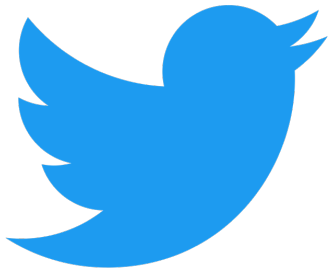
Privacy-Preserving and Distributed Machine Learning

Cláudia Brito

February 2023

claudia.v.brito@inesctec.pt

Data is Everywhere
Everything is Data



Do you accept the terms and conditions?

When you accept all the Terms and conditions:

Me who didn't read it:

The company who doesn't want you to read it:

But what about **my privacy**?

What is private data?

Private Data:

- Name
- Address
- Financial Status
- Health Records

Big Collectors:

- Hospitals
- Banks
- Finances
- Biocenters

Other Collectors:

- Browsers
- Social Networks
- Games

Do you allow us to collect your data?


NO

So, how about we do all of this in a private way?

The real problem?

Large amounts of sensitive data generated.

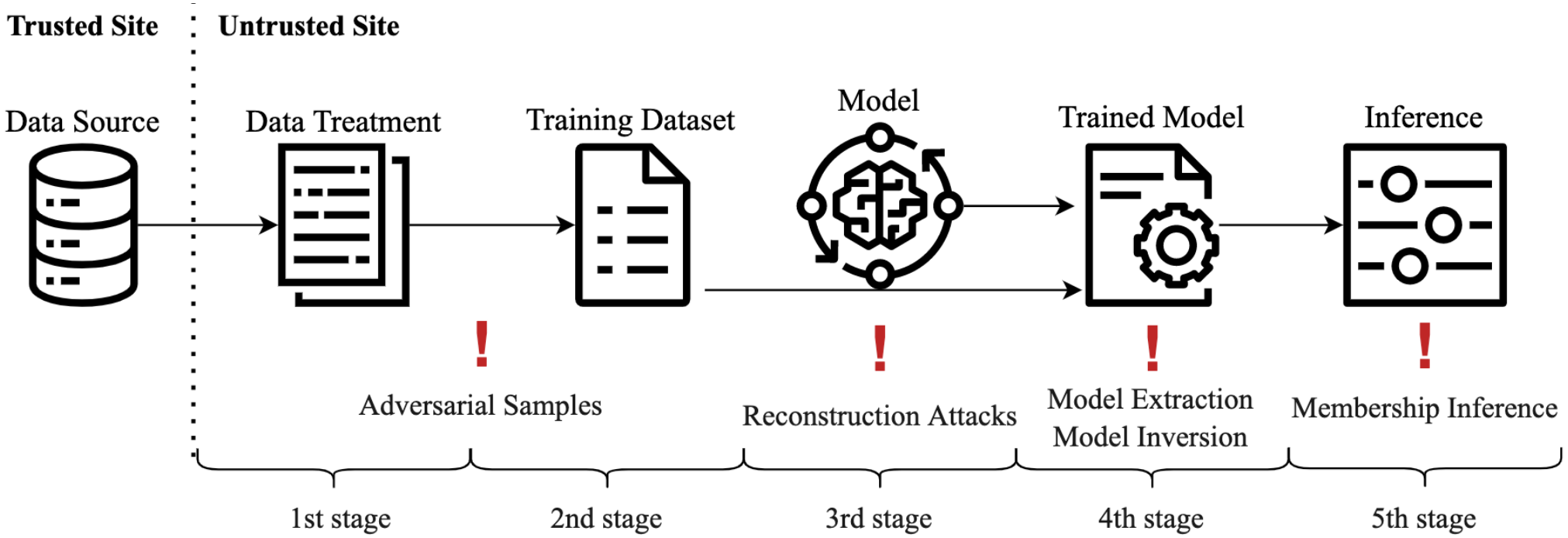
Use of AI techniques to extract valuable insights.

Regulations to avoid the misuse of sensitive data.

Why does ML comes in the picture?

