

Indicative Projects

PROJECT TITLES FOR FINAL YEAR PROJECTS IN COMPUTING; COMPUTER SCIENCE;
SOFTWARE ENGINEERING CLUSTERS

2023-24

The list below provides Final Year Project topic suggestions for students in BEng Software Engineering, BSc Computer Science, BSc Computer Security and Forensics, BSc Computer Systems and Networking

Please **find your supervisor in the list below and select the projects he/she is suggesting**. It is imperative that you discuss the topic with them and agree to the specific requirements you will be working on.

*Please use the **SEARCH** facility to find your supervisor as they may appear in more than one locations in the text below. You may also use the MS Word Navigation Pane to browse the list.*

Contents

Christopher Beckwith	7
Numerical Simulation (especially Computational Fluid Dynamics and Computational Acoustics)	7
Grid generation is often a problem	7
Cornelia Boldyreff	8
Source Code Analysis projects	8
Software Visualisation projects	8
Scheme code analysis and visualisation	8
E-learning for programming projects	9
Mutation Testing projects	9
Software metrics for OO programming and design.....	9
The Rust Programming Language	9
Other Projects	10
Diane Gan	11
Project using open source Natural language processing (Python) to detect online abuse - http://www.nltk.org	11
An investigation into a personal security issue related to social networking sites (data leakage).	11
Retrieving Hidden Friends: A Collusion Privacy Attack Against an Online Friend Search Engine.....	12
Honeypots to detect misuse	12
Identify the “pattern of life” of smart home devices.	12
IoT cyber attacks with a digital forensics investigation to identify what has happened.....	12
Identifying Botnets within the traffic generated by a network.....	13
Creating/detecting fakes – this could be pictures or video, or even fake Instagram accounts Fintsa) 13	
Investigation into risk analysis of a network	13
Elena Popa	15
Real time speech emotion recognition.....	15
Health risk prediction software solution	15
Applying social media analytics algorithms for effective prediction of movie reviews	15
Create a prototype mobile app to evaluate feasibility of combining Case-Based Reasoning with sensors for movement categorization	15
Identifying depression using Deep Learning and Conversational Agents	16
Georgia Sakellari	17
Network Monitoring and Management in IoT.....	17
Cloud-enabled energy-efficient personal computing	17
Computation Offloading in IoT Systems	17
Computation Offloading for the improvement of multi-layer Anomaly Detection in Industrial IoT (IIoT)	18

ISEC12: Network Monitoring and Management in IoT.....	18
ISEC13: Cloud-enabled energy-efficient personal computing.....	19
ISEC12: Network Monitoring and Management in IoT.....	19
ISEC13: Cloud-enabled energy-efficient personal computing.....	20
George Samakovitis	21
Graph Representation and Analysis of Transaction Paths in Cryptocurrency Networks	21
Identification of Customer Groups in Money Exchange Networks with Clustering.....	21
A Tool for Stock Portfolio Optimisation based on Investors' Risk Profiles	21
Sentiment Analysis for Predicting Stock Price Behaviour.....	22
Pattern Recognition Approaches for Identifying and predicting Stock Market Abnormal Returns	22
Recommending Intraday Stock Trading Strategies.....	23
Hai Huang	24
Validation of Factual Information in Knowledge Graph	24
Mapping from Tabular Data to Knowledge Graph (KG)	24
Entity Recognition, Linking and Typing for Knowledge Base population	24
Mapping from Tabular Data to Knowledge Graph (KG)	25
Product matching.....	25
Product classification	25
Ik Soo Lim	27
Cloud-based Wireless Live-Streaming Social Networks:.....	27
Improving Cooperation in P2P Networks:	27
Task Allocation in Multi-Agent Systems:	27
Social Internet-of-Things:	27
Jia Wand	28
Identification of leisure hubs from geo-tagged social media posts	28
Identification of tourist locations from geo-tagged social media posts.....	28
Insights from Uber movement data	28
Modelling road networks and pavements with smart navigation	28
Jing Wang	29
Unsupervised Learning for Multi-view Data.....	29
Dimension reduction for High-Dimensional Data.....	29
Semi-supervise clustering with Ordinal Priors.....	29
Robust nonnegative matrix factorization	29
Jia Wand	31
Identification of leisure hubs from geo-tagged social media posts	31
Identification of tourist locations from geo-tagged social media posts.....	31

