THE UNIVERSITY OF SYDNEY

FACULTY OF ENGINEERING

Engineering Vacation Research Internship Program

COMPUTER SCIENCE RESEARCH PROJECTS - SUMMER 2022-23

CS2022-23/1 Privacy preserving models for decision tree algorithm 4
CS2022-23/2 Faithful Visualisation for Big Complex Data 4
CS2022-23/3 Neural Interfaces for Visually Impaired as Tactile Enabling Technology 5
CS2022-23/4 Predictive Gesture Classification in Virtual Reality (VR) 5
CS2022-23/5 Drawing graphs with smooth paths and cycles 5
CS2022-23/6 Dimensionality Reduction on the Simplex: Domain Reduction made practical 6
CS2022-23/7 Explanations and Formal Language Theory 6
CS2022-23/8 Verification of Probabilistic Systems 7

CS2022-23/9 Automatic machine learning with neural architecture search for AI applications 7
CS2022-23/10 Design and Implementation of Federated Learning Frameworks for Human Activity Recognition 7
CS2022-23/11 Delta Debugging for Parsers 8
CS2022-23/12 Error Correction with Grammars (Earley) 8
CS2022-23/13 Error Correction with Grammars (GLL) 8
CS2022-23/14 Error Correction with Grammars (GLR) 9
CS2022-23/15 Optimizing Grammar Fuzzers 9
CS2022-23/16 Inferring context-free grammars for Blackbox programs 10
CS2022-23/17 Reproducible HTML Notebooks 10
CS2022-23/18 Mutation Analysis for Fuzzers 11
CS2022-23/19 Using side channels for feedback driven fuzzing 11
CS2022-23/20 Solving Test Reduction Slippage 12
CS2022-23/21 Privacy reserving federated recommendation systems 12
CS2022-23/22 Deep neural network enhancement for real-time video applications 12
CS2022-23/23 Simulation and Modeling of Fluid Dynamics with Deep Learning 13
CS2022-23/24 Machine Learning for Additive Manufacturing 13
CS2022-23/25 Human Motion Analysis and Synthesis 14
CS2022-23/26 Online algorithm design for generalised ski-rental problems 14
CS2022-23/27 Generative Modelling and Optimisation in Federated Learning 14
CS2022-23/28 FeDEQ (Federated Deep Equilibrium Models) 14
CS2022-23/29 Enhancing the alignment of medical images in augmented reality. 14
CS2022-23/30 Verification of an Asynchronous Consensus Protocol 14
CS2022-23/31 AI and NLP: Extracting data from documents for data science 14
CS2022-23/32 AI and NLP: Creating a voice interface for data visualisation 14
CS2022-23/33 Diffusion-based Data Synthesis for Out-of-Distribution Data Detection in Safety-critical Applications 14
CS2022-23/34 Detecting Dark Patterns in the Web at-scale 14
CS2022-23/35 Personalising hints using machine learning in an online programming tutor 14
CS2022-23/36 Generative Adversarial Framework for Privacy-aware Synthetic Data Generation 14
CS2022-23/37 Resilient Edge Computing for Untrusted Distributed Systems 14
CS2022-23/38 Data-driven analysis of sleep disorders using advanced machine learning 14
CS2022-23/39 Defending evasion attacks by using improved classification approaches 14
CS2022-23/40 Efficient Signed Graph Analysis 20
CS2022-23/41 Speeding Up and Scaling Large Graph Analytics 20
CS2022-23/42 An AI-Based Indoor Mobility Patterns Recognition 20
CS2022-23/43 Detecting Fake Content on social media 21
CS2022-23/44 Drones Formation Reordering in Swarm-based Services 21
CS2022-23/45 Machine learning approach for accurate indoor positioning 22
CS2022-23/46 Modelling the human mobility in Smart Cities 22
CS2022-23/47 Drone Navigation and Precise Landing using Machine Learning Approach 22
CS2022-23/48 Pattern Mining for IoT Device Usage in Smart Environment 23
CS2022-23/49 Secure and Usable Distributed Storage 23
CS2022-23/50 Analysing Security of Consensus Protocols in Popular Blockchain Projects 24
CS2022-23/51 Unconditional and Everlasting Secure Encryption 24
CS2022-23/52 Causal Complexity: Augmentation of set-theoretics for machine learning 24

CS2022-23/53 Evaluation of AI Imaging Diagnostic Medical Decision Support Systems in Radiology 25
CS2022-23/54 Apply Process mining in electronic health records 25
CS2022-23/55 Study of social dynamics for health management 26
CS2022-23/56 Extracting Table Information from Financial Document 26
CS2022-23/57 Visual Question Answering 27
CS2022-23/58 Financial markets as an approach to efficiently compute prices 27
CS2022-23/59 Artificial Intelligence for the Simulation and Control of Financial Markets 27
CS2022-23/60 Generative Adversarial Networks for Code Testing 28
CS2022-23/61 Imaging-omics for BioHeart data analysis 28
CS2022-23/62 A grading interface for marking students' short answer questions. 28

THE UNIVERSITY OF SYDNEY

FACULTY OF ENGINEERING

COMPUTER SCIENCE PROJECTS

CS2022-23/1 Privacy preserving models for decision tree algorithm

Supervisor: Dr Nazanin Borhan

Eligibility: WAM>75 and Undergraduate candidates must have already completed at least 96 credit points towards their undergraduate degree at the time of application.

Project Description:
There are always privacy considerations when publishing the results of a data mining or machine learning models publicly. However, privacy preserving models can be a good solution for this problem, without losing too much value and utility from published data. There will be always a trade-off between how much you preserved the data, and how does that effect the accuracy of the published data. In this research, you are looking at decision tree learning algorithm as a popular example of data mining models and explore different privacy preserving techniques that are available on this algorithm. There are a wide range of practical, theoretical and statical models to preserve the result of decision trees in the literature, and this research can give you a good foundation on the existing models to identify the available techniques in the literature, assess and quantify the guarantees provided by each of them, and compare the trade-offs.

Requirement to be on campus: No

CS2022-23/2 Faithful Visualisation for Big Complex Data

Supervisors: Prof. Seokhee Hong; Dr. Amyra Meidiana

Eligibility: Skills Required: Data Structure and Algorithms and Programming (Java, C++, Python, Javascript)

Project Description:
Technological advances have increased data volumes in the last few years, and now we are experiencing a “data deluge” in which data is produced much faster than it can be understood by humans.

These big complex data sets have grown in importance due to factors such as international terrorism, the success of genomics, increasingly complex software systems, and widespread fraud on stock markets.

Visualisation is a powerful tool to compute good geometric representation of abstract data to support analysts to find insights and patterns in big complex data sets.

This project aims to design, implement, and evaluate new visualisation algorithms for faithful visualisation of big complex data, to enable humans to find ground truth structure in big complex data sets, such as social networks and biological networks.

These new visualisation methods are in high demand by industry for the next generation visual analytic tools.

Requirement to be on campus: No

CS2022-23/3 Neural Interfaces for Visually Impaired as Tactile Enabling Technology

Supervisor: Dr Anusha Withana

Eligibility Criteria: You will work with the supervisor and a PhD student, and we expect you are a fast learner. Excellent skills in programming, skills in embedded systems, machine learning, knowledge in design and fabrication, and human computer interaction are added benefits.

