SONNETS : Scalability Oriented Novel Network of Event Triggered Systems

EPSRC Programme (EP/X036006/1)

David Thomas, Wayne Luk, Rishad Shafik, Andrew Brown

Mark Vousden, Graeme Bragg, Alex Yakovlev, Kabita Adhikari, Walter Distaso

Shidur Rahman, Alex Chan, Ce Geo, Ekin Erkus, Bob Pattison, Saram Abbas

June 2026

This is the second Advisory Board report for the SONNETS (Scalability Oriented Novel Network of Event Triggered Systems) programme grant. The objective of the SONNETS project is to explore a new paradigm for describing large-scale distributed systems called Event-Triggered Computing (ETC), with the objective of pioneering new AI-based financial risk-analysis capabilities for the UK.

The report contains 6 main sections:

- Section 1 — A one-page executive summary of the report.
- Section 2 — Advisory Board meeting information, including schedule and locations.
- Section 3 — An overview of the SONNETS programme's overall objectives and current progress.
- Section 4 — Engagement with stakeholders and demonstrators built over the last year
- Section 5 — An overview of the key research themes and publications
- Section 6 — Proposed objectives and operational plans for the remainder of the project

The Year 1 Advisory Board report and post Advisory Board objectives document can be found at:
https://sonnets-project.org/documents/

1. Executive summary

SONNETS (Scalability Oriented Novel Network of Event Triggered Systems) is a five-year UKRI-funded research programme. It aims to transform real-time national financial risk analysis through a novel computing paradigm: **Event-Triggered Computing (ETC)**. SONNETS integrates computing, modelling, and AI to create scalable, explainable, and deployable systems for financial surveillance and stress testing. Ultimately, the project should provide a “digital shadow” of the UK’s financial system, enabling timely, explainable, and actionable insights for policymakers and regulators.

SONNETS was originally structured around three interdependent research tiers:

- **ETC (Event-Triggered Computing)**: Cloud-deployable infrastructure for distributed, large-scale, real-time applications created using an "event-triggered" programming abstraction.
- **ETM (Event-Triggered Modelling)**: Real-time financial analysis to uncover financial interdependencies, estimate shock propagation, and generate risk scenarios.
- **ETAI (Event-Triggered Artificial Intelligence)**: Novel types of machine learning and feature extraction to provide transparent and interpretable forecasting.

The lines between these original tiers are naturally blurring as the research proceeds and the project goals are re-defined using input from stakeholders. Within this report, we address action points from the previous advisory board, including reviewing objectives, defining a vision for the end of the project, and providing a set of objectives on which the AB can provide feedback.

Achievements since the first Advisory Board in May 2025 include:

- Substantial engagement with the Bank of England (BoE) to identify real-world needs and case studies.
- Extensive fundamental and applied research resulting in 24 new peer-reviewed publications and five Best Paper Awards.
- Development of three new demonstrators, showcasing:
  - Y2Demo-Explainability: New explainable machine-learning techniques applied to finance.
  - Y2Demo-CVAEForecast: Real-time and event-triggered learning in a live system.
  - Y2Demo-ABMCalibration: Calibration of a large Agent-Based Model (ABM) to open datasets.
- Generalisation of the research to explainable risk prediction for bladder cancer patients.

Over the next year, we plan to:

- Develop a new demonstrator which combines existing explainability and real-time updates with autonomous compute scaling and distributed explainable learning.
- Perform fundamental research into plausible shock generation and injection into models.
- Demonstrate automatic calibration and optimisation of ABMs against live data.
- Develop a new tool for bladder cancer risk analysis in collaboration with NHS practitioners.
- Broaden stakeholder engagement beyond BoE to include commercial finance and other users.

Table of Contents

2. Advisory Board Meeting Information

Thank you for agreeing to attend the second SONNETS Advisory Board Meeting, which will take place on June 24th from 10:00 to 16:00 at Imperial College London.

The Advisory Board will review our research progress and future plans, and provide guidance on both the research and the overall project.

We provide information about the grant in two forms:

- An Advisory Board Report (this document), describing research and project progress so far.
- Presentations and demonstrations of the research outputs during the meeting.

2.1. Meeting location

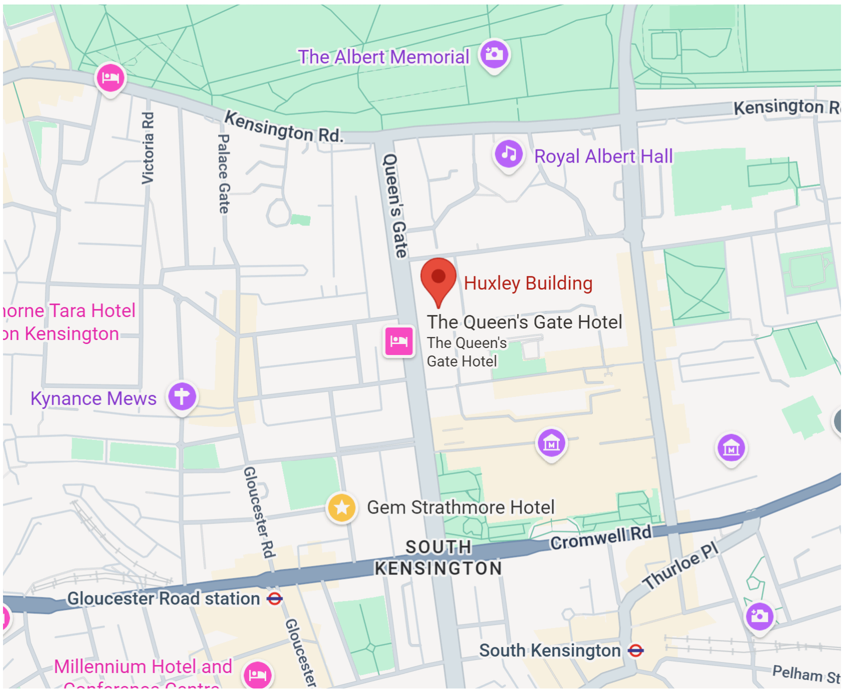
The meeting location is at:

- Huxley building: room 217 / 218
- South Kensington
- London
- SW7 2PG

Travel information can be found at: https://www.imperial.ac.uk/computing/about/getting-here/

A campus map can be found at: https://www.imperial.ac.uk/media/imperial-college/visit/public/SouthKensingtonCampus.pdf

It is recommended that you enter from the Queen's Road entrance, as then rooms 217 and 218 are immediately to your left from the entrance. The rooms are behind a security door, but we will have someone there to let you in.



2.2. Timetable

An indicative timetable for the meeting on June 24th is as follows, though the internal structure may change between preparing the report and the day of the meeting.

- 09:30 - 10:00 : Arrival and coffee
- 10:00 : Meeting starts
- 10:00 - 10:45 : Introductions and project overview
  - Introductions
  - Recap of high-level goals
  - Current goals and progress
  - Project financial position
- 10:45 - 11:15 : Published research highlights
- 11:15 - 11:30 : *Coffee break*
- 11:30 - 13:00 : Demonstrators
  - CVAE demonstrator : real-time market-driven learning and forecasting
  - ABM calibration : Calibrating a real Agent-Based Model against economic data
  - Explainable risk : interactive exploration of an explainable risk model
- 13:00 - 14:00 : *Lunch*
- 14:00 - 15:00 : Future plans and deliverables
  - Generalisation : Explainable risk prediction for bladder cancer and low-latency physics
  - Research focuses and priorities
  - Project decisions and requests for AB feedback
- 15:00 - 15:30 : Advisory Board private discussion
- 15:30 - 16:00 : Feedback of Advisory Board
- 16:00 : Meeting closes

3. Progress overview

We start with a review of action points and objectives from the previous Advisory Board (AB).

3.1. Action points from previous Advisory Board

Feedback from the previous Advisory Board meeting included the following key points. Where appropriate, we have linked forward to sections of the report which address these points.

The project needs clear directions and a clear focus:

- AP1.1 : Near-term goals established before next Advisory Board : Section 3.2
- AP1.2 : Need minimum viable product demonstrators : Section 4
- AP1.3 : Establish final overall goals at the next Advisory Board : Section 6

More engagement needed with Bank of England:

- AP2.1 : Near-term problem statement to be co-developed with BoE : Section 3.2 and Section 4.1
- AP2.2 : What are real-world problems for stakeholders? : Y2Demo-ABMCalibration and Section 4.1
- AP2.3 : How do we maximise eventual impact? : discussed in Section 6.3.3

Other comments:

- AP3.1 : Good things happened individually - what is the overall focus? : Section 6
- AP3.2 : What does an explainable Tsetlin Machine look like?: Section 5.1 and Y2Demo-Explainability
- AP3.3 : Generalisation - what are the non-financial follow-ups? : Section 6.2
- AP3.4 : Good individual technologies - how do they tie together? : Section 4 and Section 6
- AP3.5 : Financials - it is a lot of money, how is it being spent? : Section 3.3
- AP3.6 : Impact - is it purely regulatory, or are there commercial plans? : *additional impact is covered in Section 5.3, though currently still non-commercial.*
- AP3.7 : How is new IP being managed and protected? : discussed in Section 6.3.3

Report:

- AP4.1 : The report should be shorter : *we have reduced the core of the report from 44 pages to 16 pages, and have focused on the key points.*
- AP4.2 : It should have more of a focus on overall goals : *We now begin with a review of existing goals (Section 3.2), and conclude with clear project-level plans (Section 6.1) and timelines.*
- AP4.3 : Next report should contain a clear vision for intended project deliverables on which AB can give feedback : *Section 6.1 describes overall project deliverables, and Section 6.3 describes specific plans for the upcoming year.*

3.2. Objectives versus progress made

Action points [AP1.1], [AP1.2] and [AP2.1] requested documented goals for the next year, which were captured in a problem definition document (available at https://sonnets-project.org/documents/). The following table summarises those objectives, their current status, and whether the objectives were met.