Insights from Uber movement data	31
Modelling road networks and pavements with smart navigation	31
John Ewer	32
Develop a Web-based or stand-alone application based on the Covid- 19 Aerosol Transmission Estimator and undertake Comparative Analysis of Wells-Riley models with CFD based Infection Risk Modelling.	32
Develop a Stand-Alone Analysis and Summary Report Writing Tool for the Fire Field Modelling Software (SMARTFIRE).....	32
Develop a Mesh Quality Checking Tool for Unstructured Meshes.....	33
Develop a stand-alone tool to process and extract data (e.g. monitor point values vs time; graph values) from a set of VTK/VTU field data results files from a CFD simulation.	33
Manos Panaousis	35
A Tool to Advise CISOs	35
Adversarial machine learning against IoT devices.....	35
Cyber forensics visualisation tool	35
Risk assessment tool using Caldera	36
Threat intelligence tool.....	36
Cyber insurance tool.....	36
Mohammad Majid Al-Rifaie	37
I. Evolving Deep Neural Networks	37
II. Protein Folding.....	37
III. Evolving Deep Neural Networks	38
IV. Solving the Single Row Facility Layout Problem	38
V. Solving Bin Packing Problems	39
VI. Solving Travelling Salesman Problem	39
VII. FPGA Implementation of a Nature-Inspired Algorithm.....	40
VIII. GPU and Multi-Processing Implementation of a Swarm Intelligence Algorithm	41
IX. Network Alignment with Swarm Intelligence.....	41
X. Cuneiform Sign Detection and Reading.....	42
Other Projects:.....	43
Muhammad Taimoor Khan	44
Run-time security monitoring of web application.....	44
Secure-by-design web application.....	44
Model based security risk assessment of machine learning applications.....	44
Modelling classes of security threats.....	44
Model based intrusion detection	45
Governance-by-design.....	45

Ethics-by-design	45
ISEC32: Model based intrusion detection	45
ISEC33: Governance-by-design	46
ISEC34: Ethics-by-design	46
Markus Wolf	47
Maze Solving App.....	47
Holiday Scheduling System	47
Infrared Motion Detector Using Arduino	47
Image Recognition to Aid Fire Evacuations	47
Pushparajah Rajaguru	48
Machine learning algorithms for Surrogate models and uncertainty optimisation	48
Reducing computational complexity in Finite Element modelling by fast matrix to matrix algorithm.	48
Machine learning algorithms for medical image processing.....	49
Rafael Martinez Torres	50
Big Data and London Traffic flow.....	50
Music Training-Session Scheduling	50
Microsoft 10X Utility.....	50
Machine Learning and Trade Markets.....	50
Ralph Barthel	51
ISEC14: Visualising the Internet of Things through the type of data collected (Energy).....	51
ISEC15: Visualising Internet of Things data (Privacy).....	51
ISEC16: Monitoring and monetising the use of personal digital data using blockchain technologies .	52
ISEC17: A Computer Supported Collaborative Learning (CSCL) tool for Individual and Socially Shared Regulation of Learning (SSRL) during group work activities.....	52
Sadiq Sani	53
Blockchain–powered Identity Management in Internet of Things.....	53
Ransomware Attack Mitigation in Internet of Things: Control and Security.....	53
Decentralized Trust and Safety-enabled Supply Chain Platform using Blockchain Technology	53
Privacy-preserving Energy Trading and Storage Systems.....	54
Collaborative Key Management Techniques for Internet of Things Cryptocurrencies	54
Blockchain-based Documents Certification Platform.....	54
Design of Trust Techniques for Cybersecurity and Operations in Internet of Things	54
Blockchain-based Healthcare Platform	55
Peer-to-Peer Secure Situational Awareness and Decision-Making System for the Internet of Things	55
Stef Garasto	56
Prediction analytics and algorithmic fairness	56