Project Description:
Modern computers frequently use visual and auditory interfaces. For example, we can see the visual information on our screens and hear the audio computers generate. However, one important modality is less explored, that is, what we feel through our skin. We feel vibrations, temperature, and pressure through sensory receptors in our skin. This modality is particularly important to create enabling interfaces, for instance computer interfaces used by visually impaired people. In this project, we will explore how we can create enabling technologies using novel tactile interfaces, particularly an interface directly communicates with our neural system.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2022-23/4 Predictive Gesture Classification in Virtual Reality (VR)

Supervisor: Dr Anusha Withana

Eligibility: You will work with the supervisor and a student, and we expect you are a fast learner. Excellent skills in programming, skills in embedded systems, machine learning, knowledge in design and fabrication, and human computer interaction are added benefits.

Project Description:
Can we predict a player's next move in a VR game? Gestural input, for example, using your hand movements as input has become one of the most popular input technologies for rapidly developing virtual reality (VR) and augmented reality (AR) applications (eg. GearVR, HTC Vive, etc.). Gestures, such as hand movements and poses in space, are an essential part of our daily communication (ie. body language) and thus create an intuitive modality for interacting with these immersive new computer applications. In this project we focus on predicting hand movements for the purpose of pre-recognising user activities in VR. The project will build our existing work on continuous hand movement recognition (See video in the link) and the data we have collected.
https://www.dropbox.com/s/jkgik9mj4bd5e83/uist21a-sub1729-cam-i27.mp4?dl=0

Requirement to be on campus: Yes ***** *dependent on government's health advice*

CS2022-23/5 Drawing graphs with smooth paths and cycles

Supervisor: Peter Eades

Eligibility
Essential:
- Good knowledge of graph theory and continuous curve mathematics
- Some knowledge of graph drawing algorithms

Desirable:
- Some experience in either JavaScript or python

Project Description:
Planar graphs and networks are typically drawn with polyline edges (that is, each edge is a connected sequence of straight-line segments). Much effort has been spent over the past two decades in designing algorithms to minimise the number of straight-line segments in an

edge. In the best case, each edge is a single straight-line segment; trivially, a straight-line segment is a smooth curve.

However, drawing each edge as a straight-line segment ignores the *paths* and *cycles* in the graph, which can bend sharply.

We aim to draw planar graphs such that designated sets of paths and cycles are *smooth*, that is, have C^k continuity where $k > 1$. We will use cubic Bezier curves. The challenge is to maintain planarity as well as smoothness.

Specifically, we aim to show that smoothness is achievable for directed paths in upward planar graphs, and for cycles in 4-regular planar graphs.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2022-23/6 Dimensionality Reduction on the Simplex: Domain Reduction made practical

Supervisor: Dr Clement Canonne

Eligibility: Mathematical background in (basic) discrete probability/statistics, and familiarity with coding in Python or C++.

Project Description:
A recent line of (theoretical) work has proposed and relied on a new algorithmic technique, "domain reduction", to hash the (large) domain of randomly sampled data points to a much smaller domain, while preserving some statistical properties of the data distribution. This technique has led to several (theoretical) results on a variety of statistical questions in hypothesis testing under privacy or bandwidth constraints.

Despite its versatility, this technique (essentially a dimensionality reduction for probability distributions) suffers a major drawback: the theoretical analysis might be overly pessimistic, and as a result it is not in its current form clear it's efficient in practice. This project's aims are twofold:

- Understand and improve the analysis of this technique
- Implement it and empirically assess its performance

Requirement to be on campus: No

CS2022-23/7 Explanations and Formal Language Theory

Supervisor: Dr Sasha Rubin

Eligibility: Finished at least 3 years of undergraduate courses in computer science or mathematics, with an HD in COMP202022/292022 or similar course.

Project Description:
The goal is to define and study the computational problem of finding explanations of statements such as "string s is in the regular language L". This has applications to program repair and to giving explanations to actions (made by controllers, policies). The project will involve exploring different logical formalisms in which to describe explanations. A starting point would be first-order logic.

Requirement to be on campus: No

CS2022-23/8 Verification of Probabilistic Systems

Supervisor: Dr Sasha Rubin

Eligibility: Finished at least 3 years of undergraduate courses in computer science or mathematics, with an HD in COMP202022/292022 or similar course.

Project Description:
The purpose of this project is to establish precise complexity bounds on verification problems of probabilistic systems (such as Markov chains, Markov Decision Processes) from declarative specifications, especially over finite traces (such as LTLf, nondeterministic and alternating automata).

Requirement to be on campus: No

CS2022-23/9 Automatic machine learning with neural architecture search for AI applications

Supervisor: Dr Chang Xu

Eligibility: Essential deep learning knowledge and programming skills.

Project Description:
As an AutoML (Automatic machine learning) algorithm, neural network structure search can adaptively find the most suitable structure for any AI architecture such as CNN or transformer, and tasks such as classification, detection, etc. In this project, we will directly teach students AutoML knowledge and provide professional and meticulous academic guidance to help students accumulate rich AI research and application experience. This project will guide students to complete CV tasks such as image classification, image detection, etc. according to their own wishes.

Requirement to be on campus: No

CS2022-23/10 Design and Implementation of Federated Learning Frameworks for Human Activity Recognition

Supervisors: Dr Omid Tavallaie and Professor Albert Zomaya

Eligibility: Specific required knowledge, skills, and/or technology:
Python programming, Machine learning, Android programming, Amazon EC2 servers.

Project Description:
Personal mobile sensing is used for activity monitoring applications such as healthcare rehabilitation. In recent years, by applying deep learning on resource-limited devices, these applications have achieved significant attention from both industry and academia. Compared to cloud-based solutions, running deep learning algorithms on mobile devices provides several benefits including data privacy preservation and low-latency processes. On the other hand, personal mobile sensing applications are mostly user-specific and highly affected by environment. As a result, continuous local changes may seriously affect the performance of a global model generated by deep/federated learning models. To enhance running deep learning algorithms on resource limited devices, we design and implement new deep/federated learning models on android mobile devices.

Requirement to be on campus: No

CS2022-23/11 Delta Debugging for Parsers

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
In cybersecurity, your fuzzers can often come up with a very large input that causes a crash. To debug the crash, one often needs a much smaller input that reproduces the crash. HDD is an algorithm used to reduce the input size if the input conforms to a grammar. However, it does not work well if the grammar is incomplete such as in many handwritten parsers and the input can't be parsed.

In this project, we will explore input-repair based techniques to repair and then parse the input and compare it with using original delta debugging (which does not require a grammar) and coverage-based strategies to identify hierarchies in the partially parsed input.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/12 Error Correction with Grammars (Earley)

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), understanding research papers, problem solving, as well as the ability to work independently are required.

Project Description:
Data aggregation and cleaning is one of the most important steps in data science. The data may come from multiple sources and hence, may not match the required format exactly. This is especially an issue if the required format is a rich format like JSON, XML, S-Expr etc. In such cases, we may have to rely on available error correction algorithms (at best) or manual labour (at worst). While numerous error correction algorithms exist, error correction for context-free grammars is still lagging with the best-known algorithm from 1972 from Aho et al. which extends Earley parsing.

In this project, we will explore how to implement faster error correction for context-free grammars by extending the Aho's algorithm and compare it with the original.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/13 Error Correction with Grammars (GLL)

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), understanding research papers, problem solving, as well as the ability to work independently are required.

