

This work was supported in part by the Spanish Ministry of Economy and Competitiveness under TestEAMoS (TIN2016-76956-C3-1-R) and POLOLAS (TIN 2016-76956- C3-2-R) projects, and ERDF funds. We want to thank the collaboration of INEBIR (inebir.com) and G7 Innovation (g7innovation.com) in the development of this scientific project



Test-Driven Anonymization in Health Data: A Case Study on Assistive Reproduction

Cristian Augusto, Miguel Olivero, Jesús Morán, Leticia Morales, Claudio de la Riva, Javier Aroba and Javier Tuya

GIIS and IWT2 Research Group

<http://giis.uniovi.es> - <http://iwt2.org/>

University of Oviedo - University of Seville – University of Huelva



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo



Context (I)



Privacy



Usefulness



Sharing



Privacy



Usefulness



Sharing



Privacy

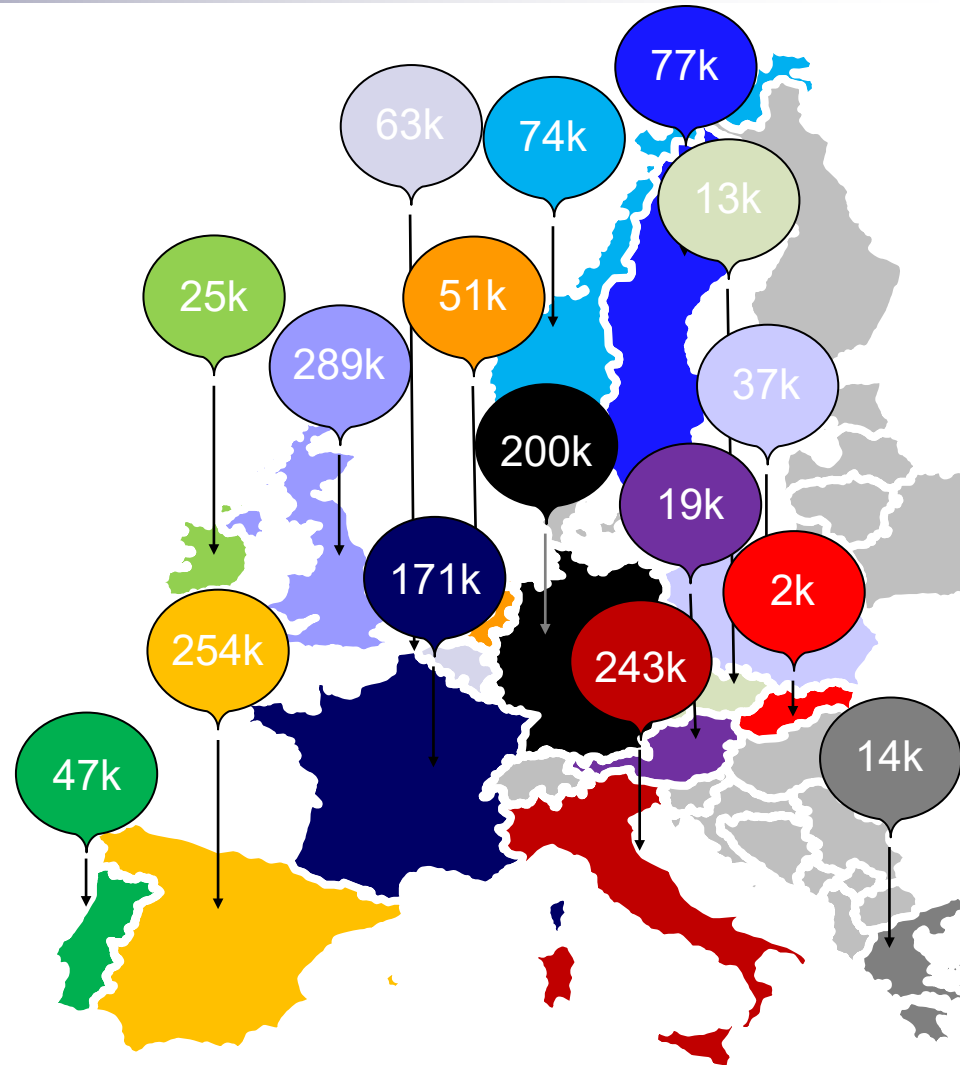


Usefulness

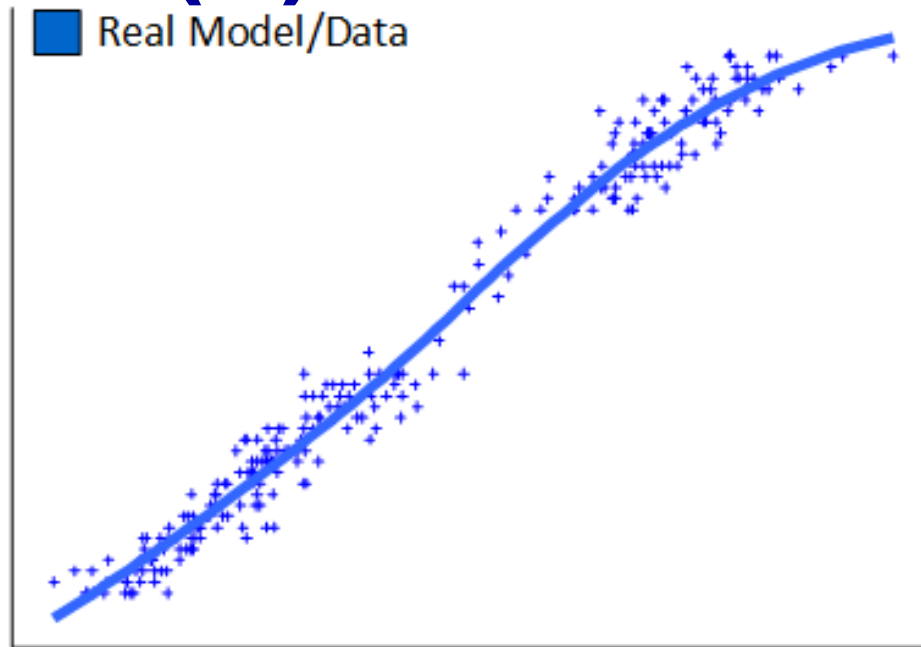


Sharing

- Governments want to share data to help researchers → data contains sensitive information about citizens.