Machine learning for Healthcare	56
Detection of idioms usage within text.....	57
Forecasting future frames in a video	57
Tuan Nguyen	59
Face liveness detection.....	59
Object detection from images captured from drones.....	59
Fish recognition	59
Object/Person re-identification.....	59
Auto generating captions for images.....	60
Colourising images or videos	60
Automatic detection of cyberbullying	60
Building a movie/book recommender system to predict the rating score of a movie/book.....	60
Detecting if the provided pairs of questions contain the same meaning	60
Text summarisation	61
TUAN VUONG	62
ISEC06: Data collection for driver training enhancement	62
ISEC07: Machine learning and data analytics for driver training enhancement	62
ISEC08: Machine learning in intrusion detection	62
ISEC09: Reinforcement learning in the Internet of Things	63
ISEC10: A software and hardware upgrade kit that makes a robot autonomous	63
ISEC11: Using AI to compose music.....	63
ISEC 45: Automation in self driving car.....	63
ISEC 46: Smart home sensor system for intruder detection & energy control	64
ISEC31: Modelling classes of security threats.....	64
Yasmine Arafa	65
Analysing Patterns of Affect in Facial Expressions.....	65
Detecting the Golden Ratio in 2D Logo Images	65

Christopher Beckwith

Numerical Simulation (especially Computational Fluid Dynamics and Computational Acoustics)

The focus will be on developing a "product". Projects in these areas could be built around making user facing applications to solve particular CFD/Acoustics problems, rather than the modelling itself

- Examples could be creating a program that models the spread of airborne diseases given a room plan, one that given some 3D geometry and wind profile can simulate the spread of pollution, etc.

Grid generation is often a problem

There are basic GUI meshers, but they are often very lacking, and decent command line meshers that have a high difficulty curve. Anything else is trapped behind thousands of pounds of licensing costs. This project would focus on a GUI frontend for one of the publically available command line meshers?

Cornelia Boldyreff

Source Code Analysis projects

These projects are on source code analysis. Either to detect plagiarism in small student programs in Java, Python or Scheme, or to detect clones in large open-source projects developed by a team of developers where the focus will be to improve the code.

In the later case, you will be expected to identify suitable open source projects to analyse. In the former case, you will be supplied with anonymised student submissions or working from a freely available code base.

Both projects will involve reviewing source code analysis tools and improving or enhancing them for their purpose: plagiarism detection or clone detection.

Resources:

Source-code Similarity Detection and Detection Tools Used in Academia: A Systematic Review
<https://dl.acm.org/doi/10.1145/3313290>

A Comparison of Three Popular Source code Similarity Tools for Detecting Student Plagiarism
<https://dl.acm.org/doi/10.1145/3286960.3286974>

How to detect plagiarism in text using Python
#python#computerscience#datascience#machinelearning

Learn how to Make a Plagiarism Detector in Python using machine learning techniques.
<https://dev.to/kalebu/how-to-detect-plagiarism-in-text-using-python-dpk>

Addressing Plagiarism: How to Reduce Occurrences in a Computing Curriculum Through Targeted Assessment
<https://www.learntechlib.org/p/221993/>

Towards a Dataset of Programming Contest Plagiarism in Java
<https://arxiv.org/abs/2303.10763>

CopyPoppy – A Source Code Plagiarism Detector
<https://ieeexplore.ieee.org/abstract/document/10001740>

A Systematic Review on Code Clone Detection
<https://ieeexplore.ieee.org/abstract/document/8719895>

Software Visualisation projects

This project is concerned with supporting small teams of Python programmers to visualise the software they are developing and identify authorship of individual code within the application. It could be developed as addition to the PyCharm IDE.