Project Description:
Data aggregation and cleaning is one of the most important steps in data science. The data may come from multiple sources and hence, may not match the required format exactly. This is especially an issue if the required format is a rich format like JSON, XML, S-Expr etc. In such cases, we may have to rely on available error correction algorithms (at best) or manual labour (at worst). While numerous error correction algorithms exist, error correction for context-free grammars is still lagging with the best-known algorithm from 1972 from Aho et. al. which extends Earley parsing.

In this project, we will explore how to implement faster error correction for context-free grammars by extending the GLL parsing algorithm (a faster parsing technique for general context-free grammars) and compare it with the original algorithm from Aho et. al.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/14 Error Correction with Grammars (GLR)

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), understanding research papers, problem solving, as well as the ability to work independently are required.

Project Description:
Data aggregation and cleaning is one of the most important steps in data science. The data may come from multiple sources and hence, may not match the required format exactly. This is especially an issue if the required format is a rich format like JSON, XML, S-Expr etc. In such cases, we may have to rely on available error correction algorithms (at best) or manual labour (at worst). While numerous error correction algorithms exist, error correction for context-free grammars is still lagging with the best-known algorithm from 1972 from Aho et. al. which extends Earley parsing.

In this project, we will explore how to implement faster error correction for context-free grammars by extending the GLR parsing algorithm (a faster parsing technique for general context-free grammars) and compare it with the original algorithm from Aho et. al.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/15 Optimizing Grammar Fuzzers

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Good Python, Java, and C Skills. Knowledge of assembly is really good to have.), problem solving, as well as the ability to work independently are required.

Project Description:
Grammar based Fuzzers are one of the most important tools in cybersecurity. The effectiveness of fuzzing is in many cases determined by the speed at which inputs can be generated, and highly performant grammar fuzzers are extremely important. When making a grammar fuzzer, there are multiple tradeoffs that can be made to make it performant. These include fixing the depth of recursion, at which point it becomes automata

that can be easily implemented in code without subroutines, or supercompiling the grammar (i.e., eliminating redundant procedure calls), or superoptimizing the produced code (i.e., assembly level optimizations using solvers), or come up with other better optimization techniques.

This project will explore how to make grammar fuzzers much more performant and compare it with the automata-based approach.

Requirement to be on campus: No

CS2022-23/16 Inferring context-free grammars for Blackbox programs

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Fuzzing Blackbox programs that use an unknown but rich input structure (e.g. conforms to a context-free grammar) has been traditionally very difficult. The problem is that any input that does not conform to the expected input structure will get rejected almost immediately and will not penetrate the actual program logic. Hence, there has been numerous recent attempts to infer the grammar from such Blackbox programs, and to use such inferred grammars for fuzzing. This has been, however, rather difficult due to a theorem by Gold which says that inferring context-free grammars from Blackbox programs is as difficult as cracking RSA.

In this project, we will explore how to get around this requirement by relaxing the opacity constraint a little bit. We will assume that the Blackbox program will tell us when the input is completely incorrect verses whether the input is a valid prefix that can be fixed by adding some suffix. This can allow us to guess at the internal state of the program, and hence, infer the input grammar.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/17 Reproducible HTML Notebooks

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor and possibly a student, and we expect you are a fast learner. Excellent skills in programming (Excellent JavaScript knowledge, and basic Python skills is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Reproducibility is the core of science, and software engineering unfortunately doesn't have a great reputation when it comes to reproducibility. This project attempts to change that. The idea is to start with the WASM powered Jupyter notebooks, bundle them into a completely self-contained HTML file which can contain implementations of algorithms, and can be passed around and modified by other researchers and students. The modified HTML files should be savable using the TiddlyWiki (**https://tiddlywiki.com/**) technique.

This project, if completed successfully, may provide a starting foundation for a workshop on reproducible software engineering, and lead to being adopted by researchers worldwide.

Requirement to be on campus: No

CS2022-23/18 Mutation Analysis for Fuzzers

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Fuzzers are test generators that specialize on producing millions of inputs at a time. The next frontier of fuzzing is to incorporate better oracles, that is verifiers of behavior. However, for this to happen, we need evaluators that can verify behavior. Mutation analysis is the gold standard for evaluating behavior. It accomplishes this by generating thousands of possibly buggy variants of the given program and checking how many of these are detected. To use mutation analysis for fuzzing, its computational expenditure needs to be brought under control. One way to do that is to evaluate multiple bugs at the same time in each execution. However, to do that, we need to ensure that the inserted bugs do not interact with each other. This project will explore the best way to identify independent bugs so that they can be combined.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/19 Using side channels for feedback driven fuzzing

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (expert C & UNIX knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Fuzzing is a premier technique in Cybersecurity and is used by software engineers to ensure that our systems are reliable, and by pen testers to identify possible avenues of attack. Naive fuzzing (blackbox) is usually ineffective in practice, and feedback from the program is necessary for effective fuzzing. Current fuzzing tools such as AFL rely on extensive program instrumentation for feedback, which unfortunately limits their use. Many systems can't be instrumented at all, and in some other systems, instrumentation can change the behaviour of the system.

One way we can avoid instrumentation is by looking at the memory contents at the end of execution, the system calls that were made, and other side channels. This project will explore such side channels especially memory and system calls to find if we can do better than AFL in detecting faults in programs.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2022-23/20 Solving Test Reduction Slippage

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Fuzzers routinely produce very large and incomprehensible inputs that can crash a program. Such inputs can't be debugged unless a much smaller input that reproduces the crash can be obtained. Delta debugging is an algorithm used to reduce the input of a test case to make it comprehensible.

One of the problems with delta debugging is that while reducing the input, new inputs may be produced that induces an unrelated crash. This is called test reduction slippage. The problem is that slippage can result in programmers time being wasted on an unrelated problem. This project will explore how to solve the test reduction slippage using execution grammars and compare it to the effectiveness of using simple coverage.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in Software engineering.

Requirement to be on campus: No

CS2022-23/21 Privacy reserving federated recommendation systems

Supervisor: Wei Bao

Eligibility: Python programming experience; Knowledge in machine learning; Knowledge in recommendation systems or social networks.

Project Description:
Recommendation systems are widely adopted in a variety of areas to predict users' preference and rating on items. It is extremely useful for online stores and social networks. Traditional recommendation system based on conventional techniques, such as collaborative filter, have become mature. However, conventional recommendation systems used users' historical rating to build recommendation models. They require users to exchange their ratings explicitly, causing privacy issues. Federated learning (FL) addresses this issue. A server and users collaboratively train a machine-learning model by privacy-aware interactions.

We are motivated to employ FL for recommendation systems. We plan to investigate how traditional mathematical models (e.g., collaborative filters) can be adaptively employed in the FL environment. In additional, we target to exam how the convergence performance is influenced in the FL environment. Finally, we aim to study the robustness of federated recommendation systems against malicious user attack (e.g., model poisoning).

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/22 Deep neural network enhancement for real-time video applications

Supervisor: Wei Bao

Eligibility: Knowledge in computer networks (especially in UDP and HTTP DASH). Knowledge in deep neural networks. Experience in Python programming.

Project Description:
Real-time videos (video conferencing, online teaching, etc.) have experienced tremendous growth during the pandemic. UDP is adopted as the underlying transport-layer protocol to

avoid delay, but it suffers from packet loss. Another issue is the limited bandwidth in the network. The sender sends low-resolution/low-frame-rate video when the network bandwidth is low, which drastically lowers users' Quality of Experience (QoE).