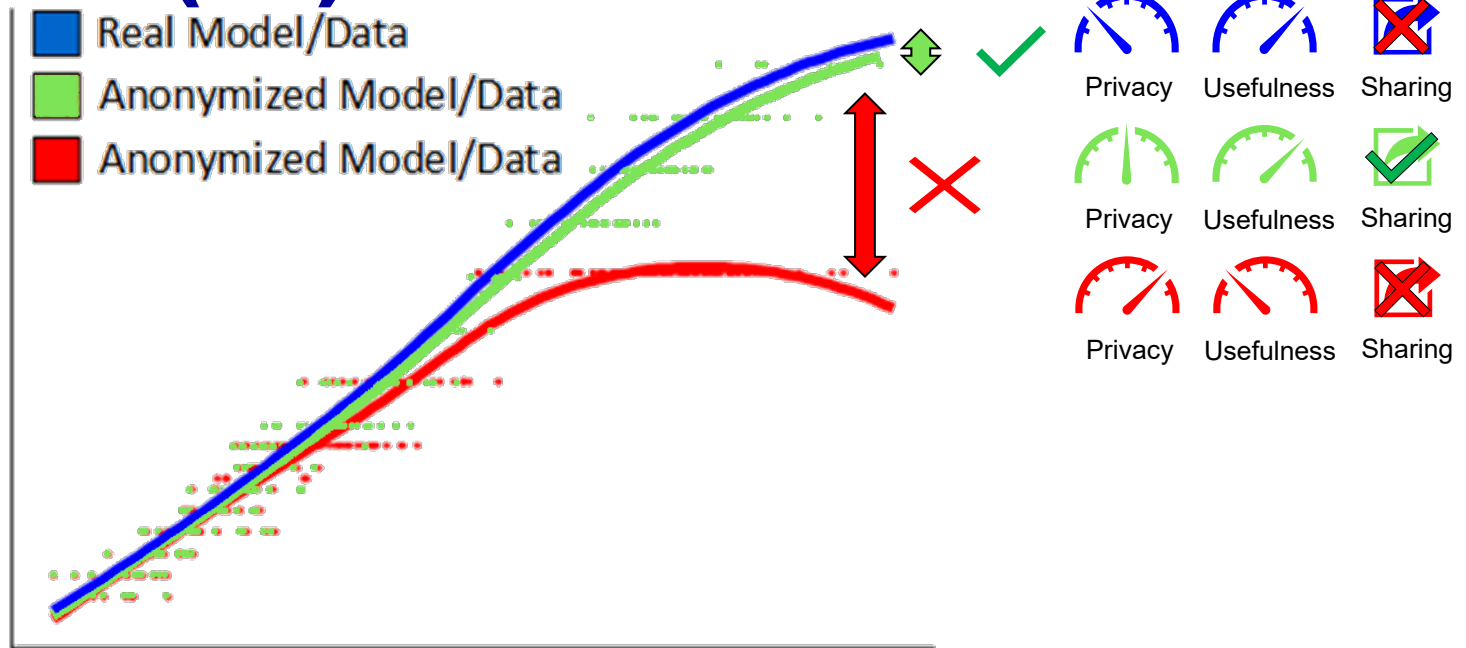
Context (II)



■ Anonymization effect

- Improve individual's privacy also loss of information
- Alters data and affects developments highly dependent of the data (i.e. Artificial Intelligence models)