Resources:

3 Tools to Track and Visualize the Execution of Your Python Code
<https://www.kdnuggets.com/2021/12/3-tools-track-visualize-execution-python-code.html>

Visualization of Source Code Similarity Using 2.5D Semantic Software Maps
https://link.springer.com/chapter/10.1007/978-3-031-25477-2_8

OpenSeesPyView: Python programming-based visualization and post-processing tool for OpenSeesPy
<https://www.sciencedirect.com/science/article/pii/S2352711022001960>

Scheme code analysis and visualisation

Research into the topic of software visualisation is required as well as specific work related to supporting collaborative software development.

A further project could consider similar support for teams working on open source projects. Here the development of a plug-in for an open source IDE such as Eclipse or NetBeans could be considered.

A further project could focus on visualisation related to impact analysis to show visually the impacts of a proposed software change to an open source software development project.

E-learning for programming projects

Enhancing DrRacket with Dodona for Learning Scheme

This project involves research into e-learning support for students of programming with specific research into the Dodona system and developing suitable materials in English for first year undergraduates learning Scheme. The learning materials can build on those used in the COMP1811 module.

(do something in English)

Further project could address material for first year students learning Python or Prolog building on those used in the COMP1811 module. The later could involve development of further materials for Prolog. These would need to consider integration with PyCharm or SWISH/SWI Prolog.

Resources:

Enhancing DrRacket with Dodona for Learning Scheme

<https://dl.acm.org/doi/abs/10.1145/3532512.3564142>

Dodona: learn to code with a virtual co-teacher that supports active learning

<https://arxiv.org/pdf/2210.10719.pdf>

Mutation Testing projects

This project involves research into software testing and mutation testing and specifically mutation test analysis for students learning Scheme with Racket.

A further project could focus on students learning Python with PyCharm and relevant mutation testing techniques.

Resources:

A Tool for Mutation Analysis in Racket

<https://www.jonbell.net/preprint/mutation23-racket.pdf>

Software metrics for OO programming and design

Building on work on OO metrics in general and specifically for UML OO based designs, this project will review the state of the art and identify specific application to student developers code and designs to support re-factoring in either Python or Java.

A further project could research the use of ML to identify design patterns in OO code.

The Rust Programming Language

This project will investigate the Rust programming language and give consideration into how easily a system could be developed to assist Python programmers to move to Rust through semi-automated translation of Python into Rust.

A further project could consider Rust IDEs and evaluate their suitability for student use. This is an exciting area of work and there is currently little published research available.

Resources:

<https://www.technologyreview.com/2023/02/14/1067869/rust-worlds-fastest-growing-programming-language/> MIT Technology Review, 2023

Learning and Programming Challenges of Rust: A Mixed-Methods Study

<https://songlh.github.io/paper/survey.pdf>

Rust IDEs <https://blog.logrocket.com/comparing-rust-ides-intellij-idea-vs-code-more/>

Other Projects

- Improving the IDE for Scheme DrRacket
- Animating Scheme functions
- Software Metrics projects
- Scheme metrics
- Python metrics

Diane Gan

Project using open source Natural language processing (Python) to detect online abuse - <http://www.nltk.org>

The objective is to identify online abuse (child grooming or bullying) using natural language processing from chat rooms, Twitter or any other online platform. The focus would be to modify the code or develop a Twitter API to collect tweets.