Goal	Description	Achieved	Details
Y2G1	Automated ingestion of live data	Yes	Y2Demo-CVAEForecast
Y2G2	Live risk analysis updated daily	Yes	Y2Demo-CVAEForecast
Y2G3	Display of risk analysis results	Partial	Y2Demo-CVAEForecast
Y2G4	System uptime of at least 7 days	Yes	Y2Demo-CVAEForecast
Y2G5	Autonomous re-calculation of analysis	Yes	Y2Demo-CVAEForecast
Y2G6	Display of risk forecast on public site	Yes	Y2Demo-CVAEForecast
Y2G7	System deployed in cloud infrastructure	Yes	Y2Demo-CVAEForecast
Y2G8	System uptime of at least 31 days	Partial	Y2Demo-CVAEForecast
Y2R1	Daily estimates of causal graphs	Partial	pivot : Y2Demo-ABMCalibration
Y2R2	Daily predictions of one-month ahead explainable risk	Partial	Y2Demo-CVAEForecast
Y2R3	Daily predictions of one-month ahead using classical risk	Partial	Y2Demo-CVAEForecast
Y2R4	Automated back-testing of risk estimates	Partial	Y2Demo-CVAEForecast
Y2R5	Incremental re-learning and re-optimisation of explainable models	Partial	found in Y2Demo-CVAEForecast pivot : Y2Demo-Explainability
Y2R6	Scenario generation of plausible and worst-case scenarios	No	pivot : Y2Demo-ABMCalibration
Y2R7	SVAR analysis of key economic variables using causal graphs	No	pivot : Y2Demo-Explainability
Y2R8	Addition of new data sources to risk analysis	No	pivot : Y2Demo-Explainability
Y2D1	FTSE 100 hourly data	Yes	Y2Demo-CVAEForecast
Y2D2	Daily data from the Bank of England	Yes	Y2Demo-Explainability
Y2D3	ONS macroeconomic data	Yes	Y2Demo-ABMCalibration
Y2D4	Bank of England interest rate changes	Partial	Y2Demo-ABMCalibration
Y2D5	News sources describing major economic shocks	No	

At the first AB it was expected that SVAR and causal analysis would form a larger part of the ongoing research, but engagement with stakeholders suggested a pivot towards explainability which has become a much bigger component through Y2Demo-Explainability. Both SVAR and causal analysis remain important parts of underlying research we expect to integrate, but were not used in key demonstrators this year. For example, we have internal proof-of-concept demos of live causal graph extraction over multiple timescales, but have not yet developed them into full demonstrators. There has also been an intentional pivot towards Y2Demo-ABMCalibration as this was of direct interest to BoE.

The existing post-AB plan envisioned a single demonstrator that encompassed all features, but this ultimately did not make sense as it limited the research into methods for interactively explaining risk which emerged as a key requirement from stakeholders. As a consequence the functionality has been split over two demonstrators: Y2Demo-CVAEForecast and Y2Demo-Explainability.

3.3. Finances

There are two main financial changes (approved by EPSRC) to report:

- Appointment of project manager : originally Southampton held budget for a project management + EDI support post, but it was difficult finding a part-time project manager with the correct experience. An excellent candidate who was already performing project management at Newcastle was found, so budget was transferred from Southampton to Newcastle to support this.
- DI (Directly Incurred) to DA (Directly Allocated) swap : Prof. Walter Distaso is a project member providing expertise on the macro-financial side of things, originally employed by Imperial with DA costs. They have since changed employment but are still important to the project. DI staff funds from Southampton are being swapped with DA staff funds at Imperial to allow Walter to continue.

A snapshot of the current financial state is summarised in the following table, which shows the directly controllable budget for each institution, and ignores estates and other costs.

Institution	Status	Staff DI	Staff DA	Travel	Equip/Cons.	Total
Southampton	Spent	£92,026	£143,488	£16,233	£6,241	£257,988
	Budget	£669,578	£704,245	£163,700	£154,500	£1,692,023
	Used	13.7%	20.4%	9.9%	4.0%	15.2%
Imperial	Spent	£198,180	£152,226	£34,404	£9,127	£393,937
	Budget	£944,718	£225,725	£118,394	£104,465	£1,393,302
	Used	21.0%	67.4%	29.1%	8.7%	28.3%
Newcastle	Spent	£186,038	£81,489	£50,417	£11,052	£328,996
	Budget	£627,378	£232,825	£58,000	£150,000	£1,068,203
	Used	29.7%	35.0%	86.9%	7.4%	30.8%
Total	Spent	£476,244	£377,203	£101,054	£26,420	£980,921
	Budget	£2,241,674	£1,162,795	£340,094	£408,965	£4,153,528
	Used	21.2%	32.4%	29.7%	6.5%	23.6%

Budget reports run at least quarter behind, so these costs are a picture from Jan-March 2026, or just over 40% through the project. The particularly high Staff DA costs for Imperial are due to the reallocation of Walter Distaso's time.

Currently the burn rate is low, which is due to:

1. Weak research fellow recruitment at Southampton for a combination of reasons: poor applicants; staff overload with teaching; time taken to organise and approve financial changes
2. Intentionally holding off on equipment purchases until research needs were clearer
3. Cloud credits have been reserved until they are needed for larger demonstrators and experiments

Actions taken to address this over the remainder of the project are:

1. Purchasing of large memory GPUs to support distributed training
2. Allocation of Staff DA time to Southampton academic staff enable more direct engagement
3. Member of Southampton academic staff (Mark Vousden) to champion and manage recruitment
4. Transfer of DI staff budget from Southampton to Newcastle to support generalisation

4. Demonstrators and engagement

In this section we highlight three distinct demonstrators that have been developed over the last year as part of our current year objectives, and also through engagement with Bank of England and other stakeholders. This is not an exclusive set of demonstrators, and more will be shown at the AB meeting, but these are most closely aligned with our core project objectives.

4.1. Engagement activities with stakeholders

Given the key goal on delivering national impact in financial risk analysis, we have focussed on collaboration with the Bank of England as a key stakeholder. Engagement with BoE was already ongoing, but we have stepped this up in response to comments from the previous AB, including:

- [AP2.1] Developing a near-term problem statement with BoE.
- [AP2.2] Understanding problems of strategic importance to BoE, and how we can maximise impact.

We have engaged in a number of ways with BoE, including:

- Internal presentations to BoE staff to share our research, identify problems of interest, and get feedback on our plans.
- Focussed scoping meetings to identify problems of interest and find internal collaborators.
- Collaboration meetings with pre-existing collaborators.
- Reviews of detailed objectives for the past year.

The result of this has been:

- Feedback on our approaches to explainable AI, guiding the development and plans for future research and demonstrators
- Guidance on focus for input data-sets and prediction tasks, for example switching from FRED to ONS data sources for econometric data, and changing target predictions from monthly to quarterly.
- Identification of integration points into existing BoE tools, for example ongoing work on integration into the Forecast Evaluation tool (https://github.com/bank-of-england/forecast_evaluation).
- A new regular weekly 30-min meetings between BoE and SONNETS to continue collaborative work started in Y2Demo-Explainability (presented in the next section).
- Identification of a new demonstrator project on ABM calibration, which is now underway and is already producing results in Y2Demo-ABMCalibration (presented later in this chapter).

4.2. Demonstrator 1: BoE explainable demonstrator

A core motivation of the SONNETS project, and this demonstrator, is interpretability and explainability. Unlike traditional opaque black-box models, Tsetlin Machines (TMs) produce predictions through human-readable logical clauses which, paired with dedicated interpretability visualisations, make the reasoning behind a forecast fully inspectable. A piece of work called TMAtlas ([pattison-2025-tmatlas]) provides some initial technical methods for how to answer this, but there is still a large gap in making explanations accessible to end users.

