How to Fight Financial Crime with AI

Data Intensive Science CDT Seminar

22nd November 2023

UNIVERSITY OF CAMBRIDGE

DiS CDT

FEATURE SPACE

Fraud is a big problem

Fraud is everywhere

40% of reported crime in UK

https://publications.parliament.uk/pa/cm5803/cmselect/cmjust/12/report.html

Fraud is damaging

DiS CDT FEATURE SPACE

Fraudsters are sophisticated

Tech

AI clones child's voice in fake kidnapping scam

'I never doubted for one second it was her,' mother says

Anthony Cuthbertson • Thursday 13 April 2023 17:05 BST • Comments

FraudGPT: The Villain Avatar of ChatGPT

Rakesh Krishnan : Tue, Jul 25, 2023 @ 08:03 AM

Threat Intelligence | Artificial Intelligence | ChatGPT | Dark Web | FraudGPT | Threat Actor

DiS CDT FEATURE SPACE

Featurespace fights fraud

We use machine learning to fight fraud

FEATURE
SPACE

OUTSMART RISK

Our work has global reach

NatWest
BARCLAYS
Danske Bank
MUFG
HSBC
AKBANK
Elavon
zeta
Hay
eftpos
Good for Australia
DiS CDT
FEATURE SPACE
MARQETA
BankservAfrica
modo
permanent tsb
worldpay from FIS
BBVA Banco
MIT
Mercadotecnia Ideas y Tecnología
círculo de crédito
TSYS
A Global Payments Company
usbank

We're based in Cambridge

DiS CDT FEATURE SPACE

Work with us!

We have a cool team

We are the **Innovation Lab** @ Featurespace:

David Sutton
Chief Innovation Officer

Ex: Research Associate in Cosmology @ Cambridge

Iker Perez
Principal Research Scientist

Ex: Assistant Professor in Statistics and Data Science @ Nottingham

Jason Wong
Lead Research Engineer

Ex: MSc in ML @ Edinburgh, Maths @ Oxford

Piotrek Skalski
Senior Research Scientist

Ex: PhD in Physics @ Cambridge

Stuart Burrell
Senior Research Engineer

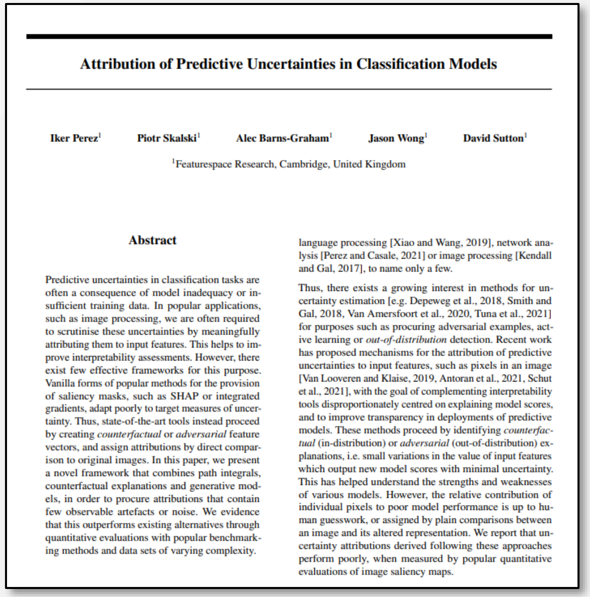
Ex: Research Associate in Mathematics @ St Andrews, MPhil in MLMI @ Cambridge

Kamal Parameswaran
Research Scientist

Ex: Senior Research Associate in Safe & Ethical AI @ Turing

We work on cool research

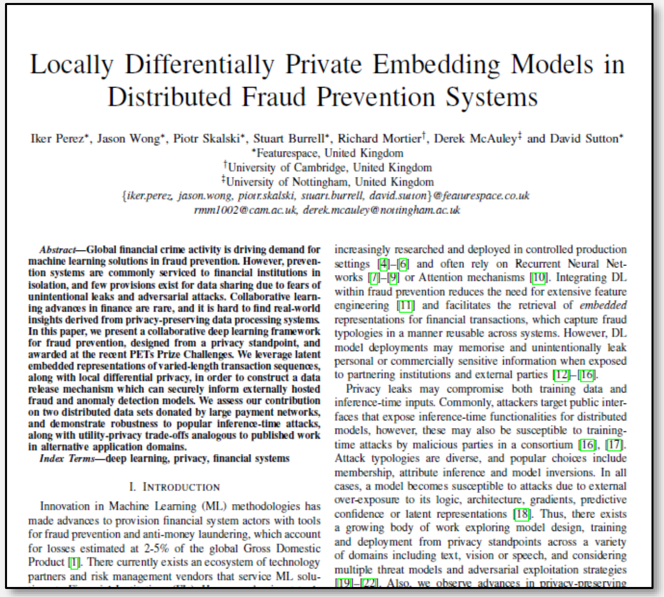
Recent publications:



Attribution of Predictive Uncertainties in Classification Models

Iker Perez[1] Piotr Skalski[1] Alec Barns-Graham[1] Jason Wong[1] David Sutton[1]
[1]Featurespace Research, Cambridge, United Kingdom