The NLTK (Natural Language Toolkit) has a lot of examples, tutorials, howtodo and ready written APIs. This can be developed into a tool or used as a proof-of-concept project to demonstrate the process. There are various NLP techniques such as tokenisation, removing stop-words, stemming and Part-of-Speech (POS) Tagging were used for pre-processing. Tokenisation is the process of splitting the sentences into words by not only using space delimiters but by also using punctuation marks. Paper to explain the tool - <https://arxiv.org/abs/1803.07640>

As part of the natural language tool kit they have an API to collect data from Twitter. This can be used to extract information about individuals to identify their pattern of life. The NLTK (Natural Language Toolkit) has a lot of examples, tutorials, howtodo and ready written APIs that can be used. <http://www.nltk.org/howto/twitter.html>

There is also a section on Sentiment Analysis which is an important aspect for identifying online abuse - <https://github.com/nltk/nltk/wiki/Sentiment-Analysis>

- Build a trained model of sentiment for a large-ish Tweet corpus
- Add a module for feature-based classification (e.g. using the Customer Product Reviews or likes)

Link to a paper written by the developers - <http://www.aclweb.org/anthology/P14-5010>

Two platforms that can be used - <http://www.nltk.org> and <https://allennlp.org/>

There are some data sets available at <https://github.com/nltk/nltk/wiki/Sentiment-Analysis>

Other information extractor algorithms such as Rapier, Support Vector Machine (SVM), Conditional Random Field (CRF) and Maximum Entropy (MaxEnt)

This project would suit a forensics and security student or a computer science student who is interested in security.

An investigation into a personal security issue related to social networking sites (data leakage).

There are a number of facets to this project, so that there are effectively 4 projects here. Social networks can pose a serious threat to an individual's personal information that they might want to keep private. There are a number of choices for this project, so this could be considered as four separate projects:

1. Using readily available tools for a specific social networking platform. There are many free tools which can be used to identify a "pattern of life".
2. Build your own app to harvest information from your chosen social networking site. This could be to look at GPS information embedded in images which have been uploaded to identify a "pattern of life". For example, your "victim" goes to the gym every week at a specific time.
3. Look at a large sample of people to identify trends. For example, using the "likes" in Facebook to trace people's preferences, fake news or looking at tweets about a specific topic and how they spread through cyber space. This could be a mapping tool.

4. Secondary leakage is a serious issue. You may think that your data is secure, but some of your “friends” may be leaking your personal information and pictures. Pictures can contain GPS coordinates.

This project would suit a forensics and security student

Retrieving Hidden Friends: A Collusion Privacy Attack Against an Online Friend Search Engine

Online Social Networks (OSNs) are providing a variety of applications for human users to interact with families, friends and even strangers. One of such applications, friend search engine, allows the general public to query individual users’ friend lists and has been gaining popularity recently. However, without proper design, this application may mistakenly disclose users’ private relationship information. Our previous work has proposed a privacy preservation solution that can effectively boost OSNs’ sociability while protecting users’ friendship privacy against attacks launched by individual malicious requestors. In this paper, we propose an advanced collusion attack, where a victim user’s friendship privacy can be compromised through a series of carefully designed queries co-ordinately launched by multiple malicious requestors. The effect of the proposed collusion attack is validated through synthetic and real-world social network data sets. The in-depth research on the advanced collusion attacks will help us design a more robust and securer friend search engine on OSNs in the near future.”

This project would suit a forensics and security student.

<https://ieeexplore.ieee.org/document/8443384/>

Honeypots to detect misuse

There are also many uses for Honeypots, which also include IoT botnet attacks, tracking hackers, intrusion detection and even social honeypots and spam detection on Twitter. The Kippo honeypot is an appropriate candidate and is free from The HoneyNet Project but there are others to use. You could even build your own honeypot.

A nice introductory paper is here - <https://onlinelibrary.wiley.com/doi/full/10.4218/etrij.2019-0155>

This project would suit a forensics and security student or a networking student who is interested in security. If you want to build your own then this would also be suitable for a Computer Science student.

Identify the “pattern of life” of smart home devices.

This project is the build a tool using an appropriate device (e.g. Raspberry Pi), which will monitor the usage of smart home devices and learn how the owners use these devices. This includes time of day when devices are used, etc. This tool can then be used to identify deviations from this pattern which would indicate malware or an intruder. For example, there is a smart TV which suddenly starts to connect to a server in China. This could indicate malware or could mean that the device has become part of a botnet. This could utilise a Raspberry Pi to monitor the smart home devices.