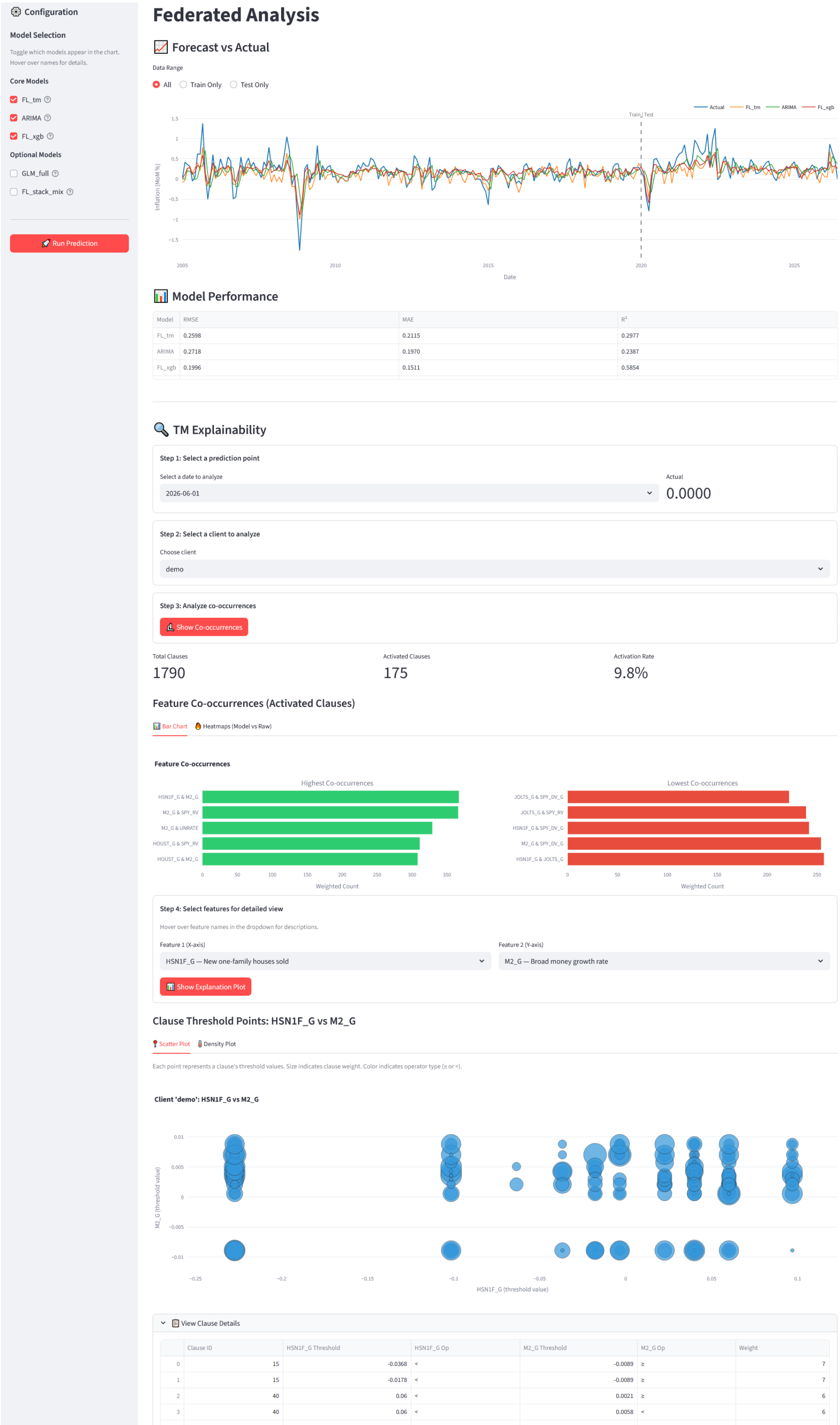


Figure 1. Explainable Risk Demonstrator

For an interactive demo visit: http://frtm.demos.sonnets-project.org/
NOTE: The demonstrator currently does not use HTTPS, so you may need to accept security warnings in your browser.

This demonstrator was developed in preparation for a presentation given at BoE in Feb 2026, with the goal of demonstrating how underlying research such as TMAtlas could be applied in the context of risk analysis. It has since been further improved, and turned into a reuseable demonstrator.

The demonstrator follows a number of steps:

1. **Model creation** : illustrates that the model is built using user input and choices.
2. **Forecasting** : shows explainable predictions versus classical statistical and existing ML approaches.
3. **Error analysis** : quantitative comparison of models on different metrics, showing explainability does not reduce accuracy compared to non-explainable models.
4. **Interactive explainability** : allows users to drill down into *why* predictions are made
    a. Users can select a particular timepoint to select a prediction
    b. The co-occurrence of features is presented to give an initial overview of how features interact
    c. Users can drill down further into pairs of features to understand the relationship to training data

The demonstrator serves three roles:

1. a domain-specific demonstration of the underlying explainability ideas from TMAtlas.
2. an interactive guide for domain experts through the steps of analysis and explanation.
3. a discussion point around the idea of "explainability" to drive further collaboration.

4.3. Demonstrator 2: ABM calibration

During scoping meetings between BoE and SONNETS we discussed various internal risk models and systems which could act as baselines for SONNETS demonstrators and research. We particularly tried to identify systems which BoE is interested in, but were currently limited by computational, operational, or deployment issues. A use case identified was the application of Agent-Based Models for modelling and understanding financial systems, in particular a model for simulating participants in the UK housing market. Essentially it tries to model each person in the UK as an agent, then simulates them as they finance, buy, and sell properties, with individual agent behaviours then aggregated to understand the overall market. This worked well as a case study because:

1. The model has already been used by BoE, so it is a real-world model of interest.
2. It supports explainable risk, as it not only provides risk estimates, but also allows exploration of where those estimates come from in terms of individual actor behaviors.
3. The ABM approach enables injections of scenarios and shocks to explore the impact of different events.

There are also a good set of research challenges, including:

- Calibration: There are a huge number of model parameters, some of which are observed in market and econometric data, some of which must be inferred. Previously calibration has been manual, and is a slow process.
- Computation: Agent-based models simulate market participants (humans) over time, and so require significant computational resources.
- Explanation: While all actor behaviours are known, there is still an explainability gap between the high-level econometric outputs, and why the underlying behaviours caused them.

The calibration problem was an initial hurdle addressed by the demonstrator, which included:

1. Identifying and updating model parameters from open data sources, including ONS and BoE data.
2. Inferring or imputing calculated parameters.

3. Rebuilding and updating the model to better support calibration.

This work has been presented back to BoE on two occasions to different groups, and is the subject of ongoing research including:

1. Making the calibration fully automatic based on live data.
2. Optimising the model computationally to support larger numbers of actors.
3. Embedding the ABM into the ETC framework to support real-time and interactive analysis.

4.4. Demonstrator 3: CVAE for financial time series

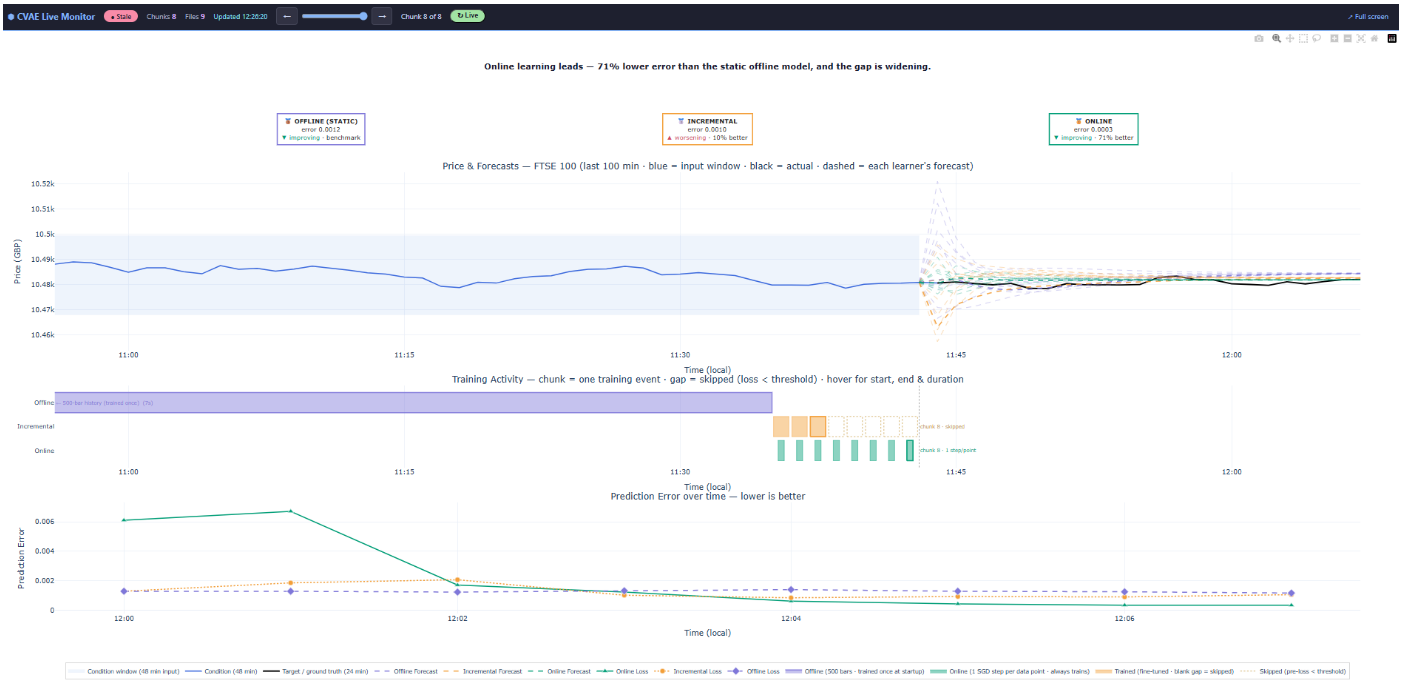


Figure 2. Real-time CVAE risk analysis and forecasting demonstrator

At the previous Advisory Board we presented a number of demonstrators related to learning and analysis, but these were all traditional academic demonstrators using offline historical data and offline learning. We also demonstrated the event-triggered-computing abstraction we had built, but only in a limited local fashion. Key objectives from the last Advisory Board were around bringing these together to provide true real-time risk analysis, including: ingesting live data ([Y2G1],[Y2D1]); daily risk updates and display ([Y2G2],[Y2G3]); and autonomous re-analysis ([Y2G5]). We also wanted to integrate the ongoing research into novel time-series predictions using conditional auto-variational encoders (CVAE), including different methods for continual and occasional training. The resulting demonstrator combines the following features:

1. A web-based interface providing forecasts of FTSE100 price movements using multiple models.
2. Live market updates at multiple frequencies, including minutely, hourly, and daily ([Y2D1]).
3. Multiple real-time training modes for the model, including offline (train once), online (train for every data-point), and incremental (train once error exceeds threshold).
4. Back-testing of accuracy across different models, allowing views of historical predictions ([Y2R4]).
5. Deployed using a combination of cloud and local compute resources ([Y2G7]).
6. The system has run continuously over live data for more than a week at a time ([Y2G4]).

The underlying CVAE model was chosen as it is a novel academic machine-learning model with no easy library wrappers requiring a complex software eco-system, so it presented validation of whether the ETC approach could be used in practice. During development of the demonstrator we were able to take the same ML model from offline development on a laptop through to production deployment with adaptive learning in the cloud using the same codebase, providing confidence in the technique.

A limitation of the current demonstrator is that:

1. Predictions are relatively poor, as input data sources are fixed and the models are manually tuned.
2. The CVAE method provides uncertainty estimates, but it is not explainable.