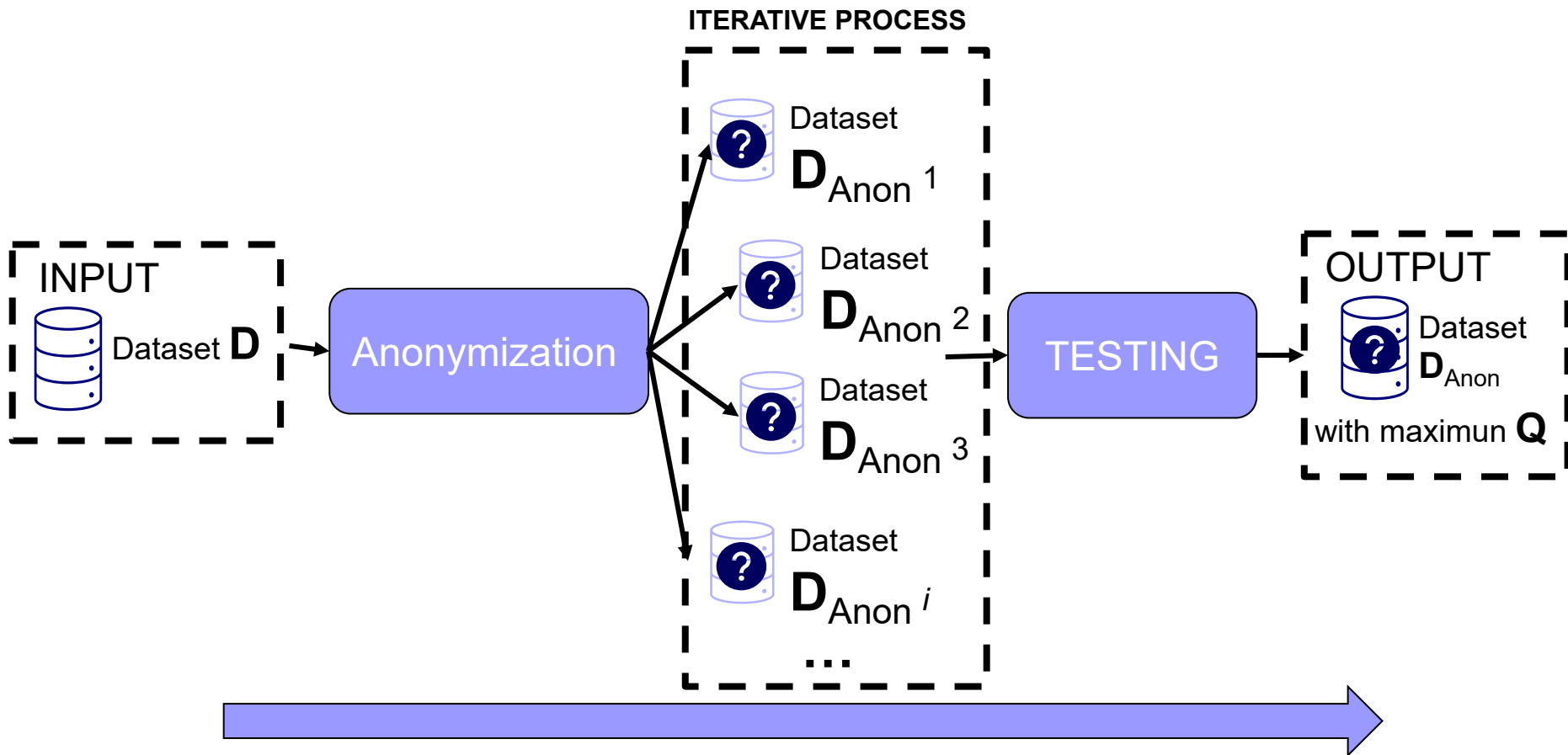
Context (III)



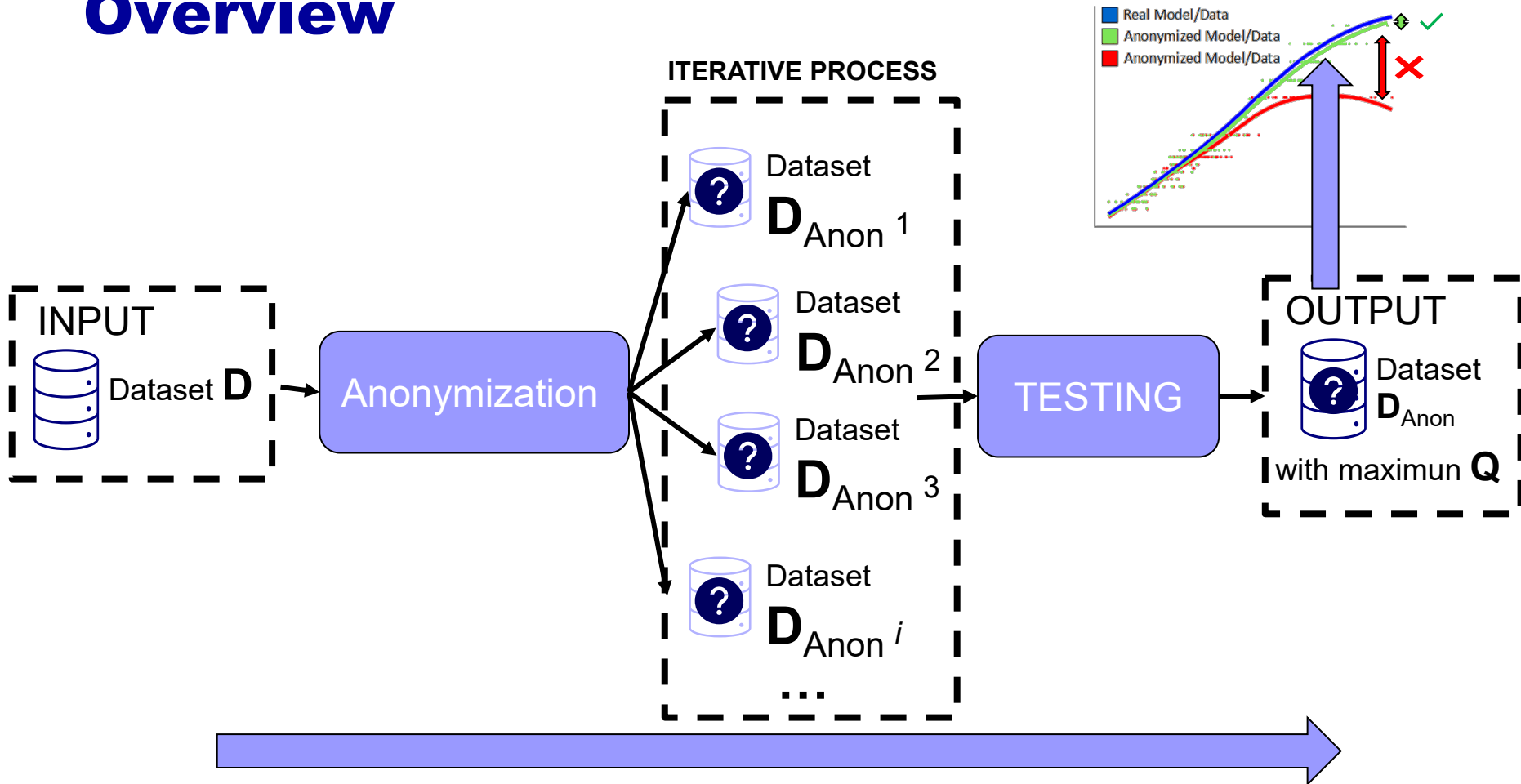
■ Fault Masking

- May be the case that an AI Model that works well in the development stage fails dramatically during production

