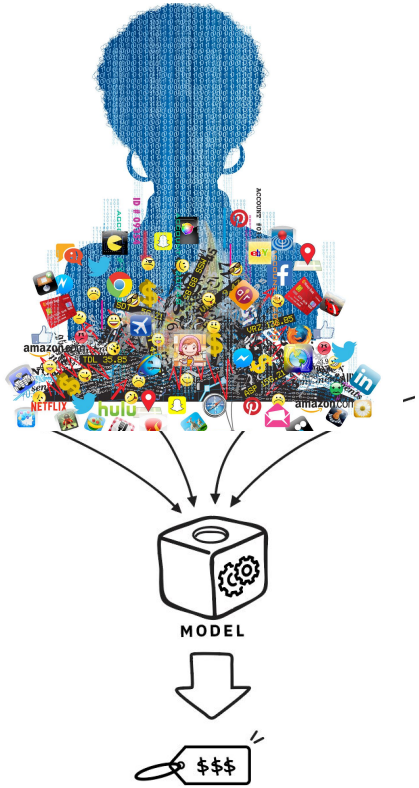
Discussion

Q: **Give examples of user data that can be used by companies and governments for monetization and surveillance.**

**please use the link in the zoom chat to enter your answers.*

Where does the value come from?

- From the ability to train/learn predictive models of behaviour.
- Once-and-for-all data transactions.



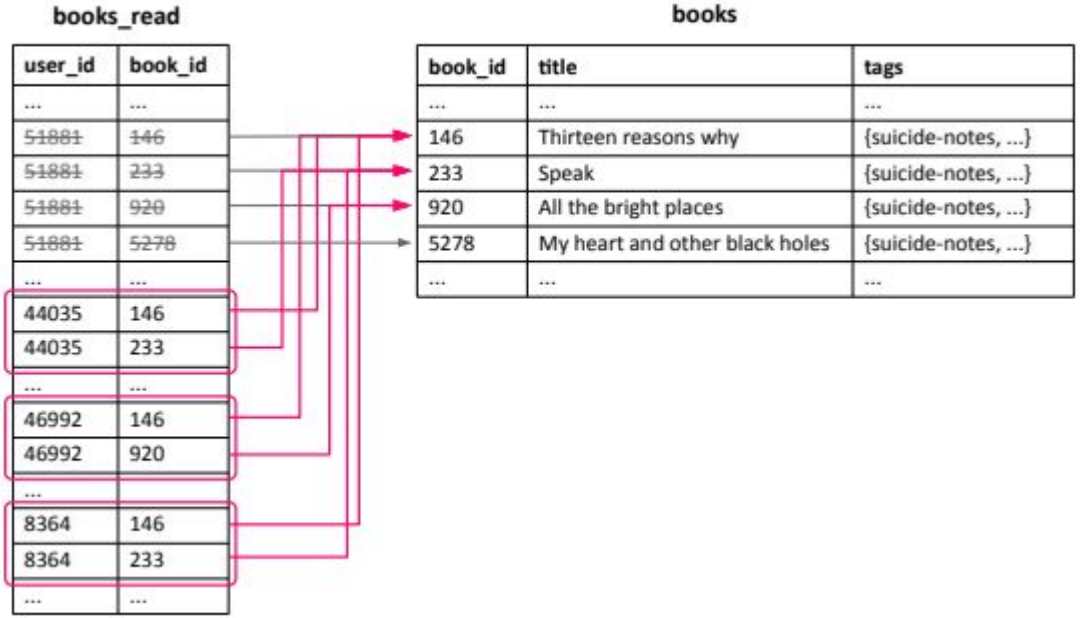
Deletion is NOT enough

books

book_id	title	tags
...	...	...
146	Thirteen reasons why	{suicide-notes, ...}
233	Speak	{suicide-notes, ...}
920	All the bright places	{suicide-notes, ...}
5278	My heart and other black holes	{suicide-notes, ...}
...	...	...