With the advances of deep neural networks (DNNs), a variety of DNN-based enhancement schemes have been designed. (1) Super-resolution DNNs recover high-resolution videos from low-resolution ones; (2) loss-recovery DNNs can fix distorted frames; (3) interpolation DNNs will recover high-frame rate from low-frame-rate videos; (4) video denoising DNNs can clear noise generated during transmission. Without the need of investing on the improvement in network infrastructure, we can still drastically improve the QoE of real-time videos by DNNs.

You are expected to design a new DNN model to realise video enhancement. The model should be adaptive to the system environment (limited bandwidth and computational capacity), with small data size and small inference delay.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/23 Simulation and Modeling of Fluid Dynamics with Deep Learning

Supervisors: A/Prof Zhiyong Wang; Prof Chengwang Lei

Eligibility:

- UG Students in their 3rd year or higher, or PG students
- Strong Programming Skills (Python)and Math Skills
- Students with knowledge of computational fluid dynamics are preferred.

Project Description:
Understanding the dynamics of fluid motion has extensive applications in a wide range of sectors, such as climate, agriculture, environment, health and medicine, and computer animation. Recently, deep learning has shown a great promising as an effective and efficient alternative for conventional resource-intensive computational approaches. This project aims to investigate physics-informed deep learning methods to simulate and model fluid dynamics. Students will gain comprehensive inter-disciplinary knowledge in deep learning and fluid dynamics.

Requirement to be on campus: No

CS2022-23/24 Machine Learning for Additive Manufacturing

Supervisors: A/Prof Zhiyong Wang; Prof Xiaozhou Liao

Eligibility:

- Strong Programming Skills (Python) and Math Skills
- WAM >= 80

Project Description:
Additive manufacturing (AM), also known as 3D printing, has gained widespread applications in many sectors, including aerospace, medicine and health, and consumer products due to its efficiency, flexibility, capability of building objects with complicated geometries with net shape, among many other advantages. However, it is challenging for quality control and manipulation of mechanical properties of parts made from AM because AM is a far-from-equilibrium process that comprises various processing parameters. This project aims to investigate advanced machine learning techniques for improving product quality of additive manufacturing. Students will gain comprehensive inter-disciplinary knowledge in machine learning and additive manufacturing.

Requirement to be on campus: No

CS2022-23/25 Human Motion Analysis and Synthesis

Supervisor: A/Prof Zhiyong Wang

Eligibility: Strong Programming Skills (Python) and Math Skills

Project Description:
Human motion is of great importance to a wide range of applications such as human action recognition for visual surveillance and character animation for gaming and VR/AR. This project aims to investigate deep learning based human action understanding and cross-modal human motion synthesis. Students will gain comprehensive knowledge in computer vision with a focus on human action recognition and computer graphics with a focus on character animation.

Requirement to be on campus: No

CS2022-23/26 Online algorithm design for generalised ski-rental problems

Supervisor: Bing Bing Zhou

Eligibility: Algorithm (e.g., COMP3027 or equivalent); Python; knowledge in competitive analysis.

Project Description:
The classical ski-rental problem [1] describes the dilemma where a customer trades off between two payment options for one commodity without the knowledge of the future: 1) renting, the customer only pays for an on-demand usage; 2) buying: the customer pays for lifetime usage at once, and no further payment will be incurred after that. To solve this problem, and an online algorithm (which can process its input piece-by-piece without knowing the future) is designed. Then, we can analyse the ratio between its performance and the best offline algorithm's performance (which is called competitive ratio).

In this project, we target to study a more complicated but more realistic scenario called the two-level ski-rental problem when a decision-maker needs to trade-off between rental and purchase of multiple commodities (or services). For each commodity, the decision-maker can rent it or purchase it (which is called single purchase). In addition, the decision-maker can purchase all instances as a combo purchase at a reduced cost. This brings new challenges to decision-makers as payment options are now correlated and cannot be addressed by existing solutions.

[1] Karlin; Kenyon; and Randall. 2003. Dynamic TCP Acknowledgment and Other Stories about e / (e - 1). Algorithmica 36(3): 209–20224.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/27 Generative Modelling and Optimisation in Federated Learning

Supervisor: Dr Nguyen Tran

Eligibility: WAM>80

Project Description:
This project will explore several aspects of distributed optimisation and generative modelling in the context of federated learning. Specifically, we will investigate how data stored on local devices in a network can be leveraged to generate new, high-quality examples (e.g., images, texts, audio, etc.), all while preserving the privacy of these devices. We will explore different sampling techniques and analyse them through the lens of optimisation to design novel generative models. The nature of decentralised learning presents several major challenges such as limited communication and statistical

heterogeneity, which complicates the design and analysis of these models. The project will aim to address these challenges both from theoretical and an empirical perspective.

Requirement to be on campus: Yes (preferred but not mandatory) **dependent on government's health advice*

CS2022-23/28 FeDEQ (Federated Deep Equilibrium Models)

Supervisor: Dr Nguyen Tran

Eligibility: WAM>80

Project Description:
Federated Learning (FL) has emerged as a cutting-edge distributed machine learning paradigm in which a network of clients collaborates to build a shared global model without revealing their private data. Unlike traditional centralized machine learning, FL obviates the need of sending raw data enabling clients to train and send only local models to the server, thereby introducing a reliable infrastructure which not only retains the trust and privacy for machine learning but also reduces the computation and communication costs. However, the inherent statistical diversity in clients' devices restricts the global model from delivering good performance on each client's task. Additionally, the resource-constrained nature of clients' devices complicates the training of FL models. In this project, we address these challenges by proposing a new scheme for personalizing FL based on representation learning, namely FeDEQ. In particular, we leverage the memory efficiency and representation power of Deep Equilibrium Models (DEQ), one of the emerging implicit deep learning frameworks, to design a FL algorithm that learns a common representation in the context of heterogeneous distributed data. By sharing the representation, clients then efficiently personalize their small local model to adapt with the local data. Furthermore, we show the inference of the proposed framework in many FL applications such as computer vision, sequence modelling, etc.

Requirement to be on campus: Yes (preferred but not mandatory) **dependent on government's health advice*

CS2022-23/29 Enhancing the alignment of medical images in augmented reality

Supervisor: Prof Jinman Kim

Eligibility: WAM>75 and should have completed computer vision related units of study. For both PG/UG students.

Project Description:
To extend the user's perception, augmented reality (AR) has been used in a variety of medical settings to augment relevant medical images, such as computed tomography (CT), onto a target location, such as a patient's body. Existing works rely on a semi-automatic alignment approach which requires the user to manually localise landmarks, e.g., using tracking markers. However, this is a cumbersome and error-prone approach whose quality is heavily dependent on the users and their input quality.

An automated alignment approach would be favourable as an enabling technology of AR in medicine where the reliance to the user is eliminated, and the alignment becomes more accurate and robust. You will work with the AR team to create an automatic alignment algorithm using the extensive sensor data from the AR headset and medical images. This project leverages an established research pipeline including collaboration with hospitals, clinical trials, and system implementations.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/30 Verification of an Asynchronous Consensus Protocol

Supervisor: Prof Bernhard Scholz

Eligibility: WAM>75 and Undergraduate candidates must have already completed at least 96 credit points towards their undergraduate degree at the time of application.

Project Description:
Fantom is a blockchain that uses a fast asynchronous consensus protocol called Lachesis to keep its ledger consistent. However, the asynchronous consensus protocol is highly complex and requires correctness proofs to protect the stored values of the blockchain.

This project is the first building block to mechanize the proof work of Fantom's consensus protocol. We will formulate the algorithms for keeping the ledger consistent in verification languages such as LTA+ and Alloy. We will compare the suitability of these model languages and create specifications checking simple properties of the Lachesis protocol.

