Matthieu Meeus

+32 496 87 12 80 matthieumeeus@gmail.com LinkedIn/Google Scholar

Education

Imperial College, London – PhD student in Privacy of AI *Oct 2022 – London, UK*

- Part of the Computational Privacy Group under supervision of Prof Yves-Alexandre de Montjoye
- Since the start of my PhD, I have seven first-author papers published/under review. For details see *Publications*.
- Currently continuing the work on privacy/memorization of Large Language Models.
- Current side-project: fine-tuning LLaMA-3 to Dutch (my native language) supported by a grant from the Flemish Super Computer for 40k GPU hours.

Harvard University – M.Sc. in Computational Science & Engineering *Class of 2020 – MA, USA*

- Three semesters of advanced mathematical programming ranging from Numerical Methods to Data Science, Optimization, Systems Development etc. – GPA 3.85/4.0
- Teaching Fellow for the core graduate-level course in Advanced Numerical Methods during Fall 2020
- Fellowship of $100,000 – Belgian American Educational Foundation – Awarded to only 7 students

University of California, Berkeley – M. Eng. in Mechanical Engineering *Class of 2019 – CA, USA*

- One-year master with focus on Advanced Energy Technology – GPA 3.98/4
- Fung Excellence Scholarship – UC Berkeley Fung Institute for Engineering Leadership

University of Leuven – B.Sc. in Engineering – major Mechanical Engineering *Class of 2018 – BE*

- Best of my class after the first year out of more than 500 – Received Best Student Award

Work Experience

Intern at Microsoft Research *May 2024- August 2024 –Cambridge, UK*

- Part of the privacy-preserving machine learning research group, working together with Lukas Wutschitz, Robert Sim and Reza Shokri.
- Working on the privacy auditing of synthetic text data generated by LLMs.

Senior Data Scientist at McKinsey & Company *March 2021- July 2022 – NYC, USA*

- As part of the People Analytics and Measurement team, I built machine learning models to optimize the core asset of McKinsey: its people.
- Leveraged transformer-based NLP models to extract features from free text resumes to improve the existing candidate ranking model by 10% (XGBoost). Now globally deployed.
- Built a 'job-matching' algorithm to detect similar jobs across the Firm, based on categorical features and job description language embeddings. Now globally deployed on the McKinsey career website.
- Built a 'resume-to-job' algorithm, that given free-text resumes recommends the top jobs within the Firm based on your qualifications. Leverages out-of-the-box document embeddings and other text features - best performance achieved with an MLP.
- Built the logic of a semantic search engine for transcripts of expert interviews, based on language embeddings and named entity recognition.
- Many ad-hoc projects including topic modelling on internal surveys, sentiment analysis, prediction of attrition, the composition of ideal teams etc.

Energy Optimization Intern at Tesla Inc. *May-August 2020 – CA, USA*

- Software development for Tesla's Energy Optimization Team, working on optimal battery discharge strategies in Python
- Simulated the performance of a new time series forecasting method: will be implemented based on my analysis.
- Built the first parallelized performance tracking algorithm for one of the team's products using Kubernetes and AWS.

Projects during M. Sc. at Harvard University *Fall 2019-Fall 2020 – MA, USA*

- Research with Prof D. Sondak on solving PDEs with Deep Neural Networks in Pytorch – Spring and Fall 2020

- Leveraged transfer learning in Computer Vision (VGG16) and NLP (BERT) to build a multimodal network for the Visual-Question-Answering (VQA) dataset, reaching a validation accuracy of 47% - built in Tensorflow
- Used Bayesian inference to predict the 2020 US primary elections using Twitter sentiment variation over time - link
- Built a ready-to-use Python package for Automatic Differentiation - link

Software Engineering Internship for EV charging at UC Berkeley/Total *Summer 2019 – CA, USA*

- Built a backend for EV charging management, data visualization and controls in DJANGO with SQL database

Capstone project at UC Berkeley: design Vertical Axis Wind Turbines (VAWT) *2018/2019 – CA, USA*

- Designed and built a VAWT and evaluated performance with combined power and wind data
- First academic team to use wind concentrator on lift-based VAWT – led to an increased output of 900%

Publications

[1] **Meeus, M.,** Jain, S., Rei, M., & de Montjoye, Y. A. (2023). Did the neurons read your book? document-level membership inference for large language models. 33rd USENIX Security Symposium (USENIX Security 2024).

[2] **Meeus*, M.,** Shilov*, I., Faysse, M., & de Montjoye, Y. A. Copyright Traps for Large Language Models. In Forty-first International Conference on Machine Learning (ICML 2024).

[3] **Meeus*, M.,** Guepin*, F., Creţu, A. M., & de Montjoye, Y. A. (2023, September). Achilles' heels: vulnerable record identification in synthetic data publishing. In *European Symposium on Research in Computer Security* (ESORICS 2023).

[4] Guépin*, F., **Meeus*, M.,** Creţu, A. M., & de Montjoye, Y. A. (2023, September). Synthetic is all you need: removing the auxiliary data assumption for membership inference attacks against synthetic data. In *European Symposium on Research in Computer Security* (DPM Workshop at ESORICS 2023).

[5] **Meeus, M.,** Jain, S., & de Montjoye, Y. A. (2023). Concerns about using a digital mask to safeguard patient privacy. *Nature Medicine*, *29*(7), 1658-1659.

[6] **Meeus, M.,** Jain, S., Rei, M., & de Montjoye, Y. A. (2024). Inherent Challenges of Post-Hoc Membership Inference for Large Language Models. *ArXiv preprint.*

[7] Shilov*, I., **Meeus*, M.,** & de Montjoye, Y. A. (2024). Mosaic Memory: Fuzzy Duplication in Copyright Traps for Large Language Models. *ArXiv preprint.*

[8] Guépin, F., Krčo, N., **Meeus, M.,** & de Montjoye, Y. A. (2024). Lost in the Averages: A New Specific Setup to Evaluate Membership Inference Attacks Against Machine Learning Models. *ArXiv preprint.*

Skillset

Technical: Excellent proficiency in AI, Machine Learning, Data Science, Scientific Computing in Python (mainly Pytorch). Experience with cloud computing on AWS, GCloud and Azure. Further proficient in multi-GPU training of large AI models, Docker, Kubernetes, Git, SQL, R, MATLAB.

Languages: Dutch (native), English (fluent), French (fluent), German (basic)

Leisure

Exploring London, adventurous traveling, running ultra-marathons, reading fiction/non-fiction.