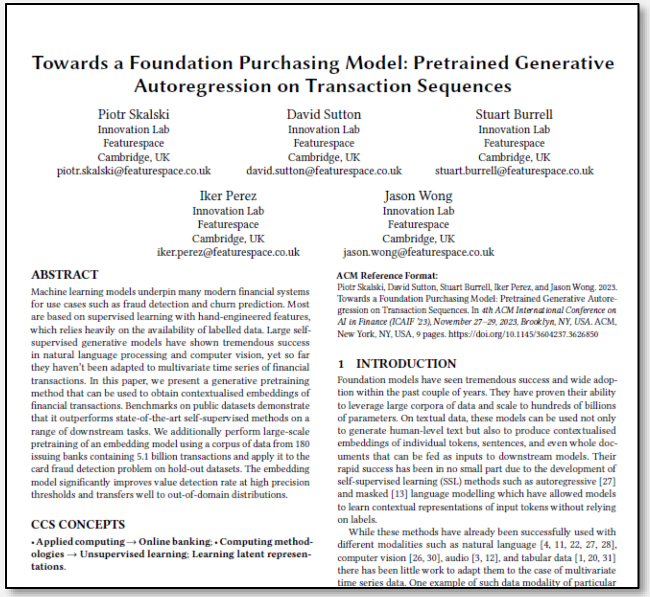
UAI 2022



Locally Differentially Private Embedding Models in Distributed Fraud Prevention Systems

Iker Perez*, Jason Wong*, Piotr Skalski*, Stuart Burrell*, Richard Mortier†, Derek McAuley‡ and David Sutton*
*Featurespace, United Kingdom
†University of Cambridge, United Kingdom
‡University of Nottingham, United Kingdom

ICDM 2023



Towards a Foundation Purchasing Model: Pretrained Generative Autoregression on Transaction Sequences

Piotr Skalski, David Sutton, Stuart Burrell, Iker Perez, Jason Wong
Innovation Lab, Featurespace, Cambridge, UK

ICAIF 2023

We will have open roles + internships

In 2024, we will hiring

- Research scientists
- Research engineers
- Interns

Register your interest @ featurespace.com/innovation-roles

Collaborative AI

Data Intensive Science CDT Seminar

22nd November 2023

UNIVERSITY OF CAMBRIDGE

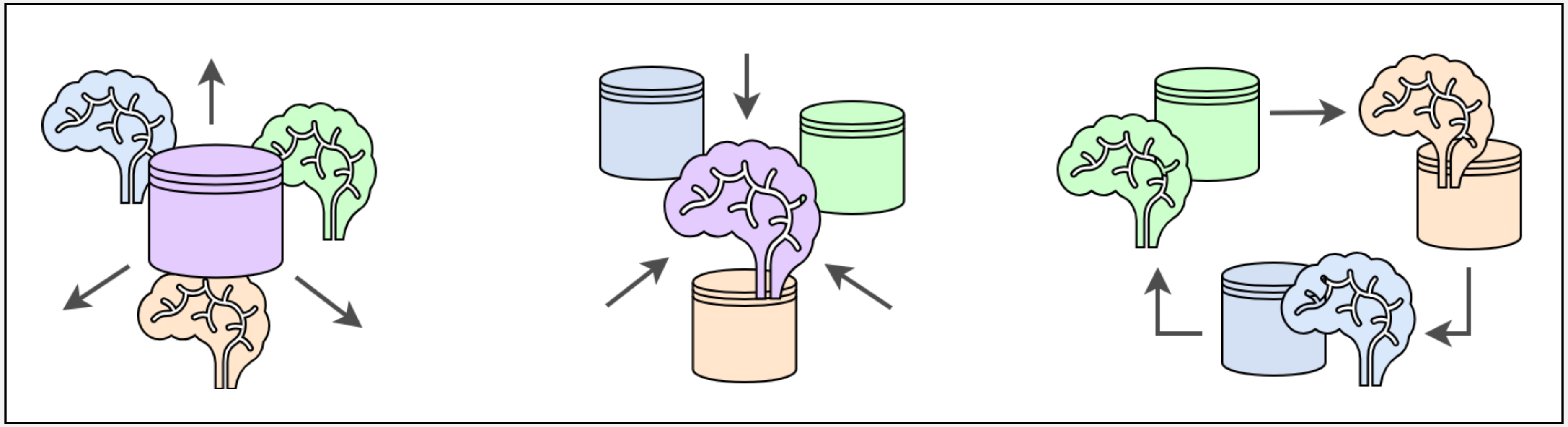
DiS CDT

FEATURE SPACE

Collaborative Financial Crime Detection

Prevention systems are commonly serviced to financial institutions **in isolation.**

- Few provisions exist for data sharing -> Fears of **unintentional leaks** and **adversarial attacks**.