Over the next year we plan to create new demonstrators that are able to autonomously select data sources and re-parameterise and tune the learners.

5. Research highlights

The demonstrators are used to unify and drive the research project, but they are all based on top of fundamental and applied research. Some of this work directly underpins the current demonstrators, some is exploratory research expected to feed into the overall project demonstrator, and there is also spin-off research as part of the generalisation work-package. Overall there have been 32 peer-reviewed SONNETS publications so far, with 26 of those published since the last AB and five of them receiving Best Paper Awards. Here we will highlight some of the research areas and themes, and highlight some specific exemplar papers.

We are trying to provide a brief summary here, so we do not attempt to cover all published papers here, nor do we include research in progress which has not been published. So for example there are no outputs from the Event Triggered Computing work-package, which has been focussed on supporting the demonstrators (though it does have two papers under development). To provide a clear delta between ABs, we also do not include papers which are under development, under review, or accepted but not yet published as of the writing of this AB report.

5.1. Explainable learning

One of the input technologies to the SONNETS project was the emerging field of Tsetlin Machines (TMs) as a novel type of learner. A very brief overview is that TMs are a new type of machine learning model which learn models using clauses, where each clause is a logical conjunction of multiple binary input features. When classifying new data, the clauses are evaluated and combined to produce a classification. They have been shown to be competitive with deep neural networks on a number of tasks, but with the added benefit of being interpretable because the internal clauses can be read as human-readable logic statements. A more detailed presentation of TMs can be found in the previous advisory board report (https://sonnets-project.org/documents/), but in the interests of space we do not try to explain them here.

A central goal of the SONNETS project is to use TMs to produce *explainable* risk models, which requires us to answer some key questions TMs including:

1. How do we extract explanations from trained models?
2. How do we present explanations to users?

Some key publications which try to attack this problem are:

[pattison-2025-tmatlas] : "*TMAtlas: An Interactive Visual Analytics Framework for Explaining Tsetlin Machine Outputs*"
This work introduced a novel framework called TMAtlas that provided interactive visual analytics to explain the outputs of TMs by maintaining traceable mappings between raw and Boolean-encoded features, and enabling structured clause analysis through a graph-based representation of a learned TM model. This work has been particularly transformative within SONNETS, feeding into Y2Demo-Explainability, and then on into an ongoing collaboration with BoE and planned upcoming demonstrator ([Y3Demo-Forecaster]) .

[tarasyuk-2026-tm-learning-dynamics] : "*Learning dynamics, pattern recognition capability and interpretability of the Tsetlin Machine*"
This work explores how TMs actually learn, particularly focusing on the TM's mechanism for constructing conjunctive clauses from training data. It provides new insight into how the TM internal state changes over time, and in particular how this might allow for better interpretability of the output classifications.

[xu-2025-recurrent-tm] : "*Recurrent Tsetlin Machine for Sequence Learning*"
This work proposes a new class of TMs called Recurrent TMs (RTMs) by extending the original TMs to support sequential learning using clause outputs from previous time steps as the recurrent state. This extension is particularly important for time-series analysis, and results in substantially better accuracy of 95% compared to existing TM machines at 57%.

[erkus-2025-latentgraph] : "*LatentGraph: A Tsetlin Machine-Based Graph Learning Framework for Interpretable Feature Extraction*"
This work introduces a new explainability pipeline for financial forecasting called LatentGraph, where multivariate data undergoes subsequent transformations from latent factor encodings all the way to TMs for simple IF-THEN rule explanations. This work is particularly important for handling financial prediction problems with large numbers of input time-series and data-sets, as it retains accuracy compared to classical approaches while adding end-to-end transparency about the decisions and classifications.

We also use the flexibility of the programme grant to consider other options, beyond TMs. One example is the CVAE approach explored in Y2Demo-CVAEForecast (which has a publication in development), but other explainable approaches have also been explored, for example:

[faloughi-2025-protohedge] : "*ProtoHedge: Interpretable Hedging with Market Prototypes*"
To avoid the black-box nature of deep neural networks, this work instead tries to create interpretable hedging strategies using the idea of "prototypes", which are representative market states. The model then hedges on the basis of these prototypes, providing an auditable trail of hedging decisions in terms of prototypes, while reaching within 0.4% of the performance of a black-box DNN. A similar approach, known as ProtoPGTn, has been developed to support interpretable time series classification [huang-2025-protopgtn].

5.2. Financial modelling and analysis

The headline driver for SONNETS is financial risk analysis, and particularly the idea of a real-time digital twin for the financial system. This cannot rely purely on machine learning, so we also consider fundamental research in the financial domain.

[leung-2025-fabs-digital-twin] : "*FABS: An Extensible and High-Performance Digital Twin Framework of AI-Driven Financial Systems*"
Agent-based models are an important tool for financial risk analysis (as discussed in Y2Demo-ABMCalibration), but they are often very slow. This work presents a new framework for building agent-based models in C++ which is designed for specifically for high performance financial simulations, achieving a 12x speed-up over existing techniques.

[jiao-2025-scalable-causal] : "*Scalable Time Series Causal Discovery with Approximate Causal Ordering*"
Causal graph analysis was a work in progress demonstrated in the first Advisory Board, and it has since been developed and published. The particular contribution here is the ability to scale to very large numbers of input time series, and is expected to be a building block used later to support autonomous learning and data-source selection.

5.3. Generalisation : Medicine

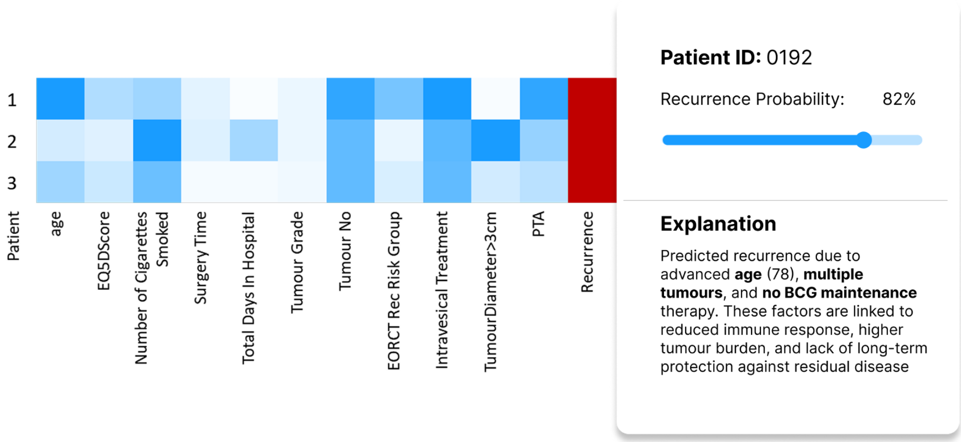
Part of the goal of SONNETS is to identify and explore other areas where the research could have impact, as part of the "Generalisation" work-package. This also relates to action point [AP3.6], around generalising impact beyond regulatory impact. An offshoot of the explainable AI research has been a sequence of work into medical applications, particularly for supporting cancer diagnosis:

[abbas-2025-ai-predicting], "*AI predicting recurrence in non-muscle-invasive bladder cancer: systematic review with study strengths and weaknesses*"
A comprehensive systematic review mapping the state of AI for bladder cancer recurrence prediction, critically evaluating each study's methodology, statistical robustness, and real-world limitations. Despite some ML models reaching 97.5% accuracy, doctors are still hesitant to implement them into routine clinical use – this review is the first to explain why.

[abbas-2025-bladder-cancer], "*Explainable AI for Bladder Cancer Recurrence Prediction*"
This work introduces an attention-based deep learning framework for NMIBC recurrence prediction that not only outperforms conventional statistical methods but discovers new factors for predicting recurrence which were overlooked by every prior model in the field. Crucially, it delivers these insights at the individual patient level, showing exactly which features drove each prediction.

[abbas-2025-clinical-tm], "*AI-Based Clinical Rule Discovery for NMIBC Recurrence using Tsetlin Machines*"
This study applies the Tsetlin Machine to 330 NHS patients from the PHOTO trial, outperforming XGBoost, Logistic Regression, and the EORTC risk tables across precision, recall, and F1-score. With an F1-score of 0.80 and predictions expressed as plain clinical rules, doctors can confidently identify exact rules that predict a patient's recurrence - something no other ML model can do currently.

This sequence of research has produced new tools for treatment of cancer, such as patient scoring tools with automatic prediction:



We plan to further develop these tools and put them in the hands of NHS practitioners in the future ([Y3O8-BCR]).

5.4. Generalisation : low-latency physics

Financial risk-analysis is intended to be real-time in the sense of day-by-day and potentially hour-by-hour, but there are also applications with much lower latency applications. A fruitful generalisation has been in the area of low-latency physics reaching down to the sub-microsecond level.

[que-2025-jedi-linear], "*JEDI-linear: Fast and Efficient Graph Neural Networks for Jet Tagging on FPGAs*"
This work considers the use of graph-neural networks in FPGAS and how to perform low-level optimisation of architectures and algorithms to minimise classification latency. This approach is not only smaller than existing approaches, but is also the first approach to achieve latencies below 60ns.

[laatu-2025-sub-microsec-transformers], "*Sub-microsecond Transformers for Jet Tagging on FPGAs, ML and the Physical Sciences*"
The transformer is a well known architecture for building large-language models, but is difficult to apply at very short latencies due to the cost of attention. This work presents the first transform operating at sub-microsecond levels using FPGAs, and has also been incorporated into the open source hls4ml framework.

[sun-2026-hgq-lut], "*HGQ-LUT: Fast LUT-Aware Training and Efficient Architectures for DNN Inference*"
The previous two works focussed on inference latency for a given model, but there is also the challenge of creating models for ultra-low latency engines with highly quantised parameters. This work addresses the problem by creating LUT-aware training approaches which taking into account the underlying FPGA hardware, and has been used to create LUT-aware DNNs for use in the LHC at CERN.

