PRIVACY-AWARE SPEECH DATA SHARING

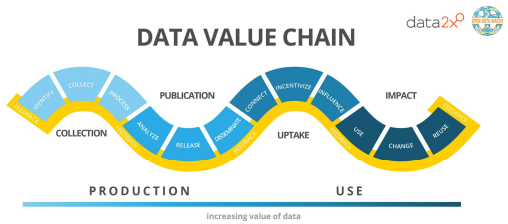
PROF.DR.IR. THIJS VEUGEN

CONTENTS

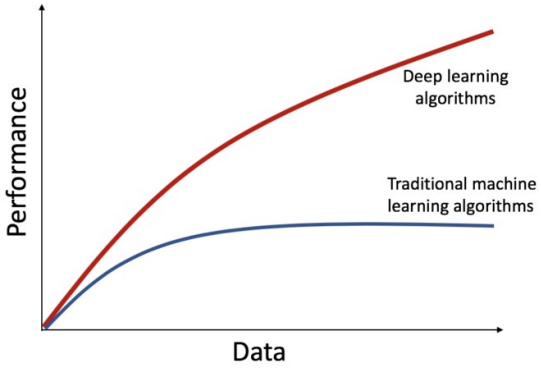
TNO innovation for life

AI MODELS
POSSIBILITIES AND CHALLENGES

- Huge potential in AI solutions



- AI models, especially deep learning models, are very data-hungry



- Data is often available, but cannot be shared and combined across organizations or even departments due to privacy and IP concerns

TNO innovation for life

SPEECH DATA SHARING
POSSIBILITIES AND CHALLENGES

- Speech + its transcriptions can be used to learn powerful Automatic Speech Recognition (ASR) systems
- Different data sources to avoid bias based on accent, dialect, gender, age, etc.
- Example application: automated help desk / call centre
- However, speech data has different types of sensitive features that make sharing across silos or organizations challenging:

Who, where, when - metadata	How - Acoustic characteristics speakers	What (Content) - Intellectual property
• Date/time • Location • Names of speakers • …	• Gender • Age • Native language • Emotional state • Health status (dementia, Parkinsons, covid-19…) • Deception • …	• Content of television show • Speech/talk/class • Discussion of medical information • Police hearing • Court case • …

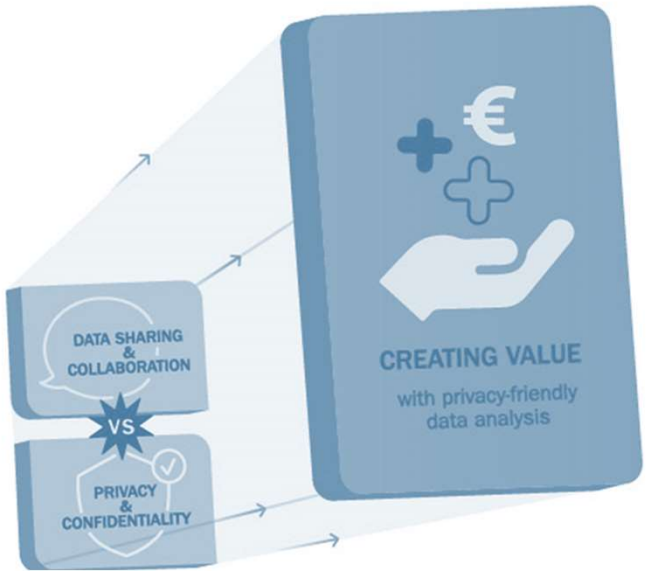
WHAT ARE WE PROTECTING?

1. Speech input data for training the model:
   Various data sources that cannot be shared, speech data is sensitive personal data
2. Trained ASR model:
   Could reveal input data, can be commercially sensitive.
3. Speech data to be transformed to text through trained ASR model:
   Speech data is sensitive personal data, and model can be sensitive
   a) Speech data remains at client and server has ASR model, how to recognise the speech without interchanging data?
      (Privacy-Preserving Machine Learning for Speech Processing, Manas A. Pathak, PhD thesis, 2014)
   b) Speech data is (securely) stored at server.
4. Text output from the ASR model:
   Text output is personal data (How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing, Samuel Sousa, Roman Kern, Artificial Intelligence Review, 2022)

We focus on the training of the ASR model: 1 and 2.

SOLUTIONS FOR SAFE SPEECH DATA SHARING
PRIVACY ENHANCING TECHNOLOGIES

- Privacy Enhancing Technologies (PET) offer ways to efficiently and safely learn models on distributed datasets
- TNO has broad expertise in secure multi-party computation (MPC), Federated Learning (FL), and synthetic data generation (SDG)
- Speech data comes with its own characteristics – which PET to use?



