PRIVACY & MACHINE LEARNING

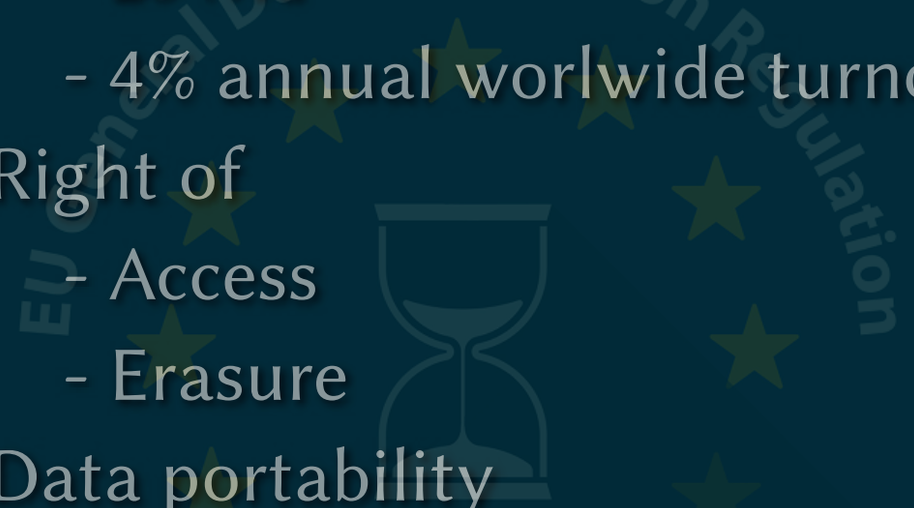
FRANÇOIS SAUSSET

Thales data science lab

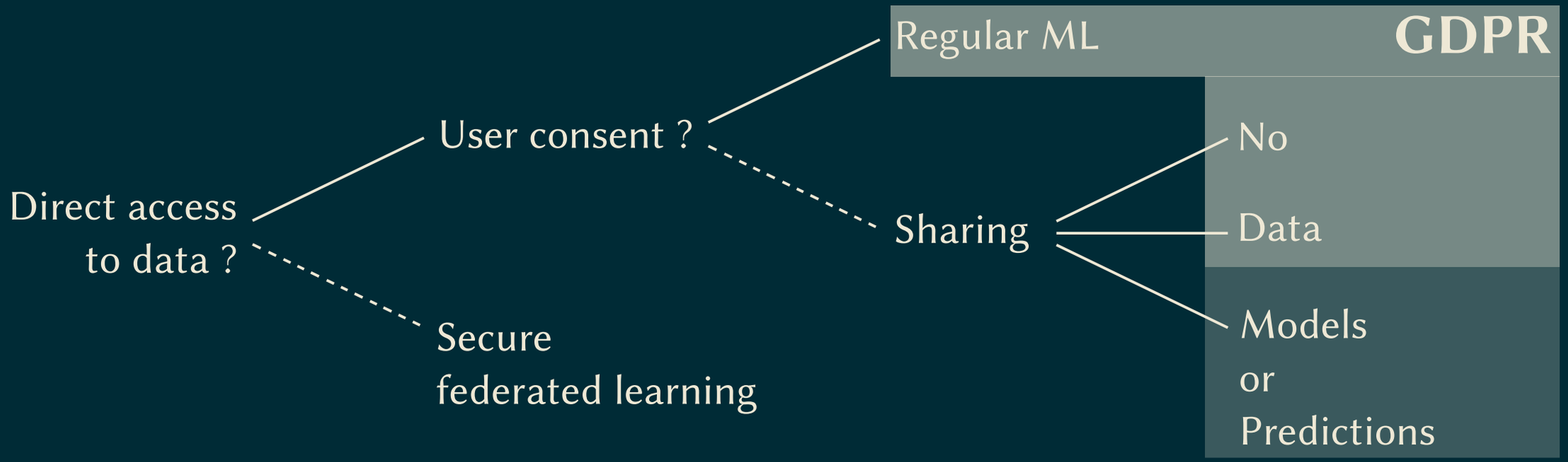
THALES


EU General Data Protection Regulation

- Apply to all EU residents data
- Sanctions up to (max of)
    - 20 M€
    - 4% annual worlwide turnover
- Right of
    - Access
    - Erasure
- Data portability
- Data protection
    - Encryption
    - Tokenization
    - **Anonymization**



MACHINE LEARNING ON PERSONAL DATA



DATA ANONYMIZATION

- **Anonymization ≠ Tokenization**
- **Remove information** to prevent people identification by
  - Isolation
  - Correlation between different databases
  - Using additional external information
- **Keep information** useful for the targeted analysis

$\Longrightarrow$ TRADE-OFF BETWEEN PRIVACY AND UTILITY

DATA ANONYMIZATION

- Latent risk of reidentification

  - Quantify it by attacking yourself !

