June 2021

Synthetic Faces to Improve Privacy and Fairness

The development of AI is creating new opportunities to improve the lives of people around the world. It also brings new ways to build fairness & privacy into these systems.

Interpreting images is hard

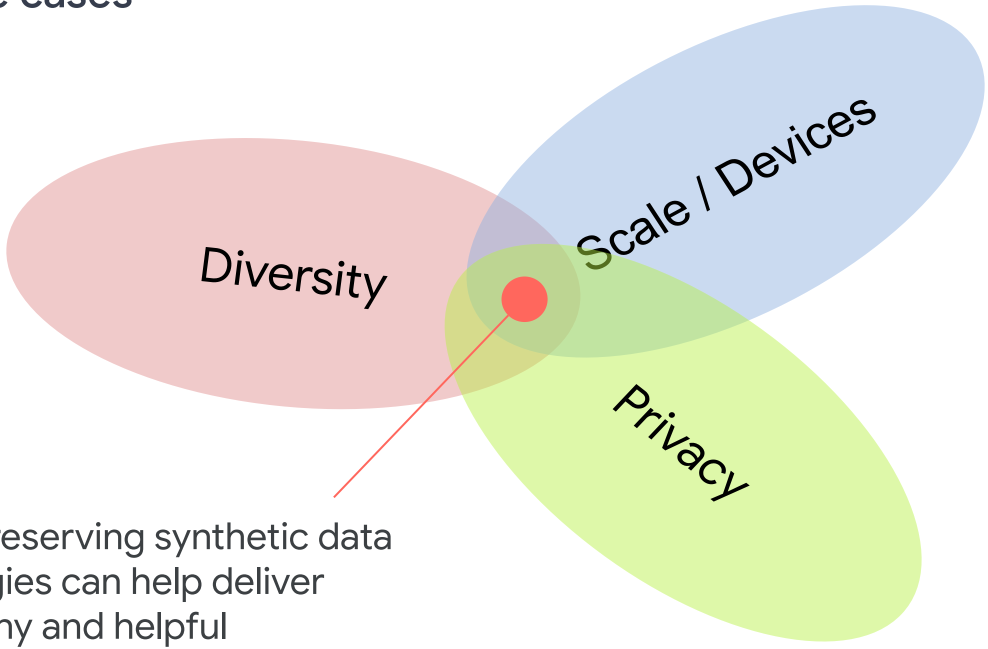
Semantic understanding of faces is key to many human-machine user experiences

- Face detection in cameras for auto-focus/exposure
- Organize personal Photos
- Attention detection for smart device interaction

Data needs / difficulties

- Large scale demographically diverse data
- Face data requires special policy/legal and privacy considerations

Synthetic data use cases



Privacy preserving synthetic data technologies can help deliver trustworthy and helpful experiences to more people with less real data

Synthetic data use cases

At Google, we are committed to the responsible development of face-related technologies. We were the first company to decide **not** to release a general-purpose facial recognition API, and are developing other face-related technologies in responsible, privacy-preserving ways that are aligned with our AI Principles.

Synthetic data can help us do that in a number of ways including:

- Ability to simulate a diverse population of users
- Amplify the effectiveness of personal data
    - Smaller datasets can be used to train generative models that can be used to generate larger amounts of diverse data
- Synthetic Counterfactual Fairness (modify hair/gender/age)