6. Future Plans

Key feedback from the previous AB was to [AP1.3] establish final goals at this Advisory Board, [AP4.2] have more of a focus on goals, [AP4.3] provide a clear vision for intended project deliverables. In this section we list measurable objectives for the final project, and then provide a set of planned goals and operational plans. We also include explicit areas where we would appreciate feedback, in addition to any other feedback the AB has.

6.1. Final project demonstrator

Action point [AP1.3] from the previous AB was to establish final overall project goals for the overall project at this Advisory Board, and [AP4.3] requested a clear vision on which the AB could provide feedback. In this section we summarise the expected outcomes in terms of functionality of the final project demonstrator, and identify different levels of functionality to support prioritisation.

Defining exactly what will be delivered at the end of the project is still difficult due to the ongoing research, but a relatively clear set of high-level requirements has emerged. Given the variety of approaches to risk analysis, data-sources, and analysis time-scales the demonstrator is expected to be a dashboard with a family of inter-related views - for example the interactive historical risk drill down shown in Y2Demo-Explainability and live views of evolving risk as in Y2Demo-CVAEForecast. We also expect to integrate views supporting injection of scenarios - for example there is ongoing work on classifying and injecting shocks interactively:

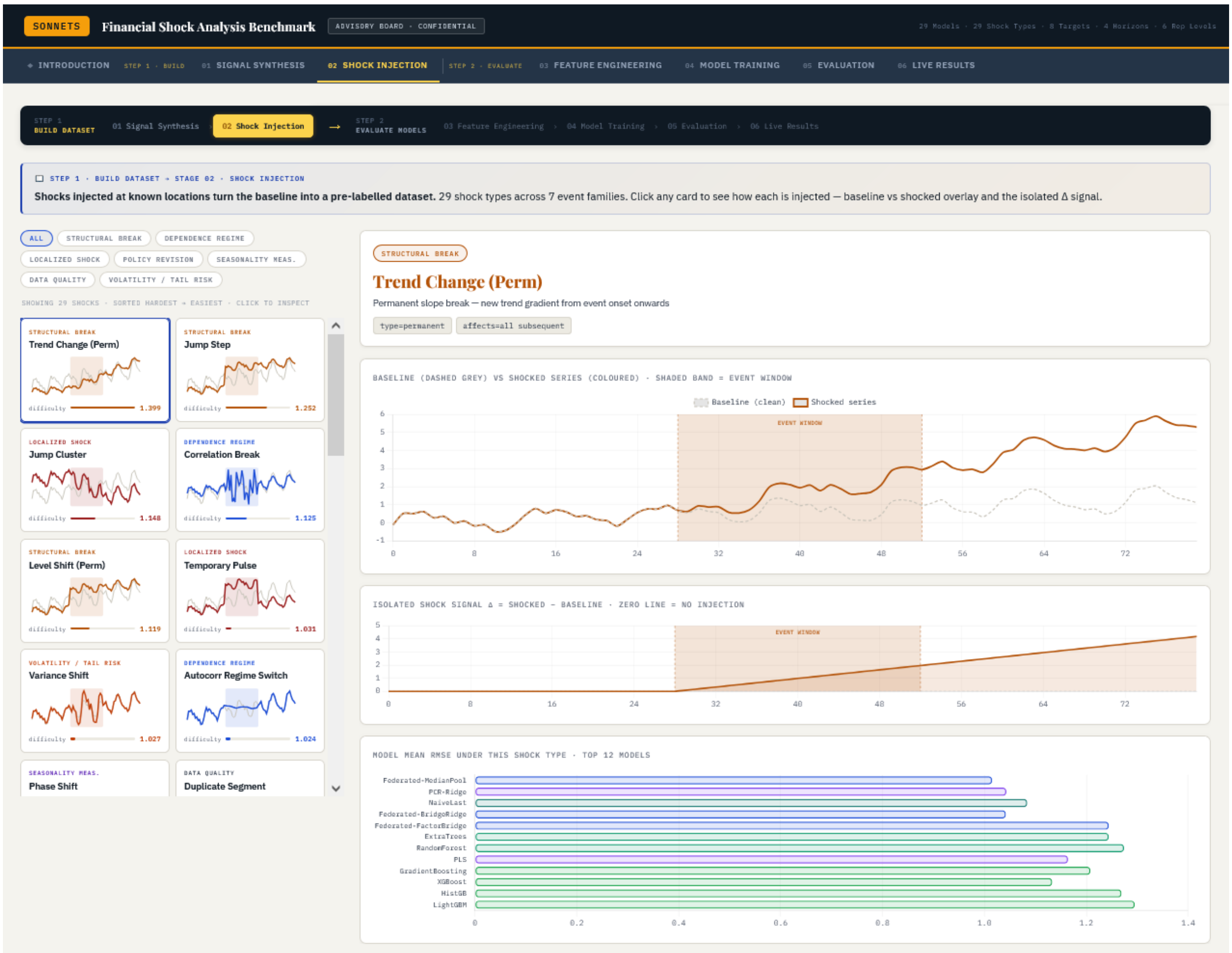


Figure 3. Prototype user-interface for user-controlled shock injection

Our overall project deliverables are rooted in the original project goals, but refined through interactions with stakeholders, demonstrator development, and the fundamental and applied research we have been performing. In defining these goals we focus on demonstrating new capabilities enabled by the research and avoiding capabilities already provided by existing systems. Those key capabilities of interest identified in combination with stakeholders are broadly:

- Explainability : systems that are able to assess risk while also explaining their reasoning
- Interactivity : ability to interact and ask "what if?" questions, getting answers *and explanations* back.

- Automatic calibration : updating and recalibrating models in real-time as new data arrives, without human intervention, but with explanations of what has changed and why.

Despite the uncertainty, we are able to list key capabilities or functions that the final demonstrator is expected to show, plus four different levels ranging from 1 (already achieved in some aspect of the project) to 4 (requires cutting-edge research).

- **PG-1RF : Risk forecasting** : there are different levels of what a risk forecast is, from a single path for a single variable upwards.
  1. Simple forecasting of key financial variable values : achieved Y2Demo-CVAEForecast
  2. Estimating confidence intervals : partial Y2Demo-CVAEForecast
  3. Estimating probability distributions of variables
  4. Joint probability distributions over multiple variables
- **PG-2LA : Live Analysis** : genuine forecasts into the future are clearly critical in practice, but often academic research focusses on forecasting historical results for convenience.
  1. One-step-ahead forecasting : achieved Y2Demo-CVAEForecast
  2. Efficient always-on forecasting
  3. On-demand horizontal and vertical scaling of forecasting
  4. Accelerated forecasting to support "what if?" scenarios
- **PG-3IE : Interactive Explanations** : a challenging question is "what is an explanation?", reinforced by the experience of presenting current explainable work to potential users.
  1. Static explanations of the current forecasts : achieved Y2Demo-Explainability
  2. On-demand static explanations of back-tested forecasts
  3. Interactive dynamic expansions of explanations
  4. Post-hoc explanations of non-explainable models
- **PG-4CB : Calibration** : a drawback of existing approaches is the need to manually update and recalibrate models, so new data needs to be integrated to automatically update and recalibrate models in real-time.
  1. Static calibration of models : achieved Y2Demo-ABMCalibration
  2. Periodic re-learning : achieved Y2Demo-CVAEForecast
  3. Event-triggered re-learning : partial Y2Demo-CVAEForecast
  4. Anticipatory learning
- **PG-5DS : Data Selection** : There are hundreds of data sources containing thousands of variables. Current risk analysis approaches use fixed sets of data sources, either chosen by humans, or selected statically at training time. Trying to look at all data sources at maximal temporal resolution is computationally infeasible and risks over-fitting.
  1. Static selection of data sources : achieved Y2Demo-Explainability, Y2Demo-CVAEForecast
  2. Optimisation of data sources through repeated training
  3. Analytical selection of data sources, e.g. using causal analysis
  4. Co-optimisation of data source selection during learning
- **PG-6SC : Scenarios** : while back-testing allows historical shocks to be examined, a key question from both a policy and outreach point of view is "What if X happened right now"?

1. Manual historical scenario injection
2. User-defined scenario injection
3. Automated historical scenario injection
4. Autonomous creation of scenarios

- **PG-7DM : Diverse Models** : there are a large set of new analysis and learning methods being developed in the project, but there are also many useful classical methods available, as well as other new domain-specific and foundation models. Integrating multiple methods makes comparisons more robust, and opens up the possibility of hybrid models.
  1. Integration of individual explainable models : achieved Y2Demo-Explainability
  2. Multiple simultaneous internal and classical models
  3. Rapid integration of external academic models
  4. Hybrid explainable models using multiple component models

- **PG-8CF : Confidentiality** : there is a need for a public-facing systems to provide transparency and public engagement, but also private versions that can incorporate proprietary data.
  1. Public data only presented without authentication : achieved Y2Demo-CVAEForecast
  2. Private data presented with authentication and access control
  3. Confidential data processed in secure enclaves
  4. Tainting of models to control access to explanations

- **PG-9UI : Web Interface** : stakeholder engagement has highlighted the importance of user interfaces to ensure impact on target end-users, and a web-based interface appears by far the most appropriate compared to demonstrators using local GUIs. In terms of a research project there is then an engineering question of how to balance engineering time:
  1. Research quality code and UI : achieved Y2Demo-CVAEForecast, Y2Demo-Explainability
  2. Open-source package deployable by experts
  3. Highly polished and unit tested UI
  4. Commercially deployable UI

This combination of functionality is distinctly different to what is currently available:

- Academic research produces offline models which are difficult to adapt for real-world use
- Regulatory tools are not public facing, and contain conservative models
- Commercial tools do not focus on explainability and do not consider national risk

In terms of the underlying and internal technologies it is difficult to pick winners at this point, but examples of the types of internal research come from the existing demonstrators and fundamental research. We avoid over-specifying the technologies — e.g. "*we will forecast 1 quarter ahead CPI deciles using SVA and Tsetlin machines*" — as we need to pivot in response to the research and stakeholder feedback.