This project needs access to smart home devices.

This project would suit a forensics and security student.

IoT cyber attacks with a digital forensics investigation to identify what has happened.

Investigate and document a number of different attacks against the components of a smart home which has a number of IoT devices present. What residual evidence is left behind post attack? Can the attack tool be identified?

This project needs access to a small scale testbed of smart home devices and a device to perform the attack. It could also be undertaken using an IoT simulator.

There are a number of IoT simulators available:-

<https://aws.amazon.com/answers/iot/iot-device-simulator/>

There are 6 available here <https://windowsreport.com/iot-simulators/>

Cupcarbon U-One - <http://www.cupcarbon.com/>

AWS IoT Services <https://github.com/aws-labs/iot-device-simulator> and

<https://aws.amazon.com/answers/iot/iot-device-simulator/>

<https://www.quora.com/What-are-some-good-simulation-tools-to-simulate-an-IoT-based-project>

<https://www.bevywise.com/iot-simulator/>

“IoT Simulator helps you test your cloud and on premise MQTT Application for functional and load testing. Simulate tens of thousands in a commodity server. Develop, test and demo your IoT Servers & managers, MQTT clients, MQTT Sensors & MQTT Devices.”

If using a simulator the project must include a comprehensive investigation to justify why you chose a particular simulator.

Ref - <https://www.sciencedirect.com/science/article/pii/S1383762116300662>

This project would suit a forensics and security student or a student who is taking forensics.

Identifying Botnets within the traffic generated by a network

There are a number of data sets which consist of packets that contain both normal traffic and assorted botnet traffic that can be used.

Use machine learning to identify the botnet packets and then design a tool to do this automatically. I have a free machine learning tool which is easier to use than R.

There is a simulator available for NetFlow at -

<https://www.gambitcomm.com/site/netflowsimulator.php>

This project would suit a forensics and security student or a networking student.

Creating/detecting fakes – this could be pictures or video, or even fake Instagram accounts (Fintsa)

“Instagram users can create two types of account to manage their self-presentation strategically.

On a real Instagram account (known as Rinsta), users highlight flattering aspects of self, whereas on a fake Instagram account (known as Fintsa), users show unflattering aspects of self.”

<https://www.tandfonline.com/doi/abs/10.1016/j.soscij.2018.12.005>

Creating the images of people who do not exist – deep fakes.

<https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2020.1754877>

Detecting fake pictures - “Creating fake pictures becomes more accessible than ever, but tampered images are more harmful because the Internet propagates misleading information so rapidly.”

<https://www.sciencedirect.com/science/article/pii/S0020025520302796>

This project would suit a networking student or possibly a forensics and security student.

Investigation into risk analysis of a network

This project uses the SecuriCAD software to model a network and evaluate the vulnerabilities present. Solutions to strengthen the network can be added and the attacker can be prevented from accessing the valuable assets of the organisation. This requires the modelling of a reasonably large network using the SecuriCAD simulation tool.

Conceptual Abstraction of Attack Graphs -a Use Case of SecuriCAD – “Attack graphs quickly become large and challenging to understand and overview. As a means to ease this burden this paper presents an approach to introduce conceptual hierarchies of attack graphs. In this approach several attack steps are aggregated into abstract attack steps that can be given more comprehensive names.”

<https://gramsec.uni.lu/2019/presentations/gramsec19paper4.pdf>

Elena Popa

Real time speech emotion recognition

The aim of this project is to perform an investigation of speech emotion systems and how the accuracy can be further improved by exploring machine learning algorithms and hybrid solutions.

The product would require a speech recognition system which could include the following high level components: a dataset; feature extraction techniques to extract the audio features; feature normalisation; a pattern recognition algorithm (machine learning algorithm or classifier) to train and predict specific emotions.