"A computer program is said to learn from ***experience E*** *with respect to some* ***task T*** *and some performance* ***measure P****, if its performance on T, as measured by P, improves with experience E."*

The ML Workflow



New Limitations and Challenges

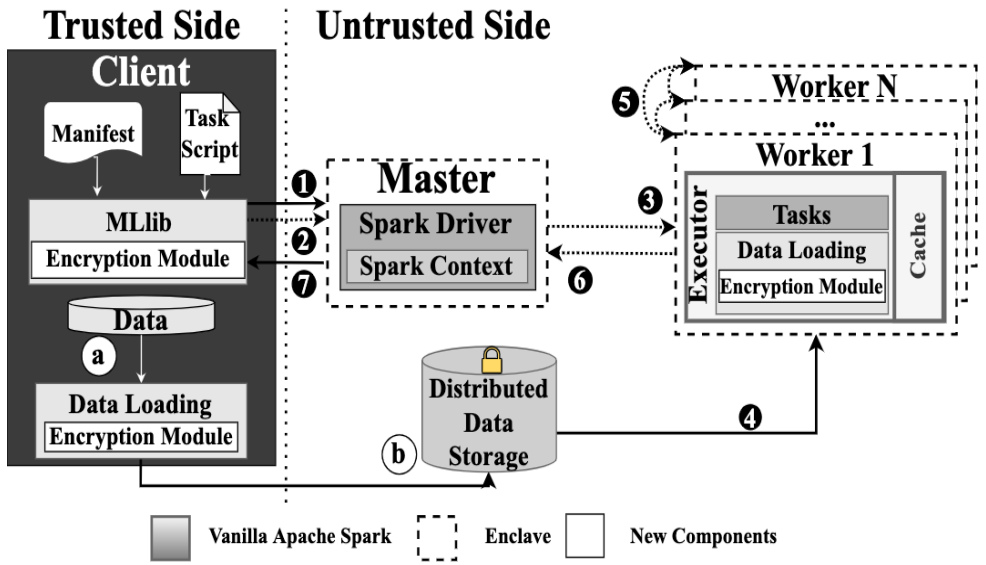
- ML datasets and models are stored and processed in plaintext.
- Large amounts of data to be processed and heavy computation.
- Common cryptographic schemes impose impractical overheads.
- Lack of local infrastructures.

Solutions

1. Outsource computation to third-party infrastructures.
2. Provide secure environments.
3. Allow the private collaboration between all the entities.

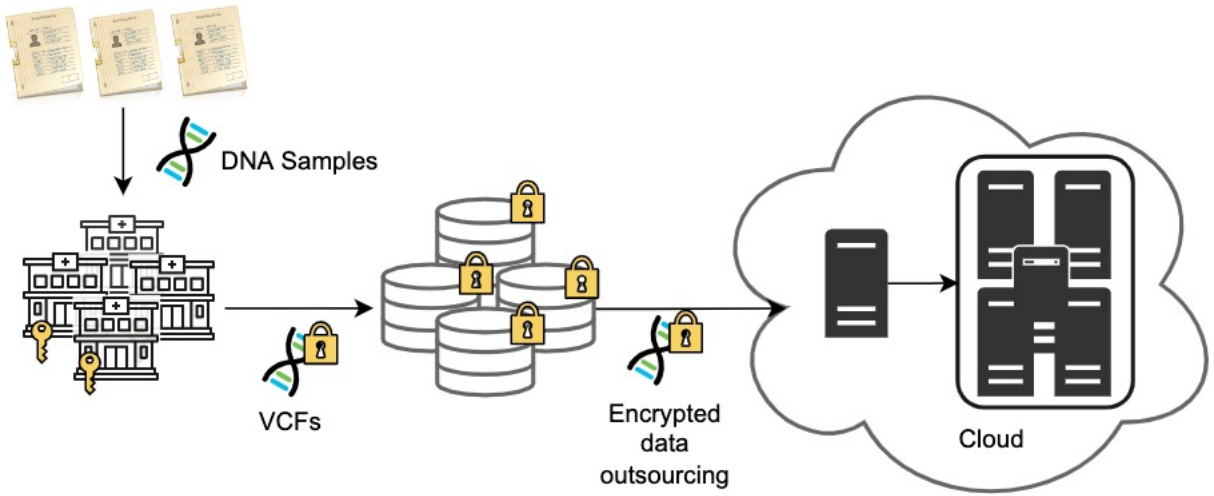
Soteria: Preserving Privacy in Distributed Machine Learning (SAC'23)

- Allows the outsourcing of computation.
- Provides secure environments for the outsourced computation.
- Redefines which computation needs to be kept private.