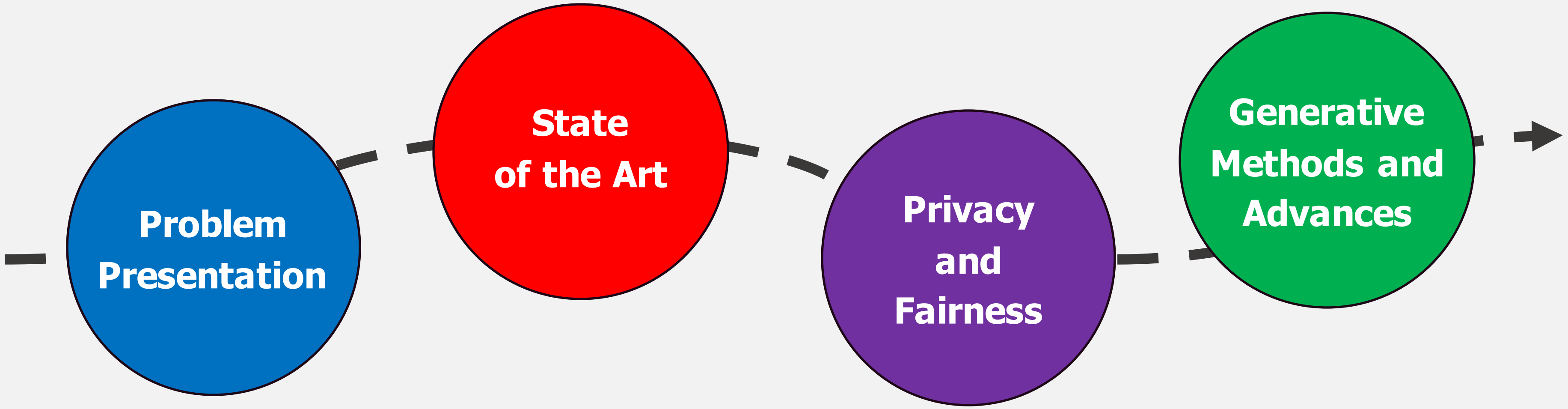


However, there exists evidence that financial data sharing can yield benefits.

Collaborative Financial Crime Detection

Institutions produce incentives for collaborative solutions to financial crime.

Let us go through the following...



Collaborative Financial Crime Detection

Crime detection systems monitor **sequences of transaction data**.

- Commonly, formulated as a supervised learning task for binary classification.

A transaction $\boldsymbol{x}_t$ recorded at time $t > 0$ is endowed with a crime label $y \in \{0, 1\}$, and conditional responses are considered Bernoulli distributed, s.t.

$$y|\boldsymbol{x}_t \sim \mathrm{Ber}(\mathbb{P}(y = 1|\boldsymbol{X}_{\leq t}))$$

where $\boldsymbol{X}_{\leq t}$ are transactions preceding and including $\boldsymbol{x}_t$.

Unbalanced Labels

Originator to Beneficiary

Tabular Data

Throughput

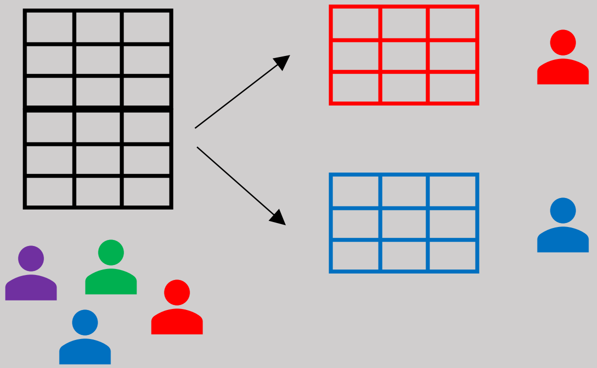
Collaborative Financial Crime Detection

A function $f : \cup_{n=0}^{\infty} \mathcal{X}^n \to [0, 1]$ estimates $\mathbb{P}(y = 1 | \boldsymbol{X}_{\leq t})$.

Decomposed...

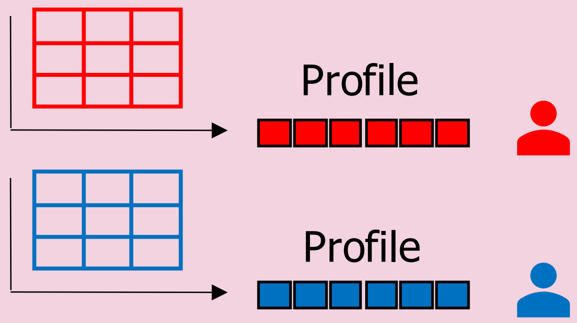
Pre-Processing

$$\boldsymbol{X}_{\leq t} \to (\boldsymbol{X}^{\mathrm{O}}_{\leq t}, \boldsymbol{X}^{\mathrm{B}}_{\leq t})$$



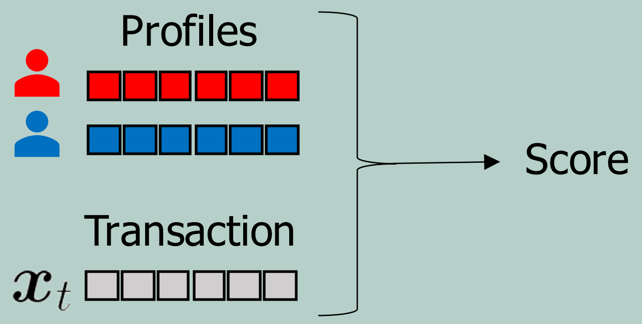
Representation Learning

$$\psi : \cup_{n=0}^{\infty} \mathcal{X}^n \to \mathbb{R}^m$$



GLM Regression

$$s : \mathbb{R}^m \times \mathbb{R}^m \times \mathcal{X} \to [0, 1]$$



Project Idea. We produce scores, observe actions and record causal effects. How to automate algorithms for optimal decision making?