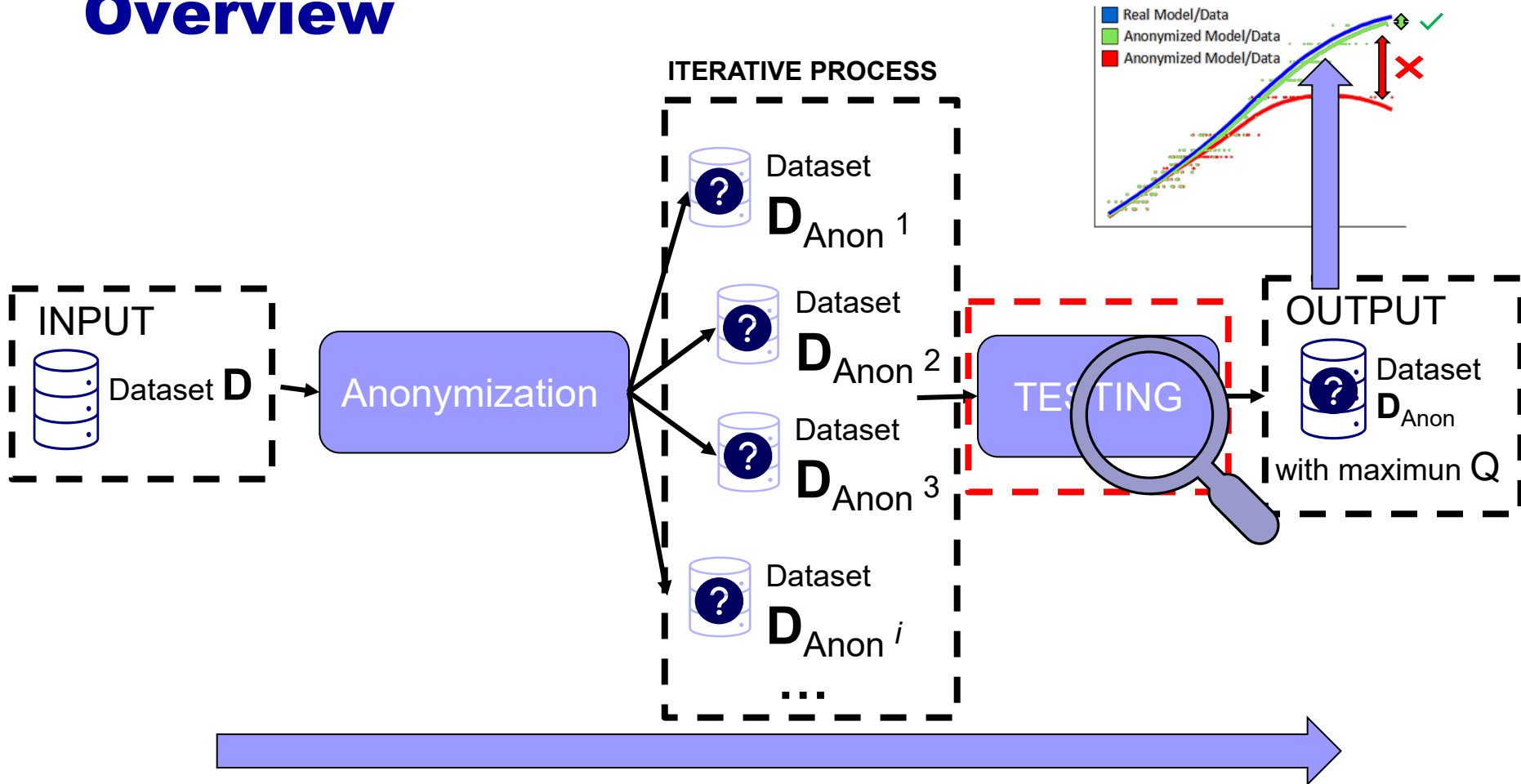
Test Driven Anonymization Overview



Test Driven Anonymization Overview



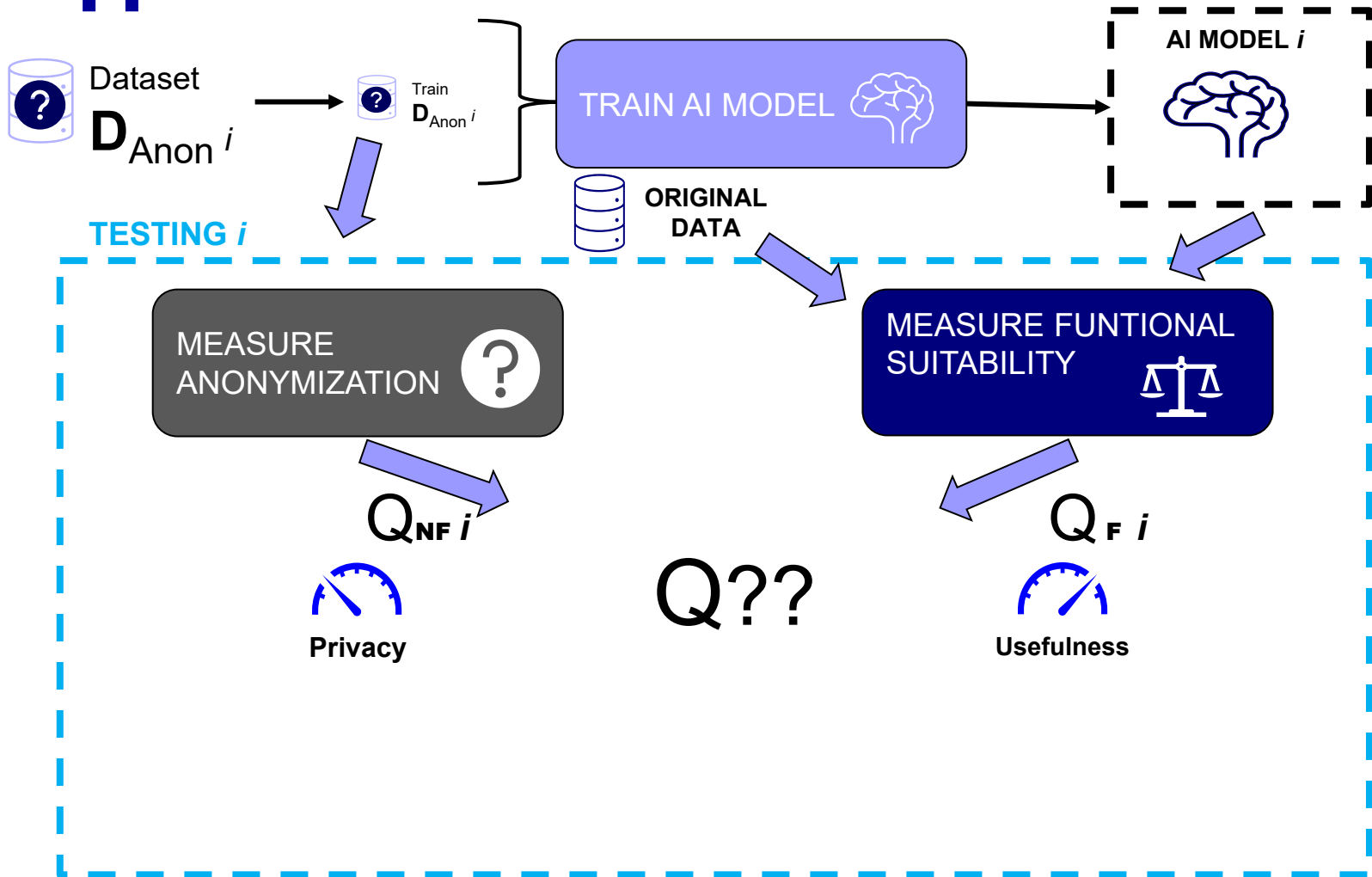
Test Driven Anonymization Overview



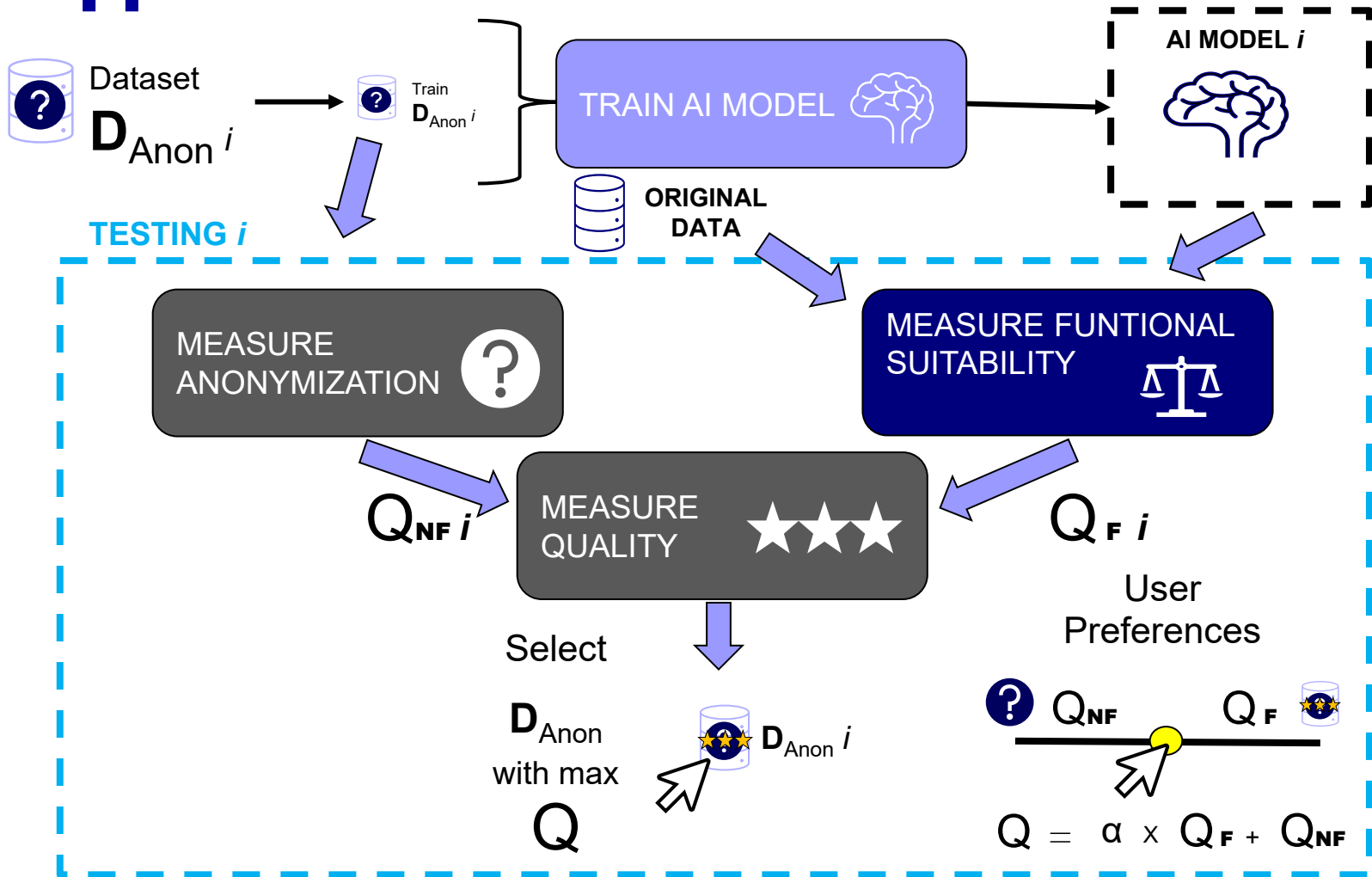
Test Driven Anonymization Approach



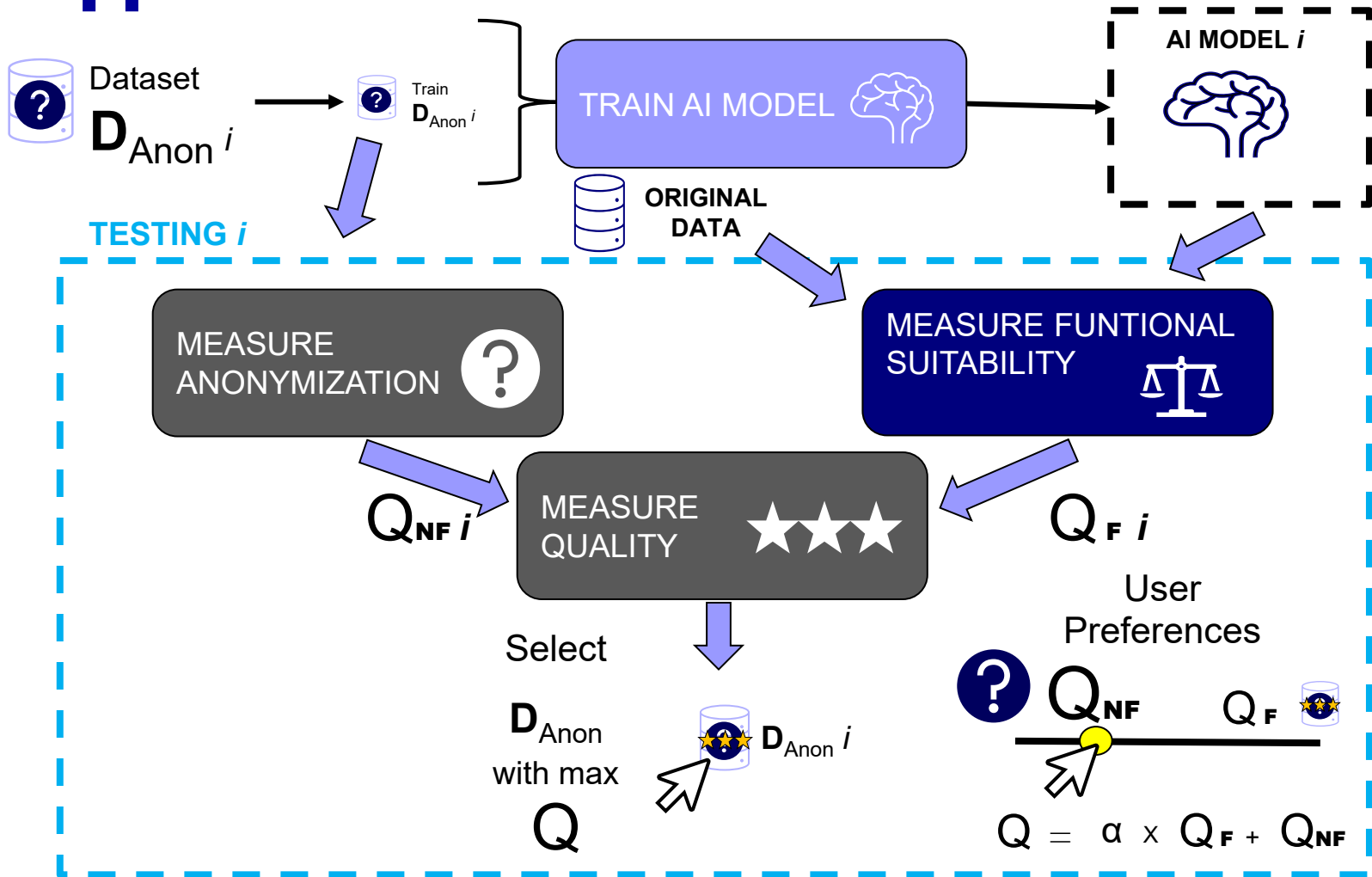
Test Driven Anonymization Approach



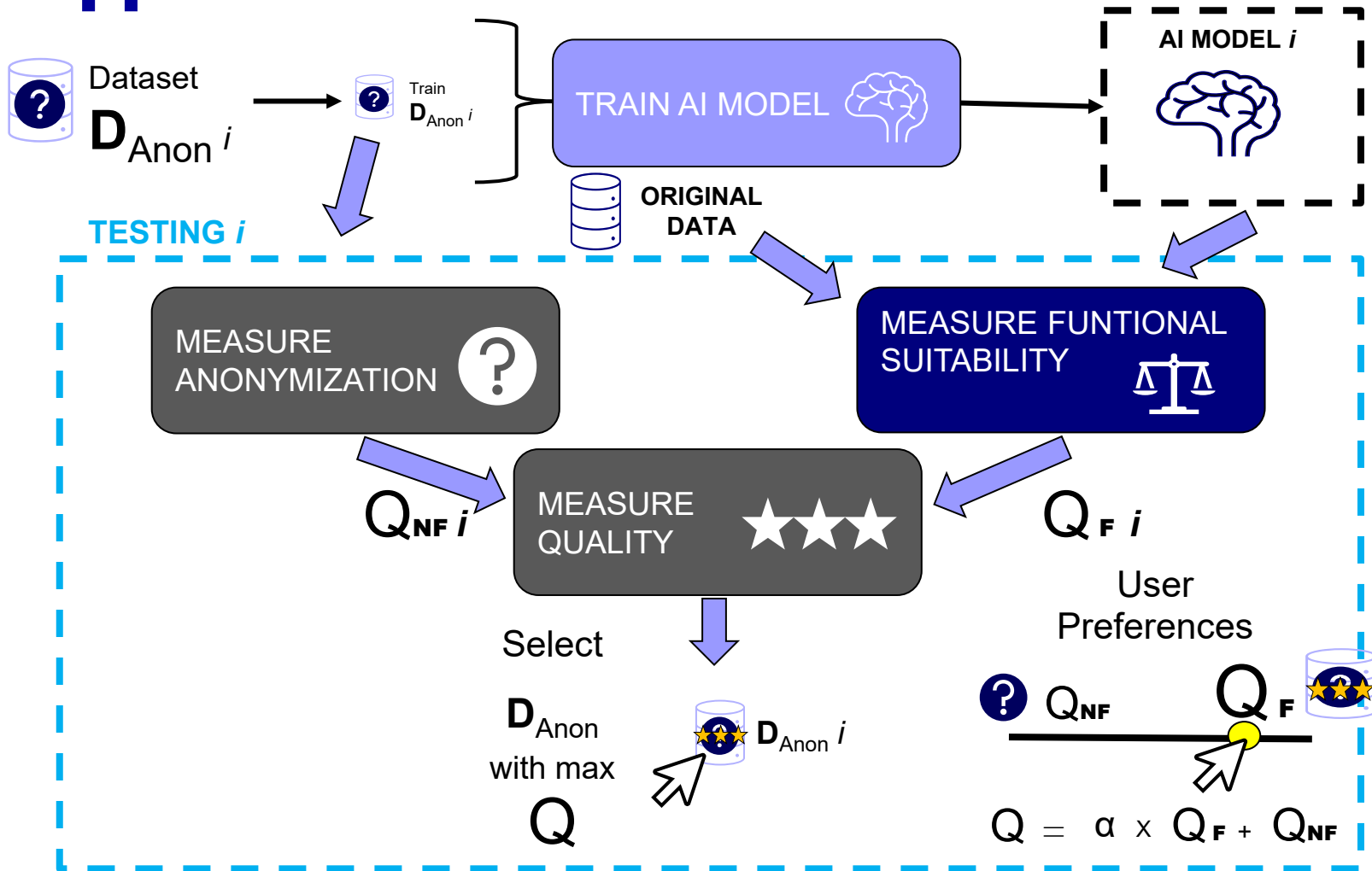
Test Driven Anonymization Approach



Test Driven Anonymization Approach