Real time speech emotion systems require a voice detector, a pre-processing component and feature extraction and normalisation. A voice detector detects the presence of the user's voice; for example, the voice detector chosen could be a hot-word detector, which detects the presence of a keyword. The pre-processing component removes the noise and normalises the audio in the speech file. The Feature extractor should extract relevant features; for example, Teager Energy Operator (TEO) can detect stressed emotions.

Some pattern recognition algorithms which can be investigated: K-Nearest Neighbour (K-NN), Support Vector Machine (SVM), Random Forest Classifier (RFC), Multi-Layer Perception (MLP) and Convolutional Neural Networks (CNN) algorithms.

Health risk prediction software solution

The aim of the project is to provide a solution that would continuously monitor the lifestyle of a person and diagnose/predict the health problems that may arise in the future. The proposed solution is an app which would continuously monitor the life style of a person; apply analytics to data gathered by a health monitoring software and warn of any potential health problems that may arise in the future.

The life style should take in consideration parameters such as diet or physical activities (and types of exercises), sleeping hours, etc.

Applying social media analytics algorithms for effective prediction of movie reviews

The aim of the project is to predict movie reviews based on social media datasets such as Twitter, etc. Social media consists of the various blog post, reviews, microblogging, surveys, news articles and internet forums. These platforms enable the daily activity of users and providing them with an opportunity to express themselves. Analysing this data is a popular topic since recent years to generate valuable insights. There are many approaches to achieve this such as artificial neural networks, swarm intelligence, sentiment analysis, text mining, deep learning etc.

Create a prototype mobile app to evaluate feasibility of combining Case-Based Reasoning with sensors for movement categorization

Case-based reasoning (CBR) is an Artificial Intelligence technique which outlines the process of solving new problems which are based on the solution of similar problems while learning from the new cases, if possible.

Accelerometers and gyroscopes in mobile devices could be used to capture data on movements. It may be possible to detect types of movements by applying case-based reasoning.

Having a mobile app that helps detect falls and automatically notifies relevant entities could save money as most people would have a mobile phone and wouldn't have to purchase

dedicated equipment. It is also very likely that someone will keep their mobile phone on them which may make it more reliable than remembering to carry specialised equipment. An app that learns over time can help detect genuine injury and reduce the potential for more serious injury or distress through more reliable and timely notification.

There are a number of smart phone apps on the market, such as Ep Detect (which detects epileptic seizures) and Fall Safety App (which detects falls by elderly and vulnerable people). These use signal processing of data from the sensors and this results in a high rate of false positives (Ramgopal et al. 2014). False positives create a considerable waste of time and effort for carers and diminishes the confidence of the user in the technology.

Case-based reasoning is an artificial intelligence technique which is able to consistently expand the case-base and learn over time, so the number of false positives decreases over time as the case base evolves.

Identifying depression using Deep Learning and Conversational Agents

The project aims to design an effective Deep Learning model to perform feature extraction to analyse user's experiences and thoughts using their conversations with a bot, and keeping a track of their feelings and sentiments every day to accurately predict/identify any signs of depressive disorder.

Some algorithms to investigate: RNN and CNN

Georgia Sakellari

Network Monitoring and Management in IoT

Description: The constant growth of IoT devices is changing the way that networks are monitored and managed. The vast number of devices and the different communication protocols (IP, RFID, NFC, ZigBee, etc.) that they are using has created implications on how they can be managed through a single network monitoring/management software system. Adding to this, the security implication makes the task of properly monitoring and managing an IoT network a very challenging one.

This project aims to develop a network monitoring system for IoT environments that will offer visualisation of network topology and network metrics in real-time.

