* Ziel
  + Review von bestehenden ML Models welche synthetische Daten generieren
* Vorteile synthethische Daten
  + Gut wenn Daten Privat gehalten werden müssen wegen Datenschutz
  + Wenn es mehr Daten braucht
* Anwendung
  + Vision
  + Voice
  + NLP
  + Healthcare
    - medGAN produziert synthetische patient records
    - Institute können ihre Datenbestände so teilen
  + Business
    - Auch Teilen von Privacy Daten
  + Education
    - Auch Teilen von Privacy Daten
  + Location and Trajectory Generation
* Privacy Risks
  + Wenn genügend von gleichen Trainingsdaten gesampled wird, besteht Chance dass diese Trainingsdaten wieder auftauchen
  + Es gibt gewisse Lösungen, jedoch keine Garantie
    - K-anonymity
    - Nearest marginal
* Fairness
  + Synthetische Daten übernehmen Bias von den Trainingsdaten, aber auch Bias von Preprocessing, collection und Algorithmen
  + Es gibt models welche versuchen Bias in Trainingsdaten zu erkennen
* Trustworthiness
  + trustworthiness of such data depends on its representativeness, privacy preservation, and absence of bias
* Evaluation
  + Human evaluation
  + Statistical difference evaluation
    - Statistische Metriken auf originalen und synthetischen Daten
  + Evaluation using pretrained moel
    - Ähnlich wie bei GAN
  + Training on synthetische Datasets und testing auf real dataset (TSTR)
  + Application specific evaluation