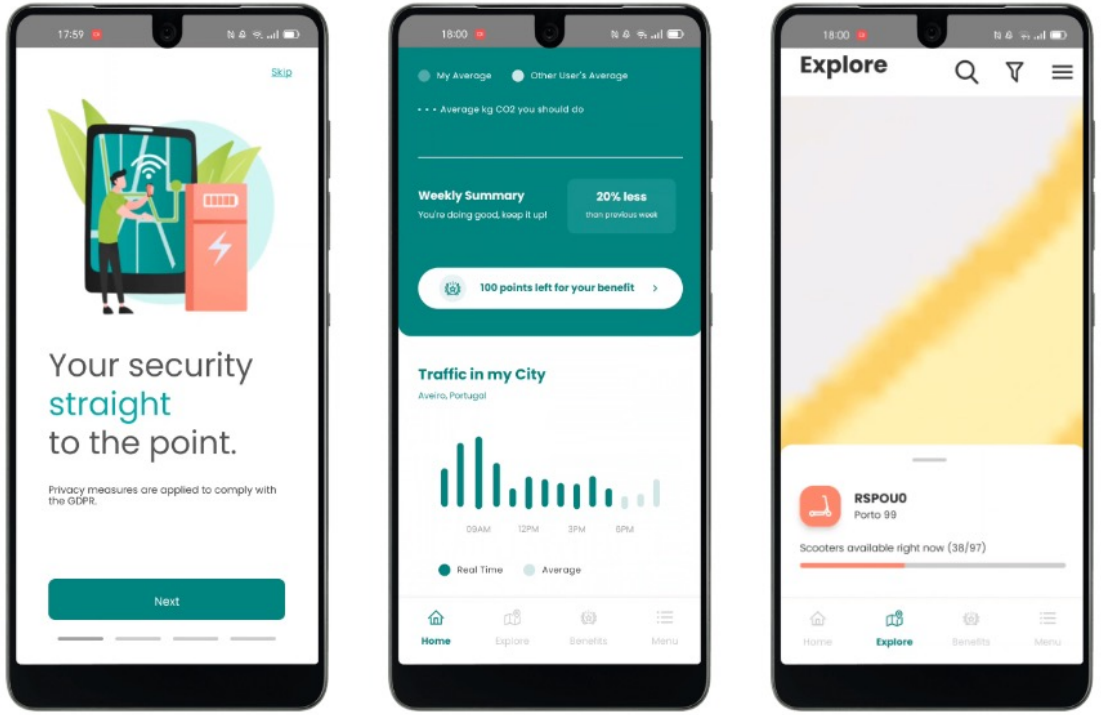
Work in Progress: Privacy-Preserving Framework for Genomic-Wide Association Studies

- Allows the outsourcing of genomic computation.
- Provides secure distributed environments for the computation of the genome.
- Promotes the collaboration between entities.



Promoting sustainable and personalised travel behaviours while preserving data privacy (TRA2022)

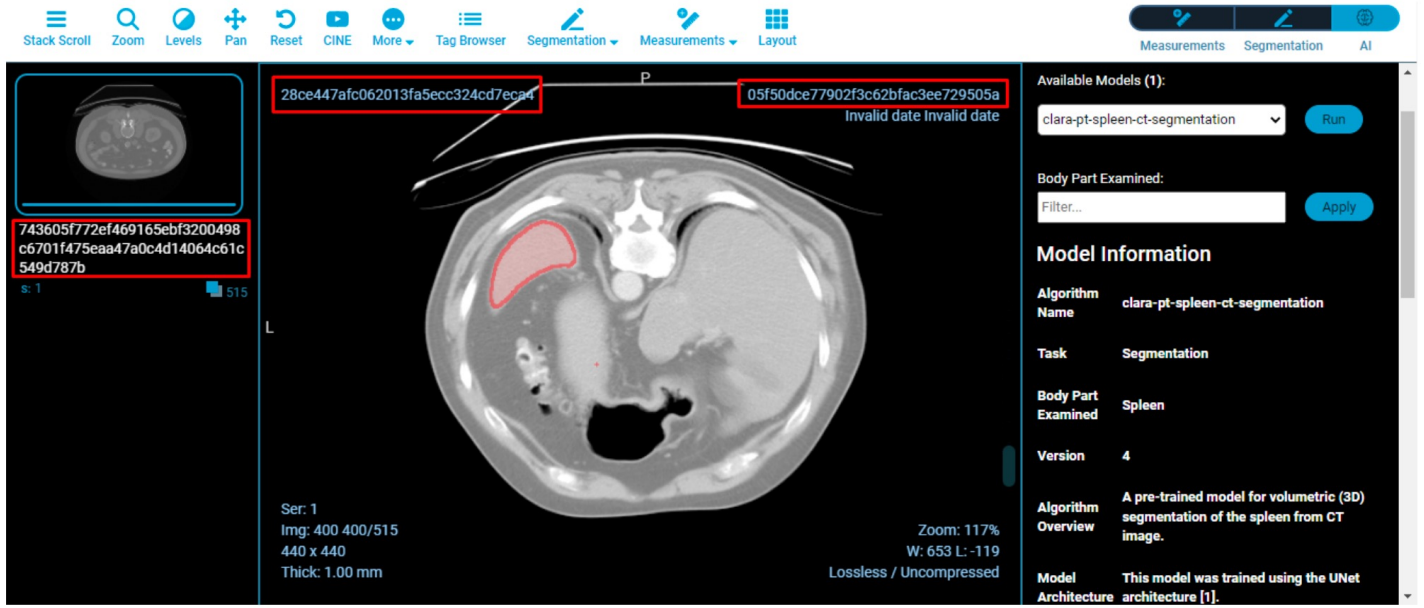
- ✔ **Assessing GHG emissions** from transport modes
- ✔ **Improving data experience** with real-time access
- ✔ **Privacy-preserving data** by not uploading it to centralised servers (leveraged by **AI approach**)
- ✔ **Local incentives to reduce GHGs**