You will carry out this project with Researchers and Engineers of the Fantom Blockchain foundation and Prof Bernhard Scholz (currently the Chief Research Officer of the Fantom Foundation). The project will also contain training sessions for the model language Alloy and LTA+. We expect that applicants have a general interest in computer science since this project will be demanding practically and theoretically.

Requirement to be on campus: No

CS2022-23/31 AI and NLP: Extracting data from documents for data science

Supervisor: Dr. Jonathan Kay Kummerfeld

Eligibility: Experience with the Python programming language.

Project Description:
Documents contain data in text and tables that could be extremely useful, but only if we can extract it into a structured format (e.g. a database).

This project is about building a system to extract information from wills and tax documents, then using that data to do economic analysis of how people leave their assets. We have a large collection of scanned documents, covering over a hundred years of wills for a specific town in the USA. The system you develop will extend OCR technology (to convert the images to text) and natural language processing (to interpret the freeform text responses).

The goal is to provide new insight into peoples' economic behaviour. This could impact laws, for example, by informing what should be done when someone does not have a will.

Requirement to be on campus: No

CS2022-23/32 AI and NLP: Creating a voice interface for data visualisation

Supervisor: Dr. Jonathan Kay Kummerfeld

Eligibility: Experience with the Python programming language

Project Description:

Writing code to create informative visualisations is challenging. There are powerful libraries such as Python's matplotlib, but achieving a specific goal often requires extensive time on sites like Stackoverflow to see how to use them.

In this project, you will be part of a team building a system you can talk to for producing complex and informative visualisations. This involves research that crosses over from Natural Language Processing (NLP) to Human-Computer Interaction (HCI). There are several components to the project:

- Crowdsourcing methods to create the data needed for AI model training
- AI models for converting a request in English into code
- User interfaces that combine AI generated actions with normal controls

Requirement to be on campus: No

CS2022-23/33 Diffusion-based Data Synthesis for Out-of-Distribution Data Detection in Safety-critical Applications

Supervisor: Dr. Suranga Seneviratne

Eligibility: Familiarity with Python and web applications.

Project Description:
Automated or computer-assisted decision-making is essential to many safety-critical applications. Such decision-making models are generally trained on observed data distributions. However, input data distributions are known to drift over time and, on some occasions, completely un-seen, out-of-distribution inputs in the form of known-unknowns and unknown-unknowns data can be fed into the decision-making models. Specifically, such unknowns can be caused by either covariate shifts (i.e., samples from the same classes but different domains) or semantic shifts (i.e., samples from other classes). As the decision-making models are not trained to handle such input data, they might lead to catastrophic failures in the decision-making process. As such, it is vital to understand how different types of out-of-distribution affect the decision-making process and explore methods of making safe and robust decisions while handling data uncertainty.

In this project, we will explore the feasibility of generating out-of-distribution data from state-of-art diffusion models, focusing on network security-related data streams such as traffic flows, website access patterns, and network telemetry data. The final outcome is to compare diffusion-based out-of-distribution data detection methods with state-of-art methods.

[1] Yang, Jingkang, et al. "Generalized out-of-distribution detection: A survey." arXiv preprint arXiv:2110.11334 (2021).

Requirement to be on campus: No

CS2022-23/34 Detecting Dark Patterns in the Web at-scale

Supervisor: Dr. Suranga Seneviratne

Eligibility: Familiarity with Python and web applications.

Project Description:
Dark patterns refer to GUI design patterns that are designed to mislead users when interacting with web-based applications. Such patterns are heavily misused, especially when it comes to collecting personal information for online advertising. For example, many cookie notifications have deceptive designs that make users confused. Eventually, the users select choices that do not provide them best privacy. Similar practices have been reported in other settings, such as opt-out mechanisms, unsubscription options, and accessing privacy policies.

In this project, we aim to understand common dark patterns and develop methods to identify such patterns at scale by code analysis or computer vision-based methods.

[1] https://dl.acm.org/doi/pdf/10.1145/3479521

Requirement to be on campus: No

CS2022-23/35 Personalising hints using machine learning in an online programming tutor

Supervisors: A/Prof Irena Koprinska and A/Prof Kalina Yacef

Eligibility:

- Machine learning skills, e.g. completed COMP3308/COMP3608 or COMP5318
- Good programming skills
- Good communication skills

Project Description:
This project is using machine learning to generate automated next-step hints in an online computer programming tutor by taking the student's current code as input and returning a suitable hint. Such techniques have been shown to improve learning gains and speed, but they are more efficient if they are personalised to the student. Using an existing variational autoencoder that transforms the program into a point in space and vice-versa, this project will consist in integrating features from the student data into the last layer of the encoding/decoding stage to personalise the neural network to that student and therefore derive better hints.

Requirement to be on campus: *Yes (*Preferred but to discuss with the supervisors if not possible) **dependent on government's health advice*

CS2022-23/36 Generative Adversarial Framework for Privacy-aware Synthetic Data Generation

Supervisor: Dr. Kanchana Thilakarathna

Eligibility: Basic knowledge of statistics and thorough knowledge on applied machine learning.

Project Description:
Gathering data for training machine learning model is a challenge in many domains due to reasons such as user privacy, ethics, time, etc. In addition to the probabilistic approaches such as Copula, recent advancements in Generative Adversarial Network (GAN) based approaches have shown outstanding performance due to their capability of generating high-fidelity datasets in many contexts such as images. However, ordinary GAN architectures are only capable of capturing the distribution of continuous and complete data but cannot be used for learning the distribution of discrete variables in time series data. We proposed a data generation framework for encrypted video traffic data realising Wasserstein GANs in the past. In this summer project, we aim to extend this data generation framework to control the reidentification risk of raw attributes, i.e. private information leakage, and also develop privacy-aware distributed learning framework from synthesized data.

Requirement to be on campus: No

CS2022-23/37 Resilient Edge Computing for Untrusted Distributed Systems

Supervisor: Dr. Kanchana Thilakarathna

Eligibility: Australian Citizenship. Basic knowledge of computer networking and software development.

Project Description:
Edge computing can be a promising technology for defence applications in delivering timely and robust ISR (Intelligence, Surveillance and Reconnaissance) if it is equipped with the recent advances in Machine Learning and Artificial Intelligence. However, there are number of challenges to make edge computing a viable solution in ISR applications such as mitigating the challenges of unreliable data, untrusted networks, security of ML models deployed at the edge, and the lack of computation resources in edge devices. This project aims to develop a prototype edge computing platform that demonstrates the feasibility of assessing trust, isolate untrusted channels/devices at the edge. This will be achieved by developing adaptive network slicing mechanisms for isolating or incorporating untrusted system nodes or channels through software-defined-networking (SDN). The demonstrator will be developed in collaborating with an industry partner on top of their edge computing devices.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/38 Data-driven analysis of sleep disorders using advanced machine learning

Supervisor: A/Prof Irena Koprinska

Eligibility:

- Machine learning skills - completed COMP3308/COMP3608 or COMP5318 with D/HD
- Excellent programming skills

Project Description:
Insomnia and sleep apnea are common sleep disorders. Insomnia is defined by difficulties falling asleep and staying asleep. Sleep apnea is characterized with periods of reduced breathing or no breathing at all during sleep. Both disorders cause daytime sleepiness, fatigue, reduced energy and motivation, and may lead to depression, heart disease, diabetes and other adverse health effects. The goal of this project is to use machine learning techniques to analyse sleep disorders, e.g. to objectively detect insomnia and normal sleepers based on EEG data and predict sleep apnea events in advance based on respiratory data to allow for medical devices to intervene. The project will use large datasets containing data from multiple signals recorded overnight.