6.2. Generalisation

There are two directions for generalisation: cancer risk prediction and high-energy physics.

The first direction is an emerging theme from the explainable financial risk work in risk prediction in medical contexts. The existing research in bladder cancer risk demonstrates promise, and is an excellent candidate for generalisation of the SONNETS work.

Bladder cancer takes one life every hour in the UK, yet survival rates have not meaningfully improved in over four decades, and current clinical scoring tools for assessing risk are outdated and unreliable. The preliminary work in Section 5.3 has established the promise of this work, but a key barrier to adoption is trust: do clinicians believe the predictions of AI?

As part of the generalisation package Kabita Adhikari will lead work on explainable AI for bladder cancer risk prediction, with two main goals:

1. Augmenting clinical decision-making by providing trusted personalised models
2. Replacing legacy clinical scoring tools with more accurate yet low cost explainable models

In terms of allocation of resources, the plans are:

- Staff costs : RA funding to be allocated to Saram Abbas, who is currently finishing their PhD. Underspent DI staff costs from Southampton to Newcastle to support this, taking advantage of flexible funding of the programme grant.
- Data acquisition costs : existing funds intended for financial data to also be used for medical data.

The second direction concerns scientific application. This research would build on our collaboration with institutions such as CERN and Caltech (described in Section 5.4) to support advances for high-energy physics experiments, enabling data analysis with low latency and high throughput. The work will be led by Wayne Luk, with two main goals:

1. Study hardware structures that support latency-critical data processing such as boosted decision trees.
2. Develop tools for automating the development of such hardware structures.

In terms of allocation of resources, the plans are:

- Staff costs : RA funding to be allocated to Terry Wang, who is currently finishing their PhD.
- Travel costs: support visits by researchers from CERN, Caltech and Bristol University to strengthen this collaboration.

6.3. Operational plans

We now identify a set of concrete objectives for the next year based on our current goals, framed in terms of research goals, plus demonstrators that integrate and use them. This does not constrain the underlying research, but is intended to make sure it is building towards the overall project goals.

Each objective is tagged as "Must", "Should", or "Could." to reflect the SONNETS team's view on priorities, but we would welcome feedback on these priorities from the AB.

6.3.1. Research objectives

- **Y3O1-DEX : Distributed explainable learning** : the Y2Demo-Explainability and Y2Demo-CVAEForecast demonstrators use a variety of learning and re-learning techniques, including federated learning, but a learning step is still a single computational task. A key challenge is to

distribute the learning in both space and time, in order to scale across multiple resources and to respond to changes arriving from different data sources.
Priority: Must. *Supports* : [PG-1RF], [PG-2LA], [PG-3IE].

- **Y3O2-SHK : Analysis of injected shocks into live scenarios** : Priority shocks are rare but important in risk analysis, so we need to be able to inject plausible shocks for testing purposes, as well as for interactive what-if analyses. Shocks are an important research challenge here due to the interaction with the explainability aspect of risk analysis.
Priority: Should. *Supports* : [PG-6SC], [PG-4CB], [PG-3IE].
- **Y3O3-AC : Autonomous calibration of agent-based models** : Y2Demo-ABMCalibration has identified ABMs as a useful method for exploring risk scenarios, but it is still manually calibrated. Automatically calibrating ABMs is a challenge but, it fits very well within the desired system where incoming data is used to update risk models, and later can be used to model injected shock impact.
Priority: Should. *Supports* : [PG-4CB], [PG-5DS].
- **Y3O4-SC : Automated scaling of compute resources** : the current underlying ETC layer is able to support distributed event-triggered compute at multiple scales, from a laptop to a cluster, but uses a fixed set of resources. In order to respond to market events without excessive cost, the system will need to automatically horizontally and vertically scale, using either local cluster management, or cloud provisioning.
Priority: Should. *Supports* : [PG-2LA], [PG-4CB].
- **Y3O5-UDI : Unconventional data ingestion** : This is a continuation of the [Y2D5] objective from the previous year, which was not explored. Data sources of particular interest include textual data such as news and social media, but might include alternative markets such as prediction markets, with a key challenge being how to incorporate them into existing or new risk forecasting frameworks.
Priority: Must. *Supports* : [PG-5DS], [PG-7DM].
- **Y3O6-HO : Hardware optimisation of models** : The TM approach has proven effective in Y2Demo-Explainability and in a large set of research outputs, but like all learning methods it slows down substantially with more data and bigger modes. Another feature of TMs is that clauses are evaluated using boolean logic, meaning there are substantial cost savings and latency reductions if specialised hardware is used for learning.
Priority: Should. *Supports* : [PG-2LA], [PG-3IE], [PG-7DM].
- **Y3O7-AO : Optimisation of ABM** : the ABM from Y2Demo-ABMCalibration is relatively slow, which limits the ability to do very large-scale simulations, though it can be scaled out to perform parallel simulations. Traditional optimisation could improve performance by 10x and would directly support BoE use of the model, but is less aligned with near term demonstrators.
Priority: Could. *Supports* : [PG-4CB], [PG-2LA].
- **Y3O8-BCR : Support tool for bladder cancer risk prediction** : develop methods and models for bladder cancer risk prediction using large-scale data-sets, incorporating explainability in a form acceptable for NHS practitioners.
Priority: Should. *Supports* : independent of main financial risk demonstrator.
- **Y3O9-SE : Broadening stakeholder engagement** : over the last year the project has heavily focussed on engagement with BoE stakeholders in order to identify problems of interest and to ensure impact. Over the next year we will broaden this engagement to include other stakeholders (e.g. NHS) and other potential commercial financial uses.
Priority: Should. *Supports* : independent of main financial risk demonstrator.
- **Y310-PHY : low latency data processing for high-energy physics** : develop methods and tools for low latency data processing that targets the latest high-energy physics experiments at CERN. *Priority*: Should. *Supports* : independent of main financial risk demonstrator.

6.3.2. Demonstrators

- **Y3Demo-Forecaster : Integration into existing BoE tools** : the work on Y2Demo-Explainability has led to an ongoing collaboration to integrate explainability into an existing BoE Forecasting tool (https://github.com/bank-of-england/forecast_evaluation). An initial example of this interface can be seen in Figure 4 (next page).
  Priority: Should. *Supports* : [PG-1RF], [PG-3IE], [PG-7DM].
- **Y3Demo-AutoExplain : Autonomous Explainable Prediction** : This demonstrator will integrate the existing work from Y2Demo-Explainability and Y2Demo-CVAEForecast into a single system, while also incorporating ongoing research, in particular [Y3O1-DEX], [Y3O4-SC], and [Y3O6-HO], though we will seek to incorporate other ongoing research.
  Priority: Should. *Supports* : [PG-1RF], [PG-2LA], [PG-3IE], [PG-4CB], [PG-5DS], [PG-6SC], [PG-7DM].
- **Y3Demo-CancerPred : Cancer Risk Prediction Tool** : This will be a prototype tool intended for use by NHS practitioners, incorporating the existing cancer risk research outputs, as well as the ongoing research from [Y3O8-BCR].
  Priority: Should. *Supports* : independent of main financial risk demonstrator.

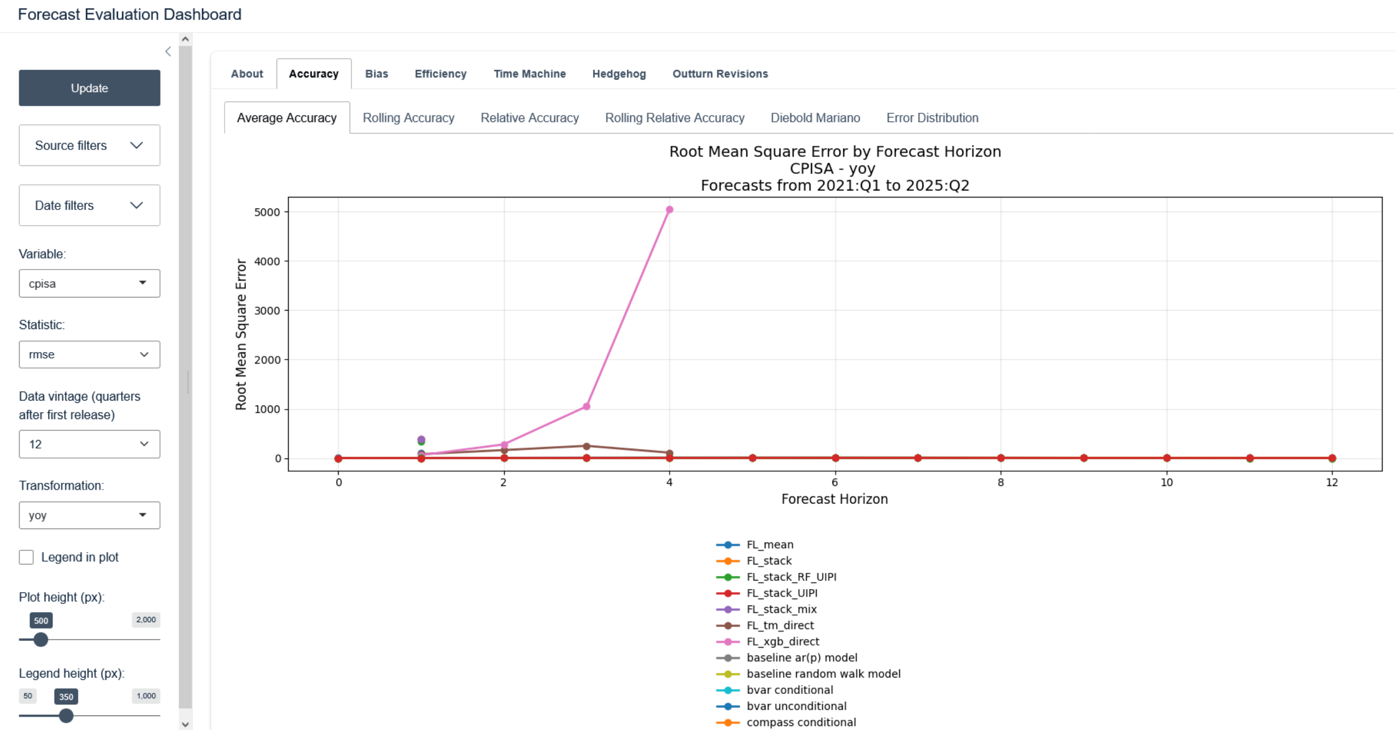


Figure 4. Prototype integration of explainable forecasting into BoE forecasting tool ([Y3Demo-Forecaster])

6.3.3. Development priorities and productisation

A current operational question is around fundamental research vs product development. SONNETS is a research project, but it also has a clear goal of creating demonstrators that will have impact. A recurring theme from the last year has been balancing research vs development, particularly in terms of acquiring skills and knowledge for modern front-end and back-end cloud development. As academics we know how to create demos, but creating polished interactive web user interfaces and cloud systems is much more time consuming.