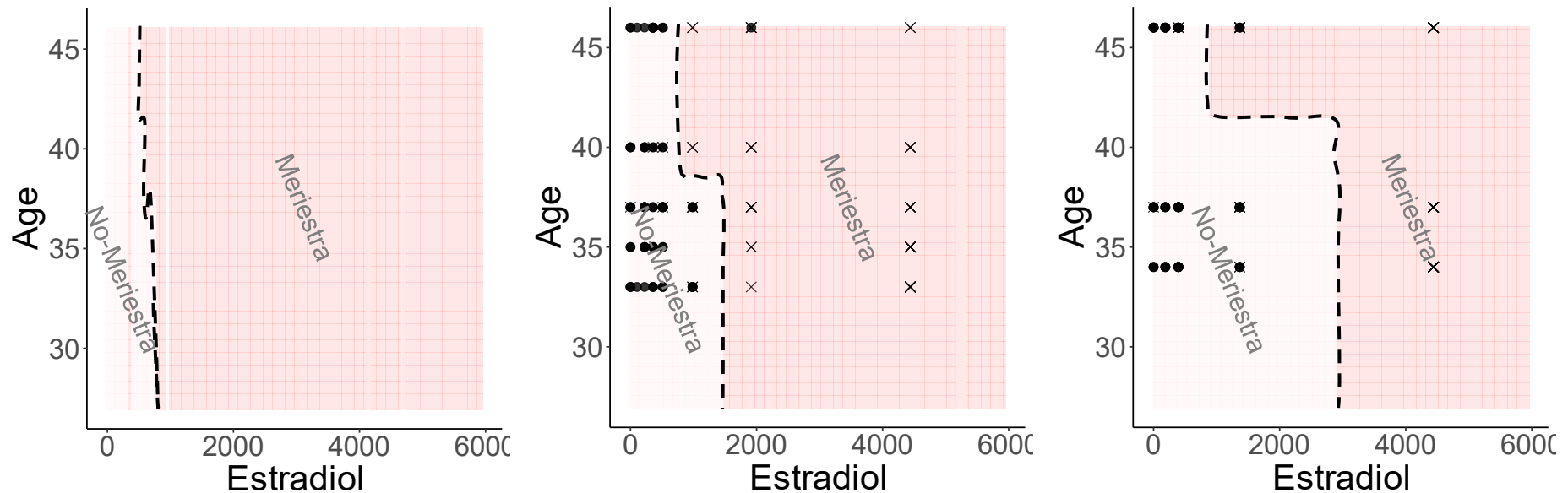
Test Driven Anonymization Approach



Case Study: INEBIR

- Institute for the Study of Biology and Human Reproduction
- Dataset:
 - 123 Institute Patients → Sensitive Data
 - Develop AI → Direct sharing not possible
 - Anonymization → Reduce search space

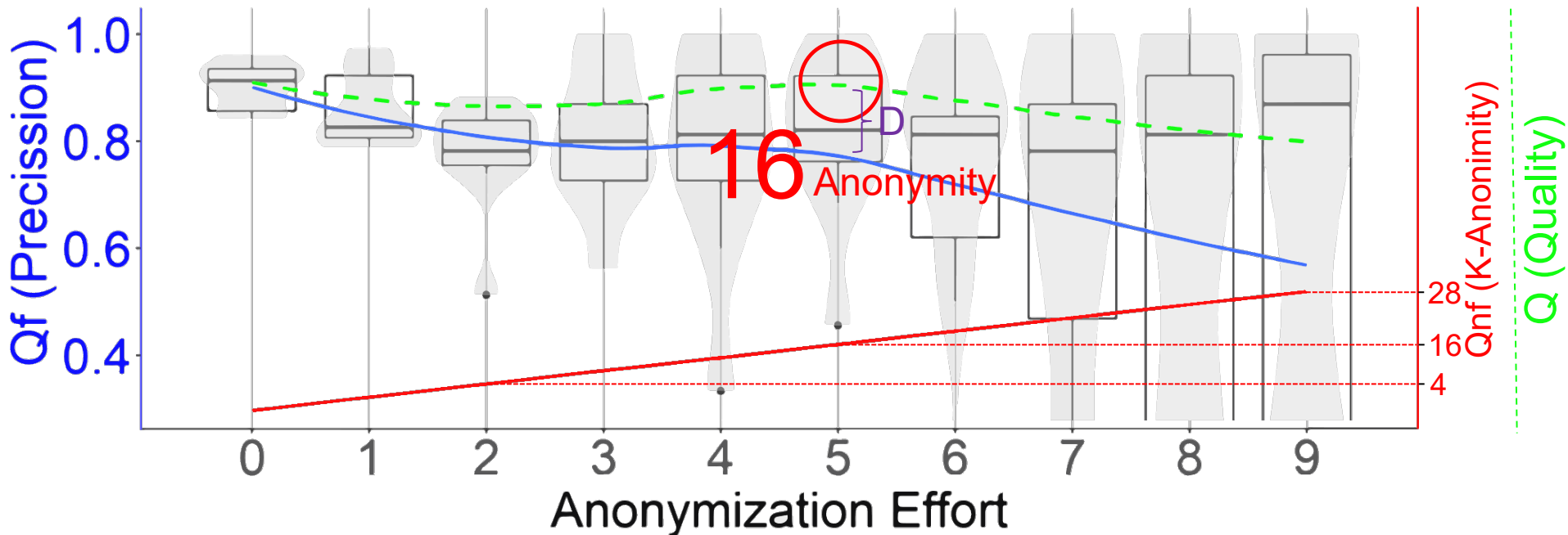
What drug we should provide?



■ INEBIR DATASET:

- Uses a Bagging AI Algorithm for classification, with two categories: Age of the patient and Estradiol Level
