Is federated learning (FL) safe?

- If an ML model is useful, it must reveal information about the data on which it was trained [1].
- Training data can be reversely inferred from the model [2].
- In FL, gradients and model parameters leak users' data [1, 3].

References

1. Melis et al. Exploiting unintended feature leakage in collaborative learning. In *IEEE Symposium on Security & Privacy*, 2019.
2. Fredrikson et al. Model inversion attacks that exploit confidence information and basic countermeasures. In *CCS*, 2015.
3. Hitaj et al. Deep models under the GAN: information leakage from collaborative deep learning. In *ACM SIGSAC Conference on Computer and Communications Security*, 2017.