Collaborative Financial Crime Detection

Why Collaborate?

Reliable feature profiles can only be generated from complete transacting histories.

- Available only to their **managing** Financial Institutions.

Profiles for externally managed accounts are **approximated.**

Collaborative Financial Crime Detection

Industry

Basic

Here is database:

- **Option A**: Contains **all** personal data for entities associated in a consortia.
- **Option B:** Contains **minimal** details for a few known criminals accounts.

Here is a **centralised** algorithm:

- Delegate full responsibility to a network or processor, no one is happy.

Academia

Complex

Sophisticated Federated algorithm:

- *It **just** needs 25 GPUs to run.*
- *Must pass gradients around 10 million times, maximum latency of 0.001 milliseconds.*
- *Everyone ensure datasets conform to the same schema and are **distributionally equivalent**.*
- *To make sure it works, please let me centralise all data just this time.*
- *The **orchestrator** rules, period.*

Concerns: Heterogeneity, Adaptability, Scalability, Efficiency and Regulatory Compliance.

Collaborative Financial Crime Detection

Why is this gap not bridged?

Privacy, Fairness and Compliance.

Collaborative models **convoluted** and **leak information** when subjected to adversarial attacks.

Privacy enhancing technologies are researched in model designs:

- **Differential Privacy**
- **Homomorphic Encryption**
- **Secure Multi-Party Computation**
- **Zero Knowledge Proofs**

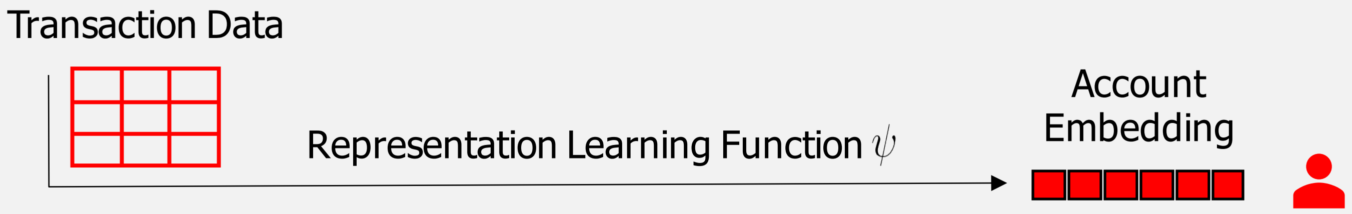
Project Idea. How to design attacks against public model APIs? What about Federated Learning models?

Collaborative Financial Crime Detection

***Simplify**: Do not train joint models, share the profiles.*

Transaction Embeddings: Numerical representations of variable-length transaction data.

- A representation learning function $\psi : \cup_{n=0}^{\infty} \mathcal{X}^n \rightarrow \mathbb{R}^m$ **is** an embedding function.



Obfuscated profiles. **Is sharing safe?**

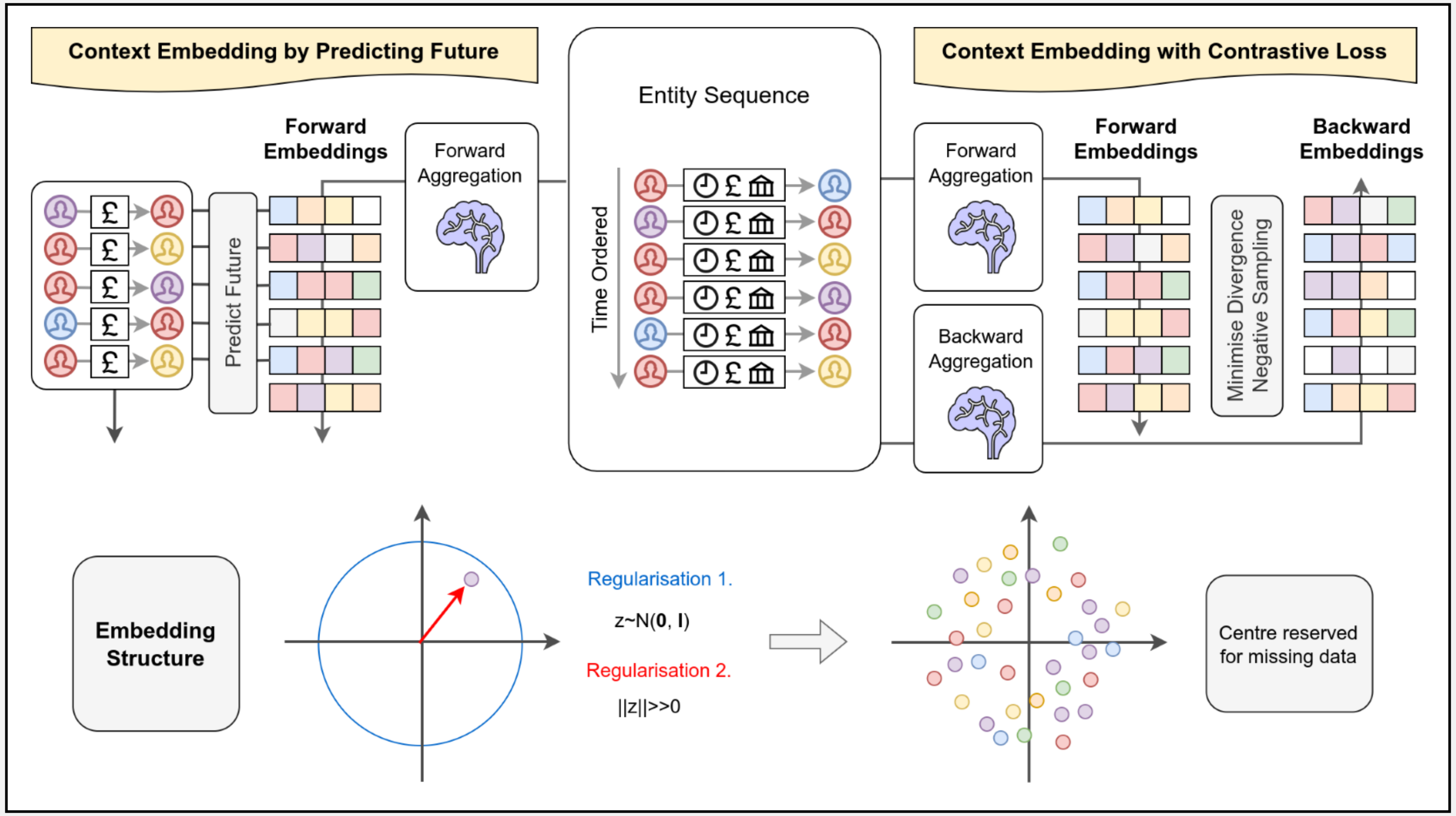
- What do recipients need? How to make them comprehensive? How protect sensitive attributes?