- Try to predict if the patient is going to need Meriembra or no

Current Work



- Reach the trade-off between functional and non-functional quality with a 16-Anonymization

Current Work

- Anonymized dataset protects the individual's privacy, not allowing reidentify the patients on it
- Is Anonymized data useful for other models?
 - Yes, i.e. Regression, Random forest...

Conclusions and future work

■ Conclusions

- TDA achieves a trade-off between privacy and functional suitability.
- TDA allows INEBIR the releasing of useful datasets for developing AI tools

■ Future Work

- Continue working with the INEBIR clinic and the Universities of Seville and Huelva.
- Evaluate the dependence between TDA and AI algorithms

This work was supported in part by the Spanish Ministry of Economy and Competitiveness under TestEAMoS (TIN2016-76956-C3-1-R) and POLOLAS (TIN 2016-76956- C3-2-R) projects, and ERDF funds. We want to thank the collaboration of INEBIR (inebir.com) and G7 Innovation (g7innovation.com) in the development of this scientific project



Any Question?

Cristian Augusto, Miguel Olivero, Jesús Morán, Leticia Morales, Claudio de la Riva, Javier Aroba and Javier Tuya

GIIS and IWT2 Reseach Group

<http://giis.uniovi.es> - <http://iwt2.org/>

University of Oviedo - University of Seville – University of Huelva



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