The SONNETS project originally assumed that research staff would be traditional research fellows, but in terms of impact it has become clear that we need to re-train substantially. This has become particularly important as the need for interactive exploration of risk has been highlighted as an important capability of interest. Contemporary AI tools have helped in terms of prototyping UIs, but the resulting code is often fragile and breaks outside the test environment.

A possible solution is to re-purpose some existing budget to explicitly include engineers rather than pure researchers. Southampton in particular has sufficient DI staff funds to support some combination of:

1. A front-end engineer responsible for developing and maintaining user interfaces, who is able to work with research teams to develop polished demonstrators.
2. A back-end and packaging engineer who is responsible for taking all the research outputs and front ends and creating fully maintainable systems.
3. Paid time for Research Software Engineers available within the university.
4. Engagement of professional consultants driven by need.

This would also support activities around IP management and commercialisation, as it would make it more possible to package up the research and identify what a "product" might look like. Currently our IP management strategy has been to publish papers and open-source code - this currently works well for engagement with BoE due to their approach to open-source. Much of the potentially exploitable IP arising from the project is in the form of codebases and implementations rather than patentable IP, so both commercially exploit arising IP or maximising public-domain IP use suggests a stronger engineering component of staff.

We would value the AB's opinion on whether this would be a useful re-purposing of the existing budget. Given the ability to recruit two more research fellows, we could focus on:

1. Recruit as normal: hire two research fellows based on research skills and experience
2. Balanced: hire one research fellow and one engineer with contemporary practical web development experience
3. Recruit for products: hire a front-end engineer and a back-end engineer

These options also have implications for the overall project deliverables, as the first option would target more fundamental research and papers, while the third leans much more towards product delivery and IP development. Our current view is option 2, but we would value the AB's opinion on this, particularly as options 2 and 3 would require more careful negotiations with university HR.

6.4. No-cost extension

A goal of the project has always been to build links with stakeholders, and for this project we particularly wanted to try to address regulatory and governmental stakeholders. Over the last year in particular we have built a relationship and development engagement, but this has taken longer than expected. This time has not been wasted, in that a great deal of contributory research has been completed, but it does mean we are further behind than planned on progress towards the overall national financial risk demonstrator.

At the same time we are underspent on budget, particularly on DI staff costs, which presents an opportunity. We already have a good team of researchers and academics in place who are working well together, and ongoing collaborations both within the project and with external stakeholders. A one-year no-cost extension is a good opportunity to maximise impact. This is not simply a case of "spending all the money", but rather a chance to consolidate the research and demonstrators, and to ensure that the final deliverables are of high quality and have impact. For the research vs product question it also allows for a better balance, as it means more initial research can be completed before focussing heavily on the final demonstrators and products.

6.5. Request for feedback

The SONNETS team would appreciate feedback from the AB on the following:

- High-level objectives Section 6.1 for the project.
- Objectives and priorities in Section 6.3 for the coming year.
- Balance between research and productisation, in terms of objectives and staffing.
- The case for a no-cost extension to maximise the achievement and impact of the project.

Our plan is to prepare a post-AB meeting report which takes into account the feedback, similar to the more ad-hoc document after the first AB, so that there is then a clear record of updated objectives to review at the next AB.

SONNETS publications

Publications since the last Advisory Board have the date highlighted in bold.

- [chan-2026-animate] Alex Chan, Moh Tarraf, Rishad Shafik and Alex Yakovlev, "ANIMATE: Automated Framework for Scalable Design of Tsetlin Machines Using 1-Safe Petri Nets", Proc. Application and Theory of Petri Nets and Concurrency, pp. 130-153, **Jun 2026**.
- [sun-2026-hgq-rt] Chang Sun, Zhiqiang Que, Thea K. Årrestad, Vladimir Loncar, Jennifer Ngadiuba, Wayne Luk, Maria Spiropulu, "*HGQ: High Granularity Quantization for Real-time Neural Networks on FPGAs*", FPGA, **Feb 2026**.
- [tarasyuk-2026-tm-learning-dynamics] Olga Tarasyuk, Anatoliy Gorbenko, Tousif Rahman, Lei Jiao, Ole-Christoffer Granmo, Rishad Shafik, Alex Yakovlev, "*Learning dynamics, pattern recognition capability and interpretability of the Tsetlin Machine*", Pattern Recognition, Volume 174, **June 2026**.
- [ito-2026-trustworthy-nn] Hiroaki Ito, Hikari Otsuka, Ryota Yasudo, Zhiqiang Que, Jose G. F. Coutinho, Daichi Fujiki, Masato Motomura, Ce Guo, Wayne Luk, "*Memory-Efficient and Trustworthy Neural Networks via Random Seed-Based Design*", IEEE Access, Vol 14, **Jan 2026**.
- [sun-2026-hgq-lut] (**Best Paper Award**) Chang Sun, Zhiqiang Que, Bakhtiar Zadeh, Qibin Liu, Kevin H. Alvarez, Wayne Luk, Maria Spiropulu, "*HGQ-LUT: Fast LUT-Aware Training and Efficient Architectures for DNN Inference*", IEEE Symposium on Field-Programmable Custom Computing Machines, **May 2026**.
- [que-2025-jedi-linear] (**Best Paper Award**) Zhiqiang Que, Chang Sun, Sudarshan Paramesvaran, Emyr Clement, Katerina Karakoulaki, Christopher Brown, Lauri Laatu, Arianna Cox, Alexander Tapper, Wayne Luk, Maria Spiropulu, "*JEDI-linear: Fast and Efficient Graph Neural Networks for Jet Tagging on FPGAs*", FPT, **Dec 2025**.
- [huang-2025-protopgtn] J. Huang, C. Guo, W. Luk, "*ProtoPGTN: A Scalable Prototype-Based Gated Transformer Network for Interpretable Time Series Classification*". Information, 16(12): 1056, **Dec 2025**.
- [leung-2025-fabs-digital-twin] Angus T. L. Leung, Ce Guo, and Wayne Luk, "*FABS: An Extensible and High-Performance Digital Twin Framework of AI-Driven Financial Systems*", Proc. ACM International Conference on AI in Finance, 27-34, **Nov 2025**.
- [laatu-2025-sub-microsec-transformers] Lauri Laatu, Chang Sun, Arianna Cox, Abhijith Gandrakota, Benedikt Maier, Jennifer Ngadiuba, Zhiqiang Que, Wayne Luk, Maria Spiropulu, and Alexander Tapper, "*Sub-microsecond Transformers for Jet Tagging on FPGAs, ML and the Physical Sciences*", NeurIPS 2025 Workshop, **Nov 2025**.
- [pattison-2025-tmatlas] (**Best Student Paper Award**) Bob Pattison, Tousif Rahman, Alex Chan, Ekin Can Erkuş, Kabita Adhikari, Ole-Christoffer Granmo, David Thomas, Alex Yakovlev, Rishad Shafik, "*TMAtlas: An Interactive Visual Analytics Framework for Explaining Tsetlin Machine Outputs*", Proceedings ISTM, **Oct 2025**.
- [erkus-2025-latentgraph] Ekin Can Erkus, Alex Chan, Walter Distaso, David Thomas, Alex Yakovlev, Rishad Shafik, "*LatentGraph: From Latent States to Rule-based Expressions for Explainable Financial Forecasting*", Proc. AI in Finance, pp 745-752, **Nov 2025**.
- [xu-2025-recurrent-tm] C. Xu, S. Duan, R. Shafik and A. Yakovlev, "*Recurrent Tsetlin Machine for Sequence Learning*" 2025 International Symposium on the Tsetlin Machine (ISTM), pp. 102-106, **Oct 2025**.
- [zeng-2025-fast-compact-tm] Y. Zeng, S. Duan, R. Shafik and A. Yakovlev, "*Fast and Compact Tsetlin Machine Inference on CPUs Using Instruction-Level Optimization*" 2025 International Symposium on

the Tsetlin Machine (ISTM) pp. 44-47, **Oct 2025**.