Collaborative Financial Crime Detection

Generative Advances: Training

- Language Approach? Next-event prediction, dual encoding…
- Vision Approach? Contrastive Learning, Adversarial Training, …

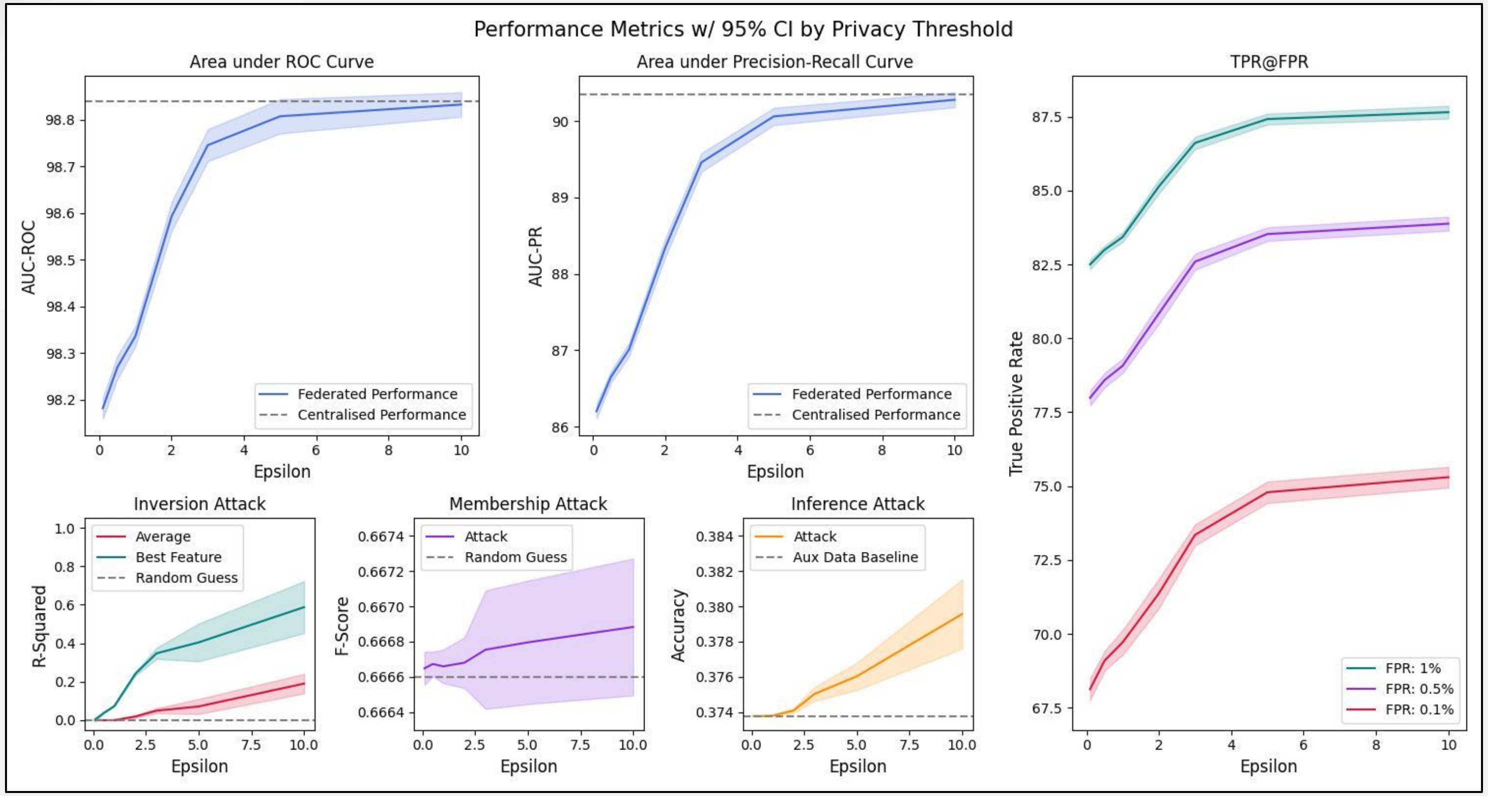
Ideas…



Project Idea. How design such a method properly? How accommodate multimodal outputs? How leverage embeddings for Generative data synthesis? Other purposes?