Requirement to be on campus: No

CS2022-23/39 Defending evasion attacks by using improved classification approaches

Supervisor: Dr. Xi Wu

Eligibility: Knowledge of DNN, knowledge of Algorithm, and good at programming

Project Description:
Deep Neural Networks (DNNs) are powerful models that have achieved impressive results on image classifications. However, they are vulnerable to be attacked by adversarial examples, which are crafted to cause prediction errors in DNNs. In order to make the networks more robust and reliable, we are aiming to design an improved region-based

classification to mitigate the evasion attack, which is well known to attack DNNs via generating adversarial examples. Specifically, in our framework, an image is considered as a matrix of Markov Chains and we would like to detect possible adversarial examples according to the Image Transition Probabilities (ITPs) in Markov Chains. Furthermore, we would like to employ some improved classifications on updated adversarial examples to get a better output prediction.

Requirement to be on campus: No

CS2022-23/40 Efficient Signed Graph Analysis

Supervisor: Dr Lijun Chang

Eligibility: Strong programming skills in C/C++

Project Description:
Signed graphs enhance the representation capability of traditional graphs, by capturing the polarity of relationships between entities/vertices through positive and negative edge signs. For example, signed graphs capture the friend-foe relationship in social networks, support-dissent opinions in opinion networks, trust-distrust relationship in trust networks, and activation-inhibition in protein-protein interaction networks. One prominent and fundamental theory in signed graph analysis is the structural balance theory. In this project, we aim to identify dense and structural balanced subgraphs from a large, signed graph.

Requirement to be on campus: No

CS2022-23/41 Speeding Up and Scaling Large Graph Analytics

Supervisor: Dr Lijun Chang

Eligibility: Strong programming skills in C/C++.

Project Description:
Graph model has been widely used to represent the relationships among entities in a wide spectrum of applications such as social networks, communication networks, and biological networks. We are nowadays facing a tremendous amount of large-scale real-world graphs with millions or billions of edges. Thus, there is a need of designing efficient algorithms for processing large graphs. In this project, our aim is to design efficient algorithms to speed up graph processing on the ever-growing large graph datasets. The problems that we will be investigating can be (1) dense subgraph (e.g., clique, k-plex) computation over a large sparse graph which finds one dense subgraph of the maximum size, or (2) dense subgraph enumeration which enumerates all maximal dense subgraphs.

Requirement to be on campus: No

CS2022-23/42 An AI-Based Indoor Mobility Patterns Recognition

Supervisor: Prof Athman Bouguettaya

Eligibility:
- Experience in programming in Python or Matlab
- Able to implement machine learning algorithms to estimate Wi-Fi Location
- Experience with backend developments (preferably NodeJS)
- Working knowledge of database technologies (e.g., MySQL)

Project Description:
Indoor mobility patterns refer to the sequences of frequently visited indoor locations, e.g., malls or offices, by a user. Detecting the mobility patterns of individuals is essential in a wide range of applications, including pandemic prevention, crowd management, and

location-based services. Several research studies were proposed to predict human mobility. However, all these approaches focus on an outdoor or macro scale human mobility across various Points of Interest (POI), locations, or city regions. Studies have shown that humans spend over 80% of their lives indoors, which results in indoor or micro-mobility.

Recent research has recognized that we cannot directly use outdoor mobility models that capture mobility at a large grid at a single spatial scale for indoor mobility modelling. This project aims to capture the indoor movement of humans by using Wi-Fi signals. Then, build a model from the collected Wi-Fi signals.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/43 Detecting Fake Content on social media

Supervisor: Prof Athman Bouguettaya

Eligibility: Programming in Python, writing scripts to pre-process data, Typical Machine learning skills, Writing and communication skills.

Project Description:
Social media has become an essential part of our lives. Social media users upload a lot of content on different social media platforms. The content shared on social media may contain untrustworthy images. Existing approaches relies on image processing and object-oriented techniques to detect fake images on social media. These solutions are costly and computationally intensive. Some recent solutions investigate comments on a post to assess credibility of an image. These solutions may not accurately determine the credibility of an image because the fake posts on social media can still get supportive comments. We leverage the metadata of an image and the related posted information to determine the trustworthiness of a crowdsourced image. This information may reflect some changes in the image. For instance, there may be some discrepancies in the metadata which might be an indication of modifications in the image. Therefore, investigating this meta-information may provide useful insights about the changes introduced in the image.

Requirement to be on campus: Yes *dependent on government's health advice

CS2022-23/44 Drones Formation Reordering in Swarm-based Services

Supervisor: Prof. Athman Bouguettaya

Eligibility: Python Knowledge.

Project Description:
Drone swarms used for delivery may fly in different formations for energy conservation purposes. Depending on the wind direction, a swarm may undertake different fly formations to reduce drag forces that consume energy. The reordering of swarm members is a common behaviour that occurs in nature (eg. bird flocks) to conserve energy. A swarm member's energy consumption is typically determined by its position in a formation. We propose reordering a drone swarm formation to reduce the load on the most energy-consuming positions and efficiently distribute energy use. In this project, we use real drones to learn the effect of reordering a formation on the energy consumption through data collection and analysis. Machine learning algorithms would be implemented to predict the optimal reordering mechanism given the different intrinsic (eg. payload and battery capacity) and extrinsic constraints (eg. wind conditions).

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/45 Machine learning approach for accurate indoor positioning

Supervisor: Prof. Athman Bouguettaya

Eligibility:

- Experience in programming in Python or Matlab
- Able to implement machine learning algorithms to estimate Wi-Fi Location
- Experience with backend developments (preferably NodeJS)
- Working knowledge of database technologies (e.g., MySQL)

Project Description:
Indoor positioning refers to detecting the location of humans or objects in indoor environments such as malls or offices. Indoor location is considered the most crucial context information because of its necessity in various real-world applications, such as path navigation, location-based services, risk management, and marketing. One way to determine the position of an object is to leverage WiFi access points. Specifically, the strength of the received signal can reveal how far the access point is. This project aims to build a model to estimate the accurate location using a collected dataset of WiFi signals.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/46 Modelling the human mobility in Smart Cities

Supervisor: Prof. Athman Bouguettaya

Eligibility:

- Experience in programming in Python or Matlab
- Experience with backend developments (preferably NodeJS)
- Working knowledge of database technologies (e.g., MySQL)

Project Description:
Understanding human mobility is essential for the development of smart cities and social behaviour research. Human mobility models may be used in numerous applications, including pandemic control, urban planning, and traffic management. The existing models' accuracy in predicting users' mobility patterns is less than 25%. The low accuracy may be justified by the random nature of the human movement. Visualizing human mobility traces would help understanding the randomness in **human movement.** This project aims to build a platform to simulate people's mobility in a smart city.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/47 Drone Navigation and Precise Landing using Machine Learning Approach

Supervisor: Prof. Athman Bouguettaya

Eligibility: Programming in Python, Machine learning skills, Writing and communication skills, Ability to work with drones on-site.