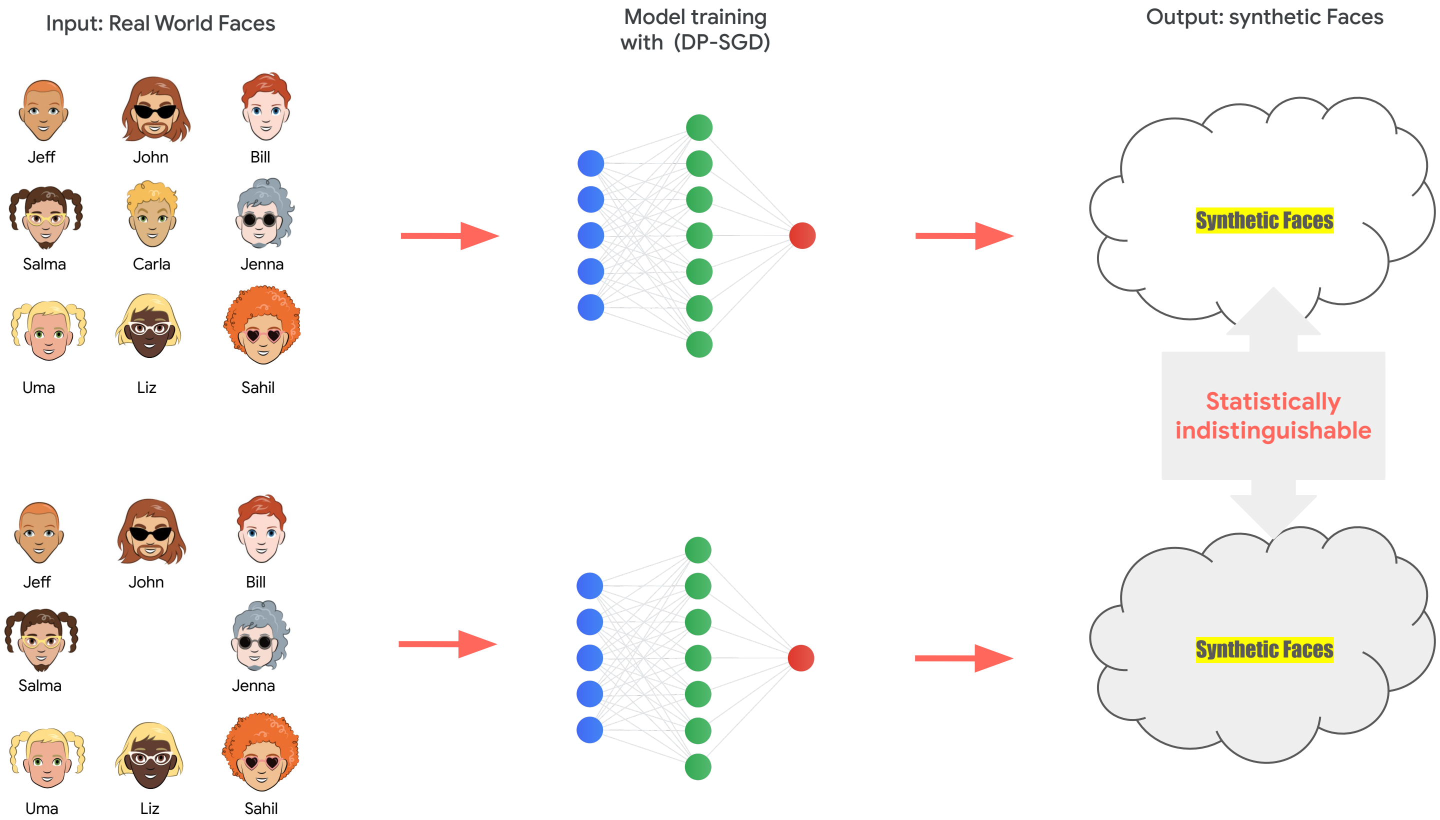
Current approaches for privacy preserving synthetic data

- Differential Privacy
  - Generative Models For Effective ML on Private, Decentralized Datasets (ICLR'20; arxiv)
    - text / image generation for debugging commonly occurring data issues
  - Pros
    - protection against data leakage (e.g., in deep learning models)
    - anonymisation (with formal guarantees)
  - Cons
    - may sacrifice utility of the synthetic data
- K-anonymity
  - is a requirement from data releases that the (quasi) identifying data of every person in the release should be identical to that of at least (k − 1) other individuals in the release
  - Cons
    - prone to linkage/background knowledge attacks

Related scholarly work

- Differential Privacy is emerging as a viable path for privacy preserving machine learning
  - Provides formal privacy guarantees, robust against re-identification (Abadi et al., 2016, Dwork et al., 2006)
- Algorithmic and legal research relating Differential Privacy with anonymization (Towards formalizing the GDPR's notion of singling out, Cohen et al., 2020)
- Differential Privacy could charter a path for algorithmic anonymization of facial imagery (e.g., Zhang et al., 2018)

How: Face Generation with Differential Privacy



Differential Privacy Guarantees

- **robust** against membership inference attack*
  - an attacker can NOT tell whether an individual face was in the training dataset or not regardless of the attack algorithm
- **robust** against memorization**
  - model can NOT output one or more input faces verbatim
- **no person-specific information** in model & synthetic data generated from it
  - model only contains aggregated, anonymized information

* arXiv:1812.02274 Differentially Private Data Generative Models
** arXiv:1911.06679v2 Generative Models for Effective ML on Private, Decentralized Datasets

Summary

- The imperative for an algorithmic way for facial imagery anonymisation
  - to protect user privacy
  - to enable building great products that rely on large scale diverse data
- Differential Privacy algorithms, when applied to, e.g., categorical data, are sufficient to render the synthetic data anonymous
- What are the implications of generating something that resembles a "real" face by chance?
  - Is this similar to randomly generating other personally identifiable information, e.g., 10 digits that could be a person's phone number?
  - Is there a difference between generated data being similar to training data or the whole domain?