Collaborative Financial Crime Detection

Transfer Learning with **Differentially-Private** Generative Embeddings...



Algorithm 1: SGD step. Bank $\beta = 1$. Peer-to-peer.

Input: Mini-batch $\{(\boldsymbol{x}_{i,t_i}, \boldsymbol{z}^{\mathrm{o}}_{i,t_i}, y_i)\}_{i=1,\dots,N}$.
Loss function $\mathcal{L}$ and learning rate $\lambda > 0$.
Slack term $\gamma \to 0^+$.

Output: Updated s_1 and $g_{1,\beta}$, $\beta = 2, \dots, B$.

for $i = 1, \dots, N$ **do**

$\rho_i \leftarrow getBeneficiary(\boldsymbol{x}_{i,t_i})$

$\hookrightarrow$ **Beneficiary Bank** ρ_i

Publish $\boldsymbol{z}^{\mathrm{b}}_{i,t_i} = \mathcal{M}_{\rho_i}(\boldsymbol{X}^{\mathrm{B}}_{\leq t_i})$

Pre-process $\boldsymbol{r}^{\mathrm{b}}_{i,t_i} = g_{1,\rho_i}(\boldsymbol{z}^{\mathrm{b}}_{i,t_i})$

Predict $\hat{y}_i = s_1(\boldsymbol{z}^{\mathrm{o}}_{i,t_i}, \boldsymbol{r}^{\mathrm{b}}_{i,t_i}, \boldsymbol{x}_{i,t_i})$

end

Update weights ω_s of s_1:

$$\omega_s \leftarrow \omega_s - \frac{\lambda}{N} \sum_{i=1}^{N} \nabla_{\omega_s} \mathcal{L}(y_i, \hat{y}_i)$$

for $\beta = 2, \dots, B$ **do** update weights ω_g of $g_{1,\beta}$:

$$\omega_g \leftarrow \omega_g - \frac{\lambda}{N_\beta + \gamma} \sum_{i=1}^{N} \mathbb{I}_{\beta=\rho_i} \cdot \nabla_{\omega_g} \mathcal{L}(y_i, \hat{y}_i)$$

where $N_\beta = \sum_{i=1}^{N} \mathbb{I}_{\beta=\rho_i}$.

end

Natural language interfaces

Data Intensive Science CDT Seminar
22nd November 2023

UNIVERSITY OF CAMBRIDGE

DiS CDT

FEATURE SPACE

Structure

- Motivation for natural language interfaces

- Discuss three open problems:

  1. Develop trust-worthy, fair, and robust agents for user support

  2. Generate natural language explanations of automated decision making

  3. Build a financial crime detective to automate complex investigations

- Q&A

Motivation

- Recent advances in LLM technology (GPT-4/Turbo, GPTs, LLama2) have the capacity to transform how users interact with software:

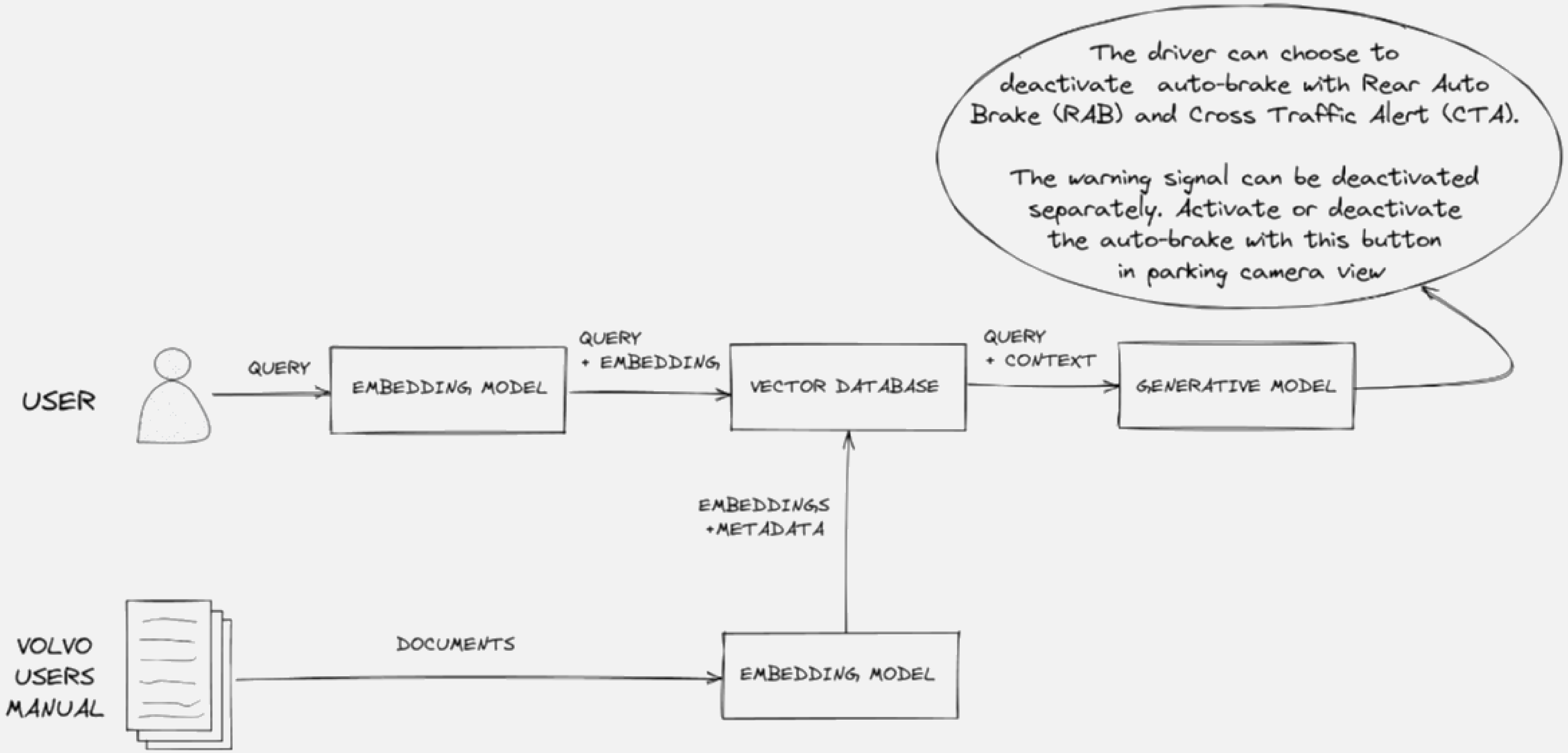
  i. *Can you do action X?*

  ii. *How do I achieve Y?*

  iii. *Why did you produce output Z?*

Problem 1. Agents for user support

- Typical approach uses *Retrieval Augmented Generation* – embedding relevant information within LLM prompts obtained via efficient similarity search over vector databases (image credit: Pinecone).



DiS CDT FEATURE SPACE

FEATURE
SPACE
OUTSMART RISK

Welcome! I'm DocBot, here to guide you through our technical systems. Ask me a question and I'll answer based on the latest Featurespace documentation, direct you to related resources, and happily discuss any follow-up questions you might have.

How may I help you today?

SUBMIT

Powered by Llama 2
Knowledge bases: ARIC Documentation

Research projects

1. How do we **robustly** and **automatically** evaluate performance?

2. How best to incorporate **user feedback** and/or **AI feedback** for continuous fine-tuning?

3. Integrating **diverse knowledge** bases (Slack, technical documentation, ...) and construct efficient retrieval regimes (recursive retrieval, reranking, ...)

4. Building an coding-assistant for a **proprietary coding language** with only sparse datasets?

5. Engineering the above given **latency**, **hardware** and **API** constraints

Problem 2. Natural language augmented explainable AI

- **Explainable AI (XAI)** is a huge field trying to pry open the **black box** of neural networks and explain **why** they produce the outputs they do.

- Simple approach is to **limit hypothesis space** to inherently interpretable models: Bayesian Rule Lists, **Sparse Linear Models**, et al.

- **Post-hoc** explainability measures attempt to explain arbitrarily complex models after-the-fact.

  i. **Feature importance**: which dimensions of the feature vector contributed most to the prediction? (Saliency, Lime, Integrated Gradients, Shap, DeepLift, and Perturbation Masks, …)

  ii. **Example importance**: which training examples contributed most to the prediction? (Influence Function, Deep K-Nearest, Neighbours, TraceIn, SimplEx, …)

- Caveat – these methods are **far from perfect**, but are continuously improving.