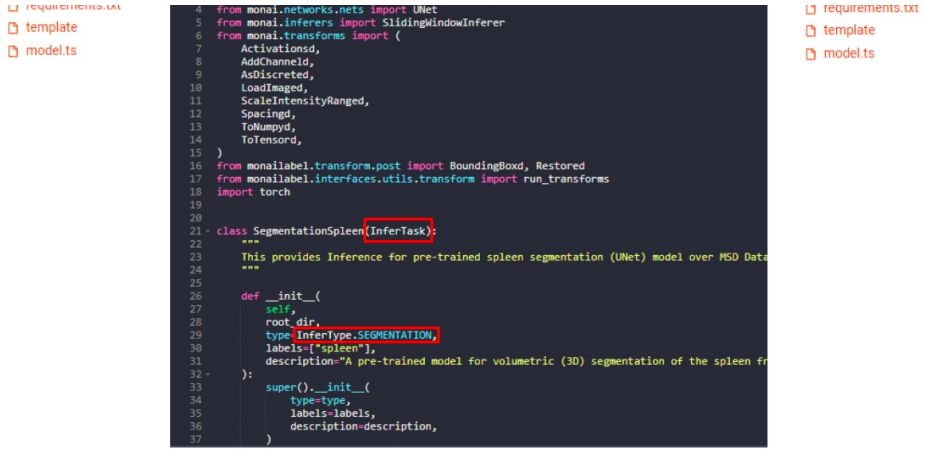
AI4CITIES franchet.ai

Other Research Works

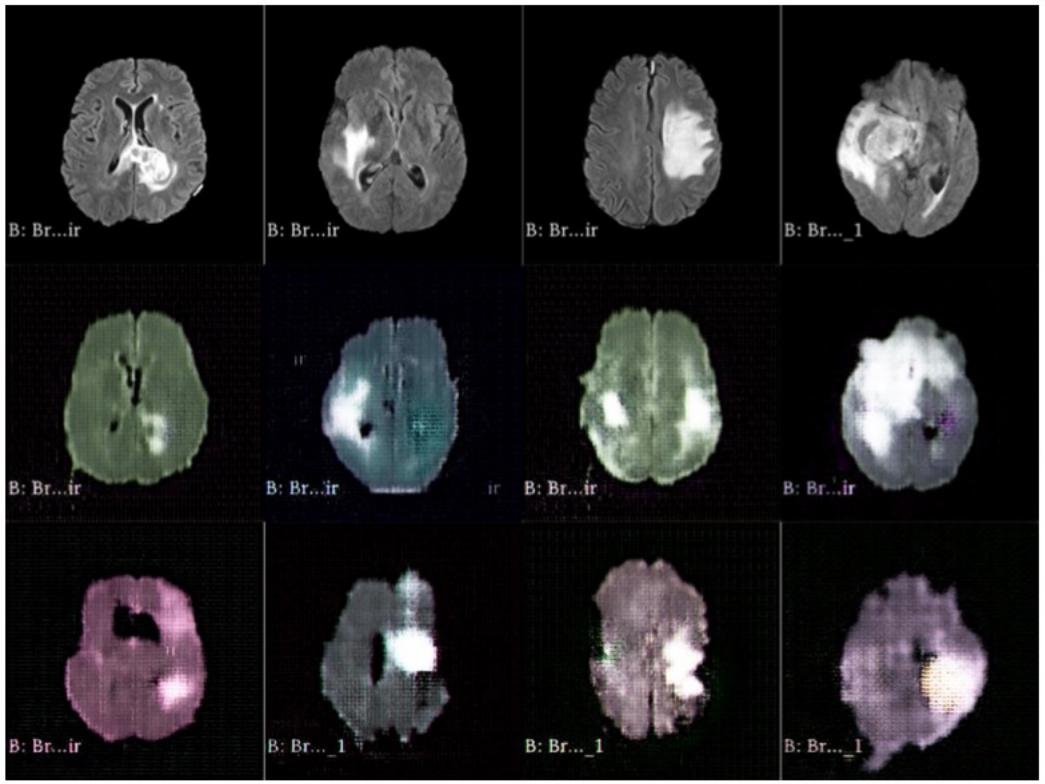
MCC - MedCloudCare



Medical platform for visualizing and analyzing DICOM images by resorting to machine learning algorithms.

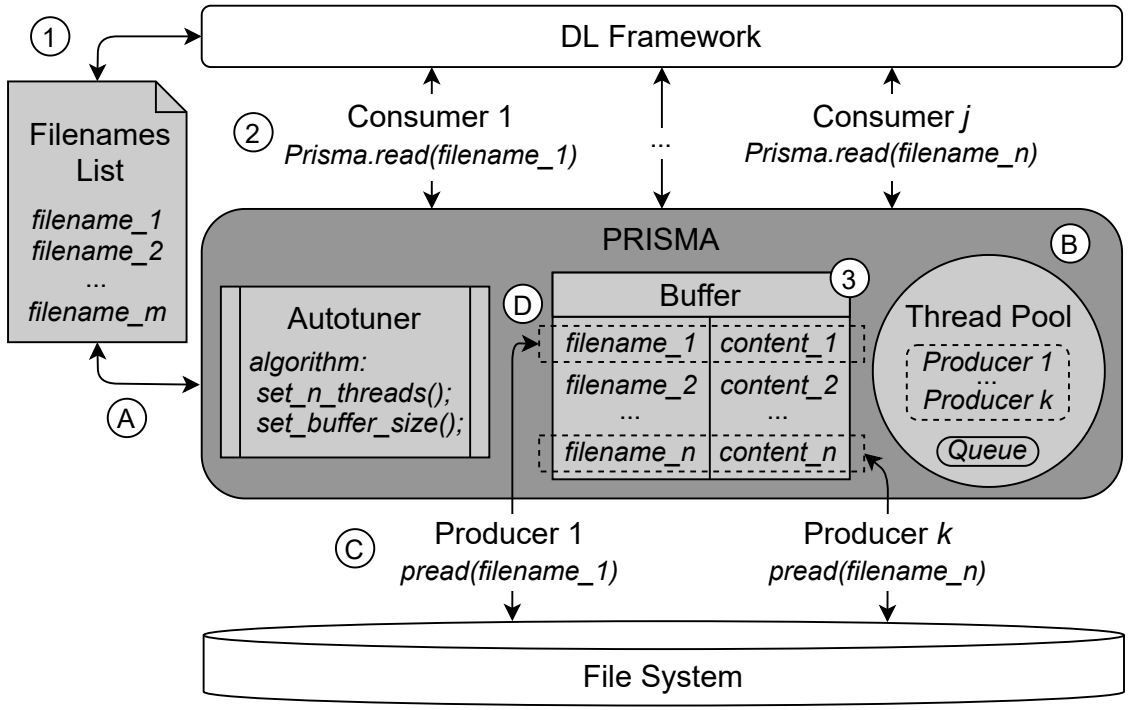


Generation of Medical Images with Deep Learning



Generation of medical images with high resolution to increase the medical data used for developing new models of diagnosis.

PRISMA



Storage optimizations for Deep Learning Workloads.

Questions?

If you have any questions send an e-mail to:
claudia.v.brito@inesctec.pt