Project Description:
Using drones in delivery services is a fast-growing industry that gained a lot of commercial attention from companies such as Amazon, DHL, and Google. Drone precision landing is critical for the environment's safety and success of various package delivery missions. We propose a machine learning-based strategy for navigating the drone to the landing position precisely. In contrast, current methods mostly focus on camera-based precision landing, in which an image target is identified, and an exact landing position is determined. When vision is obscured, these solutions frequently suffer from inaccuracy during severe weather circumstances such as fog and sandstorms. Additionally, these solutions have inaccuracies due to shadows. To overcome these limitations, we propose to

train different supervised machine learning algorithms to predict the exact landing position under diverse scenarios using a collected dataset of repeated landings using real drones. Our proposed approach is primarily based on historical data and a set of variables that may influence an accurate landing. The battery state, wind conditions, acceleration, rotation of the drones, and other parameters could all play a role in forecasting the correct landing coordinates. Common trends and behaviours analysed from the dataset could be used to develop error correction strategies for precise landings.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2022-23/48 Pattern Mining for IoT Device Usage in Smart Environment

Supervisor: Prof. Athman Bouguettaya

Eligibility: Python Knowledge.

Project Description:
Internet of Things (IoT) is fast becoming a pervasive universal computing network where everything and everyone is connected to the Internet. IoT technology is the key ingredient for smart environments such as smart homes, smart campuses, and smart cities. The objective of a smart environment is to provide its users convenience and efficiency. In a smart environment, people interact with IoT devices for various purposes. Understanding their IoT usage behavior is an essential step for providing convenient and comfortable services. We will use smart plugs, motion sensors, and indoor environment sensors to record the device interactions of the people in a smart environment. The project aims to build a device usage pattern model that mines peoples' appliance interaction records by applying data mining techniques. The proposed model will focus on discovering and structuring insightful patterns using association rule mining algorithms, clustering algorithms, and classification algorithms.

Requirement to be on campus: *Yes *dependent on government's health advice*

CS2022-23/49 Secure and Usable Distributed Storage

Supervisor: Dr Qiang Tang

Eligibility: Strong programming skills, or mathematically/theoretically mature

Project Description:
When user upload/backup data to outsourced storage (e.g., cloud), the user then has no control of it, data breach becomes universal. On the other hand, if user encrypts data first, then uploads a ciphertext, he loses usability, e.g., cannot access data from a different device as he cannot remember the key. We have recently designed a new protocol that realize the best of both, on the front end, user only needs to remember a pwd as usual, while at the back end, data is protected as properly encrypted.

We will build one exemplary application of such secure and usable storage, such as crypto wallet, or secure backup for E2EE (i.e., secure messenger).

Alternatively, we may also implement a secure distributed (e.g., multi-cloud) storage with the best efficiency so far.

References:

https://www.usenix.org/system/files/sec2022-chen-long.pdf

https://eprint.iacr.org/2020/842

Requirement to be on campus: No

CS2022-23/50 Analysing Security of Consensus Protocols in Popular Blockchain Projects

Supervisor: Dr Qiang Tang

Eligibility: Math or theoretically inclined students, or strong programming skills.

Project Description:
There are many blockchain projects, many of which rely on self-cooked consensus protocols, which might be vulnerable to all kinds of attacks. Many of them are not even well-documented.

We will understand how they work exactly, and investigate potential threats of some of them and consider how to secure them in rigorous manner.

Requirement to be on campus: No

CS2022-23/51 Unconditional and Everlasting Secure Encryption

Supervisor: Dr Qiang Tang

Eligibility: Math or theoretically inclined students.

Project Description:
Honey encryption originally offers a new tool for password-based encryption; recently, we observe that the concept could be broadly useful to circumvent the classical barriers of Shannon's one time pad.

We will investigate how to make it useful to offer unconditional and everlasting security.

Related reference:

https://eprint.iacr.org/2014/155.pdf

https://eprint.iacr.org/2016/174

https://drops.dagstuhl.de/opus/volltexte/2021/14342/pdf/LIPIcs-ITC-2021-23.pdf

Requirement to be on campus: No

CS2022-23/52 Causal Complexity: Augmentation of set-theoretics for machine learning

Supervisor: A/Prof Simon Poon

Eligibility: Good knowledge in data science techniques like Association Rule Mining, Clustering, and SVM. Programming skilling and understanding in statistics and algebra. Students interested in honours and postgraduate research studies are encouraged to apply.

Project Description:
Assessing effects from complex interactions of multiple study factors are common in empirical studies. Changing one factor may have little effect on study outcome if other factors remain unchanged. Furthermore, such interactions may extend to different configurations with complex synergistic relationships amongst many seemingly unrelated factors. From the view of statistical association analysis, especially correlation analysis with the well-known correlation coefficients has been a useful technique for identifying relationships from observational data. However, conventional statistical methods cannot account for situations in which only specific combinations of variables reveal their impact on the outcome (conjunctural causation) or all paths that lead to an outcome need to be simultaneously uncovered (equifinality). These methods also fall short in explaining situations in which a given combination of variables contributes to the presence of an outcome but at the same time is irrelevant for the absence of that outcome (causal asymmetry). In this project, we address these issues by integrating set-theoretic approaches

in conjunction with statistical approaches (like regressions) for discovering meaning configurations from observational data with limited diversity.

Reference: A Configurational Analysis of Risk Patterns for Predicting the Outcome After Traumatic Brain Injury https://pubmed.ncbi.nlm.nih.gov/29854144/

Requirement to be on campus: No

CS2022-23/53 Evaluation of AI Imaging Diagnostic Medical Decision Support Systems in Radiology

Supervisor: A/Prof Simon Poon and Dr Neysa Petrina

Eligibility: Good knowledge of Health Information System implementation & evaluation. Interest in usability research of technology in clinical context. Good knowledge in mixed method research.

Project Description:
The implementation of artificial intelligence (AI) technology in radiology systems as support tools for clinical decision-making promises to substantially expedite the medical screening process, reduce workload burdens, and improve the quality of care provided. However, the utility of AI in a real-world radiology context is significantly impacted by human factors related to daily clinical practice which go beyond diagnostic accuracy such as user acceptability and system usability. Neglect of these in the adoption of past healthcare technologies (e.g. electronic health records, computer-aided detection software) has resulted in suboptimal outcomes such as increased clinician distrust and frustration, and degraded decision-making quality even where the technology was not functionally flawed.

This is part of an ongoing industry-collaboration research project. The aim is to systematically investigate factors contributing to Radiologists acceptance or rejection of AI, in particular Radiologists behavioural responses to AI technology in the diagnostic decision-making process. This is a health information systems project. Student enrolled in combined social/health science and engineering/computing degrees and planning to pursue honours or masters by research are encouraged to apply.

Requirement to be on campus: No

CS2022-23/54 Apply Process mining in electronic health records

Supervisor: A/Prof Simon Poon

Eligibility: Strong interest in medical and health informatics, and process-oriented data science. Good knowledge in data mining and machine learning, and programming skills. Students interested in honours and postgraduate research studies are encouraged to apply.

Project Description:
In recent years, process mining has been increasingly applied to healthcare to gain insights on complex, multidisciplinary clinical pathways. It has been used to inform efforts to improve patient quality of care, improve coordination between departments, and optimize the use of resources [1] It bridges the gap between clinical pathways prescribed by guidelines and the actual processes patients are undertaking. This project aims to apply process mining (PM) methods to generate process models for clinical pathways from electronic health records (eMR). These process models will allow us to study causes of pathway constraints and improvement strategies based on data-driven performance analysis. Process mining algorithms (like Heuristics, inductive and Fuzzy miners) would be used to extract process models from the event logs extracted from eMR. Different data-driven process models will be systematically evaluated to terms of accuracy and fitness in representing clinical pathways.