SEVERAL OPTIONS

Protect speech data	Protect ASR model
• Input data transformation • Federated Learning • Homomorphic Encryption	• Differential Privacy • Trusted Execution Environment

PROTECT SPEECH DATA FOR TRAINING THE MODEL
INPUT DATA TRANSFORMATION

Before training, transform speech data such that it becomes less sensitive:
(Understanding the Tradeoffs in Client-side Privacy for Downstream Speech Tasks, Proceedings, APSIPA Annual Summit and Conference 2021)

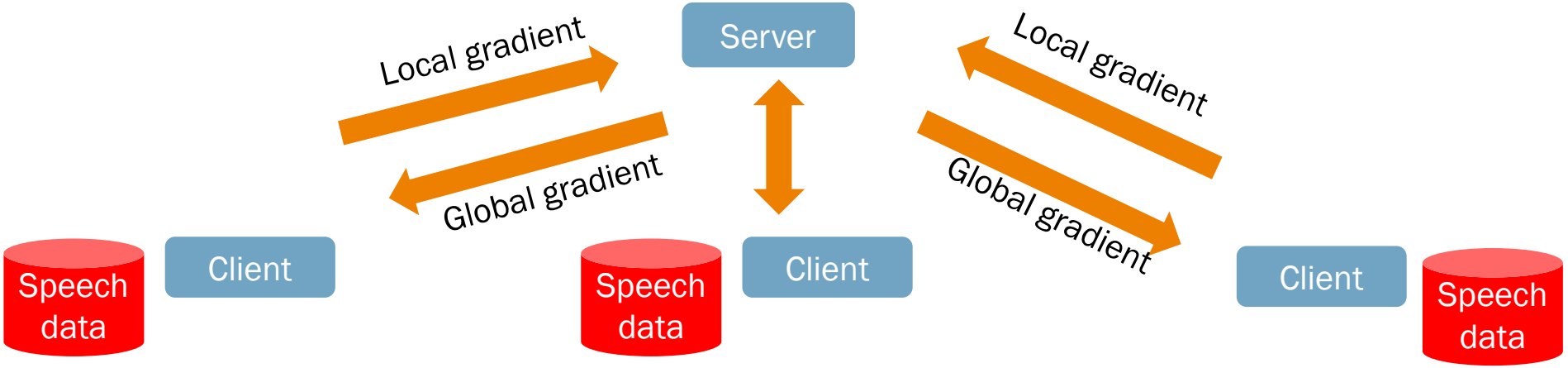
- Client-side transforms:
  signal processing techniques
- Local differential privacy:
  Add (a little) noise to speech data

easy to perform

still sensitive data, probably combine with other privacy-preserving training techniques

PROTECT SPEECH DATA FOR TRAINING THE MODEL
FEDERATED LEARNING

- A server trains the model through iterative updates from clients



scalable solution

(Y. Gao et al., "End-to-End Speech Recognition from Federated Acoustic Models," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 7227-7231, doi: 10.1109/ICASSP43922.2022.9747161)

gradients leak information on speech data

(B. Hitaj, G. Ateniese, and F. Perez-Cruz. Deep models under the GAN: Information leakage from collaborative deep learning. In ACM CCS, 2017)

PROTECT SPEECH DATA FOR TRAINING THE MODEL
HOMOMORPHIC ENCRYPTION

- Local gradients are sent encrypted to the server
- The server combines them to an encrypted global gradient without decryption
- The clients compute the new local encrypted gradient without decryption

no leakage of gradient information

computations with encryptions require more computational effort

Sav, S., Pyrgelis, A., Troncoso-Pastoriza, J., Froelicher, D., Bossuat, J., Sousa, J.S., and Hubaux, J. POSEIDON: Privacy-Preserving Federated Neural Network Learning, 2020, https://arxiv.org/abs/2009.00349

PROTECT THE TRAINED MODEL
DIFFERENTIAL PRIVACY

- Global differential privacy:
  Add (a little) noise to the (encrypted) model
- Mini-batch stochastic gradient descent with differential privacy:
  Add (a little) noise to global gradients (M. Abadi et al., “Deep learning with differential privacy,” in Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, 2016, pp. 308–318.)

(a little) protection of the trained model

risk of accuracy loss, needs to be combined with either homomorphic encryption or federated learning

PROTECT THE TRAINED MODEL
TRUSTED EXECUTION ENVIRONMENT

- Use sophisticated hardware to store the model (Trusted execution environment)
- Secure computations during ASR (Trusted Computing)

no leakage of trained model (and input data from that)

need sophisticated hardware

CONCLUSION

- Speech data is sensitive personal information
- We need a PET (federated learning, homomorphic encryption, MPC) for training distributed speech data. Trade-off: security (HE) vs. Scalability (FL)
- If the model is sensitive, we need additional PETs for protecting it
- If the recognised text is sensitive, we need additional PETS for protecting it


BEDANKT VOOR
UW AANDACHT
TNO innovation for life