- [gorbenko-2025-tm-diagnosis] Anatoliy Gorbenko, Olga Tarasyuk, Jingjing Zhang, Rishad Shafik, Alex Yakovlev, Matthias Eberl, "*Using Tsetlin Machine for Decoding, Visualization and Minimization of Local Immune Fingerprints in Peritoneal Dialysis Infections*", 2025 International Symposium on the Tsetlin Machine (ISTM), **Oct 2025**.
- [abbas-2025-clinical-tm] Saram Abbas; Naeem Soomro; Ole-Christoffer Granmo; Rishad Shafik; Rakesh Heer; Kabita Adhikari, "AI-Based Clinical Rule Discovery for NMIBC Recurrence through Tsetlin Machines", 2025 International Symposium on the Tsetlin Machine (ISTM), **Oct 2025**.
- [gorbenko-2025-gtrsb-tm] (**Best Paper Award**) A. Gorbenko et al., "*Optimising Tsetlin Machine for Traffic Sign Recognition: A Study of Image Pre-processing, Booleanisation and Ensemble Methods*", International Symposium on the Tsetlin Machine (ISTM), pp. 86-93, **Oct 2025**.
- [jiao-2025-scalable-causal] Ziyang Jiao, Ce Guo, Wayne Luk, "*Scalable Time Series Causal Discovery with Approximate Causal Ordering*", Mathematics, vol. 13, no. 20, **Oct 2025**.
- [faloughi-2025-protohedge] (**Best Paper Award**) Lisa Faloughi, Ce Guo, Wayne Luk, "ProtoHedge: Interpretable Hedging with Market Prototypes", Proc. ACM International Conference on AI in Finance, pp 202-210, **Oct 2025**.
- [duan-2025-ethereal] S. Duan, R. Shafik and A. Yakovlev, "*ETHEREAL: Energy-efficient and High-throughput Inference using Compressed Tsetlin Machine.*", IWASI, **Aug 2025**.
- [abbas-2025-bladder-cancer] Saram Abbas, Naeem Soomro, Rishad Shafik, Rakesh Heer, Kabita Adhikari, "*Attention-enabled Explainable AI for Bladder Cancer Recurrence Prediction*", Proc. IEEE Cong. Engineering in Medicine and Biology, pp 1-7, Jul **2025**.
- [tarasyuk-2025-prediction] Olga Tarasyuk, Anatoliy Gorbenko, Matthias Eberl, Nicholas Topley, Jingjing Zhang, Rishad Shafik, Alex Yakovlev, "*Prediction of the infecting organism in peritoneal dialysis patients with acute peritonitis using interpretable Tsetlin Machines*", Bioinformatics Advances, Vol 5, No. 1, **Jun 2025**.
- [guo-2025-resbench] Ce Guo, Tong Zhao, "*ResBench: A Resource-Aware Benchmark for LLM-Generated FPGA Designs*", HEART, pp 25-34, **June 2025**.
- [khomenko-2025-dist-places] Victor Khomenko, Maciej Koutny, and Alex Yakovlev, "Distributed Places and Safe Net Reduction", Application and Theory of Petri Nets and Concurrency (PETRI NETS), pp. 265-286, **Jun 2025**.
- [que-2025-trustworthy-dl] Zhiqiang Que, Hongxiang Fan, Gabriel Figueiredo, Ce Guo, Wayne Luk, Ryota Yasudo, Masato Motomura, "*Trustworthy Deep Learning Acceleration with Customizable Design Flow Automation*", HEART, pp 1-13, **May 2025**.
- [mao-2025-dynamic-tm] G. Mao, T. Rahman, S. Maheshwari, B. Pattison, Z. Shao, R. Shafik, and A. Yakovlev, "*Dynamic Tsetlin Machine Accelerators for On-Chip Training Using FPGAs*", IEEE Transactions on Circuits and Systems, pp 1-14, **May 2025**.
- [abbas-2025-ai-predicting] S. Abbas, R. Shafik, N. Soomro, R. Heer and K. Adhikari, "*AI predicting recurrence in non-muscle-invasive bladder cancer: systematic review with study strengths and weaknesses*", Frontiers in Oncology, Vol 14, Jan 2025.
- [chan-2024-event-driven-tm] A. Chan, A. Wheeldon, R. Shafik and A. Yakovlev, "*Design of Event-Driven Tsetlin Machines Using Safe Petri Nets" in "International Conference on Applications and Theory of Petri Nets and Concurrency*", pp. 357-378, 2024.
- [que-2024-low-latency-vae] Z. Que, M. Zhang, H. Fan, H. Li, C. Guo and W. Luk, "*Low Latency Variational Autoencoder on FPGAs,*" in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 14, no. 2, pp. 323-333, 2024.
- [wang-2024-trustworthy-codesign] Q. Wang, Z. Que and W. Luk, "*Trustworthy Codesign by Verifiable*

Transformations," 2024 IEEE International Test Conference in Asia (ITC-Asia), pp. 1-6, 2024.

- [guo-2024-bayesian-opt] C. Guo, H. Wu, and W. Luk, "*Resource-Constraint Bayesian Optimization for Soft Processors on FPGAs*", Proceedings of the 14th International Symposium on Highly Efficient Accelerators and Reconfigurable Technologies, pp. 27-36, 2024.
- [zhang-2024-hardware-dropout] Z. Zhang, H. Fan,H. Chen, L. Dudziak,and W. Luk, "*Hardware-Aware Neural Dropout Search for Reliable Uncertainty Prediction on FPGA*", Proceedings of the 61st ACM/IEEE Design Automation Conference, article no. 301, 2024.
- [vandebon-2024-diverse-designs] J. Vandebon, J. G. Coutinho and W. Luk, "*Auto-Generating Diverse Heterogeneous Designs*", IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 116-123, 2024.

Appendix A: External talks

This summarises external talks presenting work based on SONNETS. We do not include the 17 presentations associated with peer-reviewed conference papers listed earlier.

- Internal Presentation: "Tsetlin machines for Explainable Financial Risk", Delivery: R. Shafik, A. Chan, **Feb 2026** : Internal presentation at the Bank of England which gave an overview of Tsetlin Machines and presented explainable risk analysis based on TMAtlas.
- Invited Presentation: “Energy-efficient and Explainable Machine Learning using Tsetlin Machine”, Delivery: R. Shafik, **2025** : Talk given at Durham University.
- Plenary Talk: "Sustainable computation in under-specified or unconstrained HPC workloads", Deliver: D. Thomas, **Oct 2025**, Invited presentation at workshop on Sustainable Scientific Computing at the Lorenz Centre in Leiden to ~50 academics.
- Internal Presentation: "Improving Agent-Based Modelling for UK Housing Market", Delivery: C. Guo, M. Stoddard, **May 2026** : Internal presentation at the Bank of England to showcase the latest advances of the agent-based model study.
- Invited Talk: "Neuromorphic inspired national financial risk", Delivery: D. Thomas, **Mar 2026** : Invited talk at launch of the new ECS Centre for Neuromorphic Technologies, relating the event-triggered aspects of the previous POETS grant and SONNETS, delivered to around 100 academics from across the UK.
- Workshop talk: "Domain-Specific Acceleration Targeting Green AI for Healthcare". Delivery: W. Luk, **Nov 2025** : invited talk at Workshop on the Challenges and Solutions for Energy-efficient AI Systems.
- Invited Panelist: "AI for Designing Novel Architectures". Delivery: W. Luk, **Nov 2025** : Invited panel speaker at the Reach Emerging Architectures in Computing Horizons Conference.
- Keynote talk, "Towards 40 Years of ASAP: From Systolic Arrays to Application-Specific Processing". Delivery: W. Luk, **July 2025** : keynote at IEEE Int. Conf. on Application-specific Sys., Architectures, and Processors (ASAP).
- Keynote talk, "Domain-Specific Acceleration Targeting AI for Science". Delivery: W. Luk, **Jul 2025** : keynote at IEEE Int. Conf. on Signal Processing, Communications and Computing (ICSPCC).
- Workshop Talk: "Multi-FPGA multi-core overlay architectures for executing graph-based applications", Delivery: D. Thomas, Jan 2025 : Invited presentation at Workshop on Reconfigurable Computing as part of HiPEAC. This covered work from the POETS grant, and then the ongoing work in the SONNETS grant.
- Invited Presentation: “Energy-efficient and High Throughput Machine Learning using Tsetlin Machine”, Delivery: R. Shafik, Feb 2025, Invited presentation at EDGE AI Symposium in Austin Texas.
- Academic Talk: "Tsetlin Machines: stepping towards energy-efficient, explainable and dependable AI", Delivery: A. Yakovlev, Feb 2025 : Talk given at Durham University. Covered Tsetlin Machines as an example of advanced machine learning, and how they fit within the SONNETS project.
- Academic Talk: "Optimising Diverse Accelerator Designs", Delivery: W. Luk, Jan 2025 : Invited presentation at WRC 2025 : Covered a set of existing work on optimising compute accelerators, with a call to arms for future deployment and optimisation using SONNETS.
- Workshop Talk: "Accelerating graph algorithms (and more) using 50,000 threads across 50 FPGAs", Delivery: D. Thomas, Jan 2024 : Invited talk at HiPEAC, which covered aspects of the previous POETS grant, and the change in goals and direction for SONNETS.
- Academic Talk: "SONNETS: National financial risk modelling for the UK", Delivery: D. Thomas, Feb 2024 : Selected talk at NVidia event in Southampton to around 200 academic GPU users, talking about

the overall goals of SONNETS.

- Industrial Talk: "Using Event-Triggered Computing to Program the Heterogeneous Cloud", Delivery: D. Thomas, 2024 : Invited talk given at ARM campus in Cambridge, mainly targeting about 40 computer architecture practitioners. The talk covered hardware design from POETS, and then laid out the goals and motivation for SONNETS.
- Forum Talk: "From POETS to SONNETS", Delivery: G. Bragg, Mar 2024 : Talk given to UK Design Forum, which is a national-level forum bringing together industrial and academic attendees from the general area of digital and hardware design in the UK. This talk was promoting the new SONNETS grant, and establishing areas for future potential collaboration.
- Keynote Talk: “Empowering Green Machine Learning: One Logic Conjunction at a time”, Delivery: R. Shafik, 2024, Keynote at NILES Conferences in Egypt 2024,