[1] Munoz-Gama, J. et al. (202022), Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127, 103994. URL - https://www.sciencedirect.com/science/article/pii/S1532046422000107

Requirement to be on campus: No

CS2022-23/55 Study of social dynamics for health management

Supervisor: A/Prof Simon Poon

Eligibility: Good programming skill and experience in simulation studies.
Interest in network science and social network analysis. Students interested in honours and postgraduate research studies are encouraged to apply.

Project Description:
The Guardian Angel concept proposes the idea that a large, decentralised social support network can more effectively motivate individual in a connected community to collectively improve a population's health and lifestyle habits than a traditional centralised system of a few localized hubs, e.g., health care professionals (e.g. clinicians), monitoring a large number of spokes (e.g. patients). The Guardian Angel idea is to organise patients, so that each individual has a guardian angel (or more than 1 guardians) from among the other patients (defined as children) and is assigned to be a guardian angel for someone else. The angel has the ability to engage with the other person's activities and then encourage and motivate the person to continue on a positive trajectory (to accomplish health goal). Every person tries to motivate someone and is motivated by someone else so that the health and lifestyle of the whole community improves.

The aim of this project is to develop an agent-based model using simulation to study the dynamics of different social network topologies.

Reference: Peacock A, Cheung A, Kim P, Poon SK. Socialising Health Burden Through Different Network Topologies: A Simulation Study. Stud Health Technol Inform. 2017;239:112-118. PMID: 28756445.

URL: https://pubmed.ncbi.nlm.nih.gov/28756445/

Requirement to be on campus: No

CS2022-23/56 Extracting Table Information from Financial Document

Supervisors: Dr Josiah Poon and Dr Caren Han

Eligibility: Students must have finished COMP5046 (Natural language processing) and COMP5329 (Deep learning) or have equivalent experiences in the application of image processing and natural language processing techniques.

Project Description:
Tables of information are everywhere. They can be found in academic literature, business reports, news articles or even sport webpages. The numbers in a table are useful for analysis but the extraction is tedious and error-prone manual process. Of course, it is not simply a copy-and-paste of numbers, but also to acquire the semantics of a value in each cell according to the column and row labels. This task turns out to be easier said than done. There were some attempts at this task, but they all had a lot of restrictions & assumptions. A successful project not only benefits the world, but you may have a company selling this service.

Requirement to be on campus: No

CS2022-23/57 Visual Question Answering

Supervisors: Dr Josiah Poon and Dr Caren Han

Eligibility: Students must have finished COMP5046 (Natural language processing) and COMP5329 (Deep learning) or have equivalent experiences in the application of image processing and natural language processing techniques.

Project Description:
Visual question answering (VQA) is a multimodal task that processes both image and text features. Given an image and a question, a VQA model is expected to predict the correct answer to the question based on the image contents. This requires a model's understanding in both images and questions. Current VQA models use different techniques to integrate the image features and question features together for an answer representation, which is then decoded for the answer prediction. In this project, we aim to work on the state-of-the-art VQA models, making modifications to achieve an improved performance with higher accuracy.

Requirement to be on campus: No

CS2022-23/58 Financial markets as an approach to efficiently compute prices

Supervisor: Dr Michael Harre

Eligibility: A working understanding of statistical measures (correlation, p-values, common distributions), an interest in financial markets and designing algorithms for complex problems as well as excellent coding skills.

Project Description:
This objective of this project is to produce a theoretical analysis and simulation of a modified "minority game", a theoretical representation of financial market behaviour. The primary tools will be the Fisher Information Matrix, transfer entropy, and ideas based on unfolding recurrent neural networks. The minority game can be understood as a financial market model that is mathematically related to a fully connected, recurrent network that is most "efficient" for a certain value of a hyper-parameter alpha. For different values of alpha and different network topologies the efficiency of a market can be tuned, but relatively little is understood about how this efficiency is related to network computation and information flow through the network. The outcome of this work will relate the efficiency of a market to its ability to allow information to flow through the market as a function of network parameters.

Requirement to be on campus: No

CS2022-23/59 Artificial Intelligence for the Simulation and Control of Financial Markets

Supervisor: Dr Michael Harre

Eligibility: A working understanding of statistical measures (correlation, p-values, common distributions), an interest in agent-based modelling and financial markets as well as excellent coding skills.

Project Description:
This project has 2 parts to it:

1. You will write software to simulate financial markets using the "minority game" as the basis. This involves simulating "fat tails", "clustered volatility", and "criticality", each of these properties contribute to the complexity of modelling markets. Your task is to design AIs that trade in this market to reduce market inefficiencies and stabilise dynamics.
2. Then, you will take financial market data for individual stocks and measure periods of clustered volatility, strong correlations, and fat tails, and re-examine

the statistics when correlated stocks are removed from data to test hypotheses about the relationship between correlations and market inefficiency.

These parts can be taken independently and studied in more detail, or in combination and comparisons between them used to evaluate theory against real data.

Requirement to be on campus: No

CS2022-23/60 Generative Adversarial Networks for Code Testing

Supervisor: Dr. Basem Suleiman

Eligibility: Selected candidate(s) must meet the following required knowledge and skills:

- Good understanding of unit testing, code coverage, and machine/deep learning
- Designing and implementing Generative Adversarial Networks (GAN) models
- High familiarity with keras or pytorch
- UI design and implementation (preferred)

Project Description:
The goal of this project is to investigate the use of Generative Adversarial Networks (GAN) to generate automate code tests for software applications. The scope of the project includes researching relevant research studies and background information and critically analyse its contributions, design, and implement new GAN models to generate unit test code, conduct experimental analysis and evaluation of the designed models and compare the results with related approaches. Relevant dataset should be also generated (consisting of simple source codes and their unit test code counterparts) so it can be used in the evaluation. The project would also involve designing a simple UI that allows to use/interact with the developed GAN models.

Requirement to be on campus: No

CS2022-23/61 Imaging-omics for BioHeart data analysis

Supervisors: Prof Jinman Kim and A/Prof Luping Zhou

Eligibility: WAM>75 and should have completed computer vision related units of study for both PG/UG students.

Project Description:
This project will focus on developing approaches for imaging-omics. One of the key imaging-omics technologies that we plan to investigate deep learning algorithm for cardiac imaging data that is available in our BioHeart data.

Currently much of the initial data processing is done manually and extremely time consuming. There is a real need to automate image quantification to ensure scalability in our workflow. This project will join a multidisciplinary team to develop an automatic imaging workflow to enable fast processing of cardiac CT imaging data.

Requirement to be on campus: Yes *dependent on government's health advice

CS2022-23/62 A grading interface for marking students' short answer questions

Supervisors: A/Prof Kalina Yacef and Prof Judy Kay

Eligibility: High level skills required:

- HCI / UX skills (such as taught in INFO3315 Human Interface Design), and front-end web development.
- Good programming skills
- Good communication skills

Project Description:
We are developing a human-in-the-loop system to assist human graders assess short answer questions and code comments, by suggesting automated, machine-learning derived feedback and rubric item selections to the human grader. This summer project is to build and test a prototype of a suitable grading interface that should provide the human with the additional information without disrupting their workflow. The student will work in collaboration with PhD student Joe Godbehere.

Requirement to be on campus: Yes, (preferred but to discuss with the supervisors if not possible) and dependent on government's health advice