DiS CDT FEATURE SPACE

Research project

- **Creating natural language model explanations from post-hoc XAI outputs**

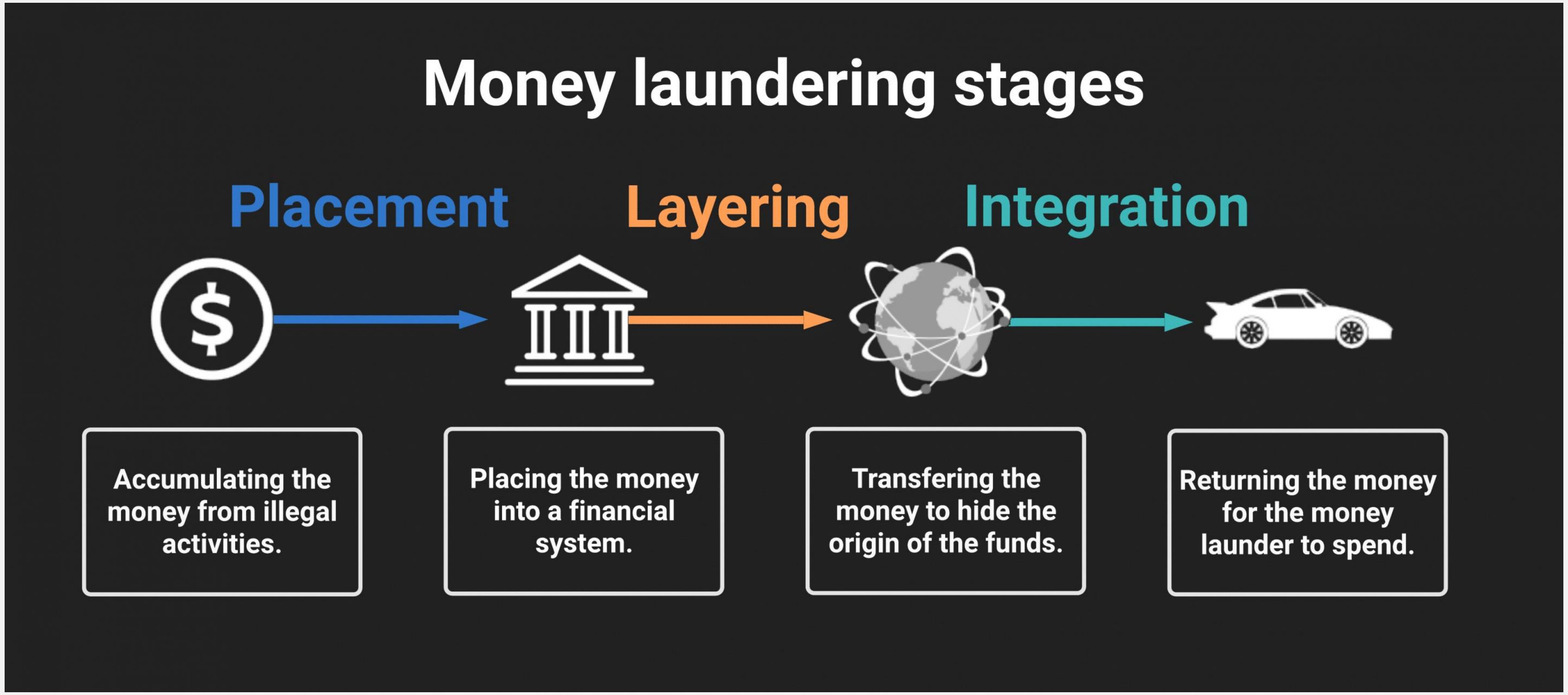
  i. Can we use an LLM to reliably convert heatmaps of feature importance scores into natural language explanations?

  ii. Can we use an LLM to explain a fraud prediction based on natural language summaries of relevant training examples?

	Event time	Time since previous	POS entry mode	Merchant country	Amount	AVS result	CVV result	Merchant fraud rate	MCC fraud rate	Merchant country fraud rate	Score
5	2017-11-06 14:29:30	10 days 15:08:00	e-commerce	UK	67.54	no match	match	-6.00	-6.00	-7.00	0.95
4	2017-10-26 23:21:17	16 days 14:16:00	e-commerce	UK	60.43	no match	match	-6.00	-6.00	-7.00	0.96
3	2017-10-10 09:05:41	28 days 16:13:00	keyed	UK	375.94	full match	match	-8.00	-8.00	-7.00	0.04
2	2017-09-11 16:53:00	4 days 02:53:00	e-commerce	Luxembourg	34.26	full match		-12.00	-11.00	-11.00	0.16
1	2017-09-07 14:00:26	21 days 23:04:00	keyed	UK	25.47	full match	match	-8.00	-15.00	-18.00	0.18
0	2017-08-16 14:56:02	0 days 00:00:00	keyed	UK	50.14		match	-7.00	-13.00	-16.00	0.33

Problem 3. Build a financial crime detective to automate complex anti-money laundering investigations

- The capacity of governments and financial institutions to crack-down on crimes such as anti-money laundering (AML) is severely bottlenecked by manual labour.



Research project

- Can we fine-tune a LLM to **accurately** and **verifiably** assess and summarise suspicious behaviour in a transaction history suspected of AML?

- How can we **evaluate** the resulting LLM to assess whether it is reporting in a **fair** and **unbiased way** with respect to protected attributes? (Learning from human preferences, OpenAI 2017, Constitutional AI, Anthropic 2022).

- Can we build out a fully-fledged investigator agent; equipped with **tools** (API access, web-search) to aid an investigation?

Thanks for listening!

Feel free to add my on **LinkedIn** or email stuart.burrell@feauturespace.co.uk if you'd like to talk more.

References

- Bai, Yuntao, et al. "**Constitutional AI: Harmlessness from AI feedback.**" *arXiv preprint arXiv:2212.08073* (2022).
- Christiano, Paul F., et al. "**Deep reinforcement learning from human preferences.**" *Advances in neural information processing systems* 30 (2017).
- Lundberg, Scott M., and Su-In Lee. "**A unified approach to interpreting model predictions.**" *Advances in neural information processing systems* 30 (2017).
- Adadi, Amina and Mohammed Berrada. "**Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI).**" IEEE Access 6 (2018): 52138-52160.
- LlamaIndex, Liu, J., https://github.com/run-llama/llama_index

Work with us!

FEATURE SPACE

featurespace.com/innovation-roles