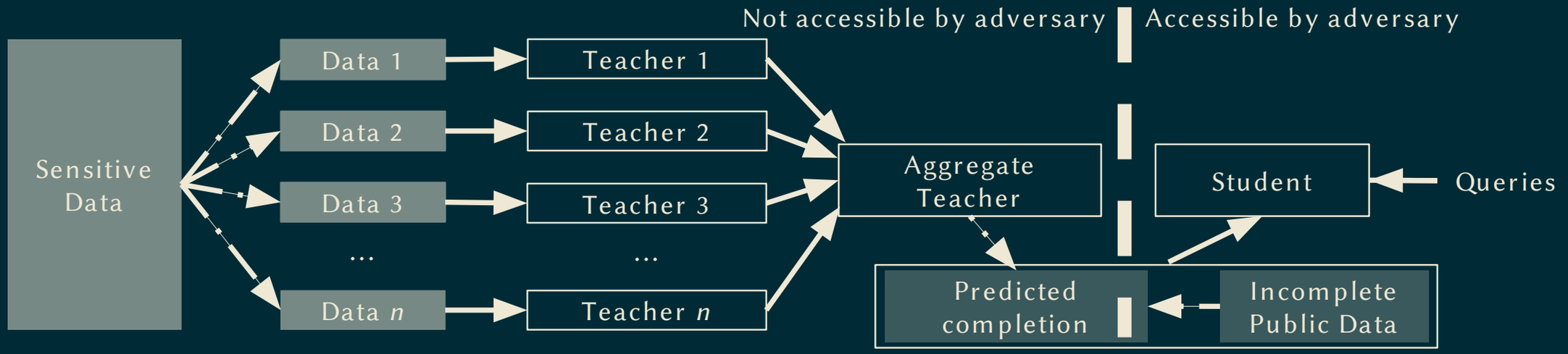
- Main methods

  - Gather people & create typical persona
  - Keep people, but disturb information
  - Generate relevant fake data

  (GANs: arXiv:1803.03148,...)

- **Anonymization is use-case specific !**

PRIVATE LEARNING



- Split ML model (arXiv:1610.05755, arXiv:1802.08908)
- Fair AI (arXiv:1712.08197)
- New dropout for Deep Learning (arXiv:1712.02629)
- New SGD for Deep Learning (arXiv:1712.09097)

SECURE FEDERATED LEARNING

OPENMINED.ORG

Deep learning

+

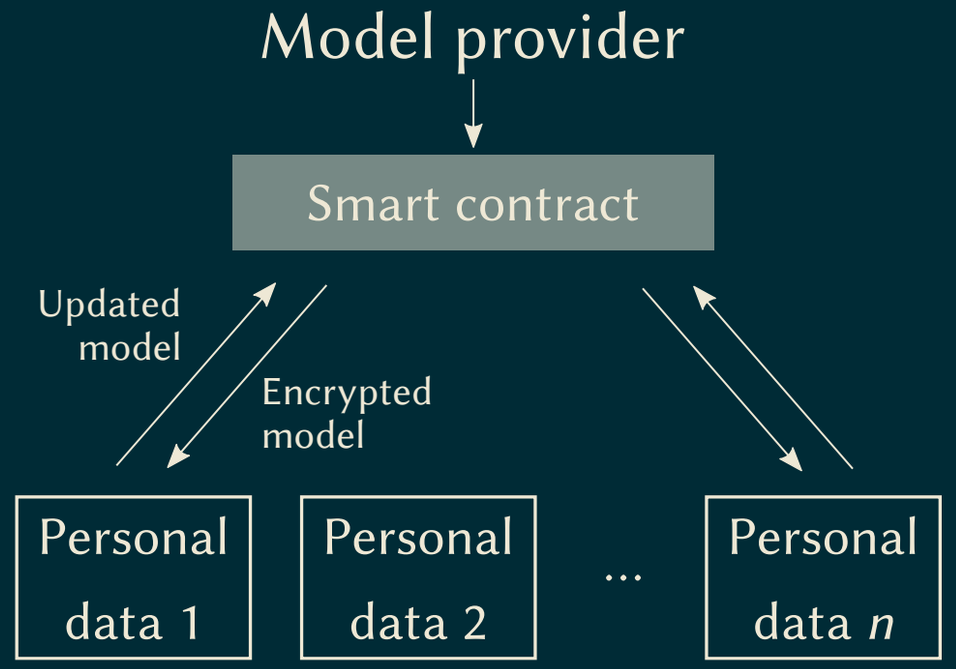
Federated learning

+

Homomorphic encryption

+

Smart contracts (Blockchain)



TAKEAWAYS

- **Not as simple as using tokens or encryption !**
- New field of research
- Growing rapidly
- Many ways to ensure privacy when doing ML
    - Data transformation
    - ML model architecture
    - ML distribution/learning architecture