Skills required:

- Knowledge of network performance metrics
- Programming skills to create a network visualisation tool
- Very good research and academic scholarship skills

Cloud-enabled energy-efficient personal computing

Description: Cloud computing has been gaining considerable ground in companies and organisations, both small and large. Yet, the uptake of cloud computing for personal use has been limited to specialised applications, such as online file storage and email and more recently online games. The bulk of personal computing applications though are still run locally on PCs. The purpose of this project is to demonstrate the practical environmental benefits of an integrated PC-cloud computing paradigm, where a user's applications are run on the PC or a cloud system depending on an energy consumption/computation profile. It will develop a software toolkit that will allow running PC applications either locally or on a private cloud environment. A series of PC applications with different energy and performance requirements will then be tested to evaluate which characteristics of an application denote where that application should run.

Skills required:

- Good analytical skills
- Good cloud computing knowledge, preferably OpenStack Cloud
- Programming skills to enable PC applications to be either cloned or migrated to a cloud environment
- Very good research and academic scholarship skills

Computation Offloading in IoT Systems

Description: Edge computing is the paradigm by which computation and processing of data is performed at the edge of a network, closer to the source of the data, rather than a remote cloud data centre. As a result, edge computing is now considered the enabling element of the future Internet of Things (IoT). The choice of whether a task should be processed near the edge or offloaded at a distance cloud infrastructure is currently largely based on the original design of the system or best effort. The purpose of this project is to demonstrate the practical benefits

of dynamically offloading the computation of IoT devices to more powerful devices. It is expected to compare existing dynamic offloading mechanisms. The comparison should take into consideration the various tradeoffs, e.g. in terms of performance, communication and energy, when deciding whether edge, cloud or even local IoT devices' resources should be used for performing a task. These tradeoffs depend on the nature of the IoT task, as well as the current load and the condition of the networks utilised for computation offloading, which change continuously.

Skills required:

- Networking skills to create a real testbed consisting of IoT devices, and edge device and a cloud device or Programming and networking skills to create a virtual network either using VMs or containers
- Knowledge of network performance metrics
- Very good research and academic scholarship skills

Computation Offloading for the improvement of multi-layer Anomaly Detection in Industrial IoT (IIoT)

Anomaly detection is the process of identifying unexpected/unusual events in data sets (e.g. network data, IoT monitoring data, sensor data, etc.), which are not considered "normal". Today, many anomaly detection methods are using machine learning or deep learning approaches. However, this comes at a computational cost, especially if the anomaly detection is in resource restricted devices such as IIoT. To address the resource restrictions, many recent approaches consist of multi-layer detection systems, where different complexity models are deployed at each layer of the IIoT architecture (IIoT device, edge device within the same network, cloud). Therefore, at the low processing IIoT device, a simple machine learning method (e.g., decision trees) is used, based largely on local data. If the detection at this level has low confidence, the next layer detection is triggered, i.e. the edge device model, where a more advanced machine learning method is used, and more crucially, also data from multiple IIoT devices can be used. If once again, the confidence is low, the third layer, i.e. cloud, is triggered, where typically a deep learning method is used, and different data types from several sources, hubs and networks can be used.

This multi-layer architecture can be seen as a task offloading system, where the offloading decision is taken only based on the confidence of the detection mechanism's outcome. The purpose of this project is to demonstrate the practical benefits of a more dynamic and context-aware mechanism of offloading based on multiple parameters, such as the confidence of the detection, the processing availability, the network conditions, the energy consumption, etc.

- Programming and networking skills to create a virtual network either using VMs or containers to run an existing anomaly detection mechanism
- Knowledge of network and computer performance metrics
- Very good research and academic scholarship skills

ISEC12: Network Monitoring and Management in IoT

Supervisor: GEORGIA SAKELLARI and GEORGE LOUKAS (<https://group.isec>)

Description: The constant growth of IoT devices is changing the way that networks are monitored and managed. The vast number of devices and the different communication protocols (IP, RFID, NFC, ZigBee, etc.) that they are using has created implications on how they can be managed through a single network monitoring/management software system. Adding to this, the security implication makes the task of properly monitoring and managing an IoT network a very challenging one.