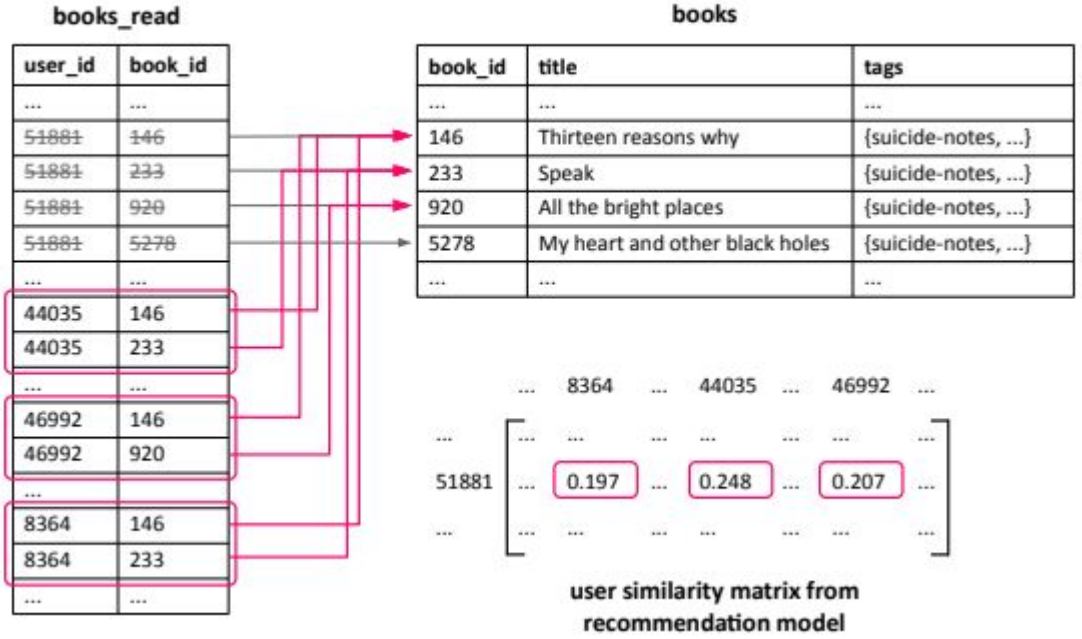
Schelter, S. *"Amnesia"-A Selection of Machine Learning Models That Can Forget User Data Very Fast.*

Deletion is NOT enough



Schelter, S. *"Amnesia"-A Selection of Machine Learning Models That Can Forget User Data Very Fast.*

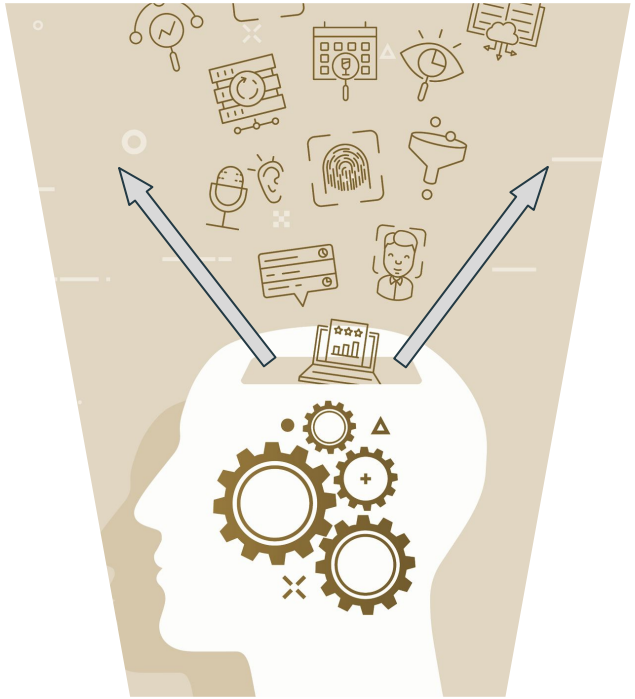
Deletion is NOT enough



Schelter, S. *"Amnesia"-A Selection of Machine Learning Models That Can Forget User Data Very Fast.*

We need “forgetting”

- Models need to scrub away information pertaining to a data deletion request.
- Models need to efficiently unlearn data.



Challenges to algorithmic forgetting

- Lack of incentives.
- Potential for unintended consequences.
- Technical and algorithmic challenges.

Discussion

***Q:* Discuss challenges to implementing flexible self-management of data.**

These may include examples of lack of regulatory frameworks, unintended consequence or technical challenges.

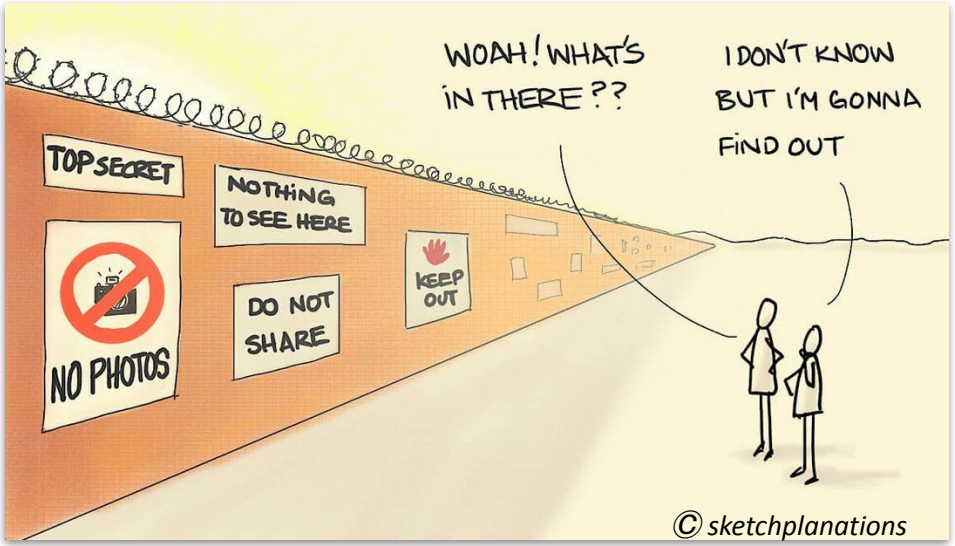
**please use the link in the zoom chat to enter your answers.*

Misaligned incentives

- Current regulatory frameworks do not mandate data to be deleted from aggregated statistics and trained & deployed ML model.
- Lack of audit and verification mechanisms.
- Data aggregations and data derivatives are often not considered personal information.
- Companies have little incentives to design algorithms that are more ‘deletion’ compliant at the cost of performance loss.

Unintended consequences: **Streisand Effect**

When trying to suppress something draws more attention to it



Algorithmic challenges

- No agreed upon technical definition of certifiable data removal.
- Model retraining upon data deletion request can be prohibitively expensive.
- Current data and model management lifecycles is not designed with flexible deletion in mind.

How to achieve Machine Unlearning

“Gold standard”

*Delete data point **{X}** from dataset **D** and retrain ML models from scratch using **D\{X}**.*

+ Simple and straightforward definition.
- Often practically infeasible due to high costs of model re-training.

How to achieve Machine Unlearning

“Statistical indistinguishability”

Modify an existing model slightly to ‘remove the influence of a point’

- Many competing definitions.
- Requires lots of assumptions to work.

How to achieve Machine Unlearning

“Data-deletion compliant algorithms”

Design new kind of algorithms with ‘forgetting’ in mind:

1. Federated Learning
2. Differentially Private Algorithms
3. Distributed Learning

Discussion

Q: **How to reconcile regulatory aspirations with implementation challenges to successfully guarantee individual agency in data ownership?**

Thank you

Questions?

Mihaela Curmei:
mcurmei@berkeley.edu
https://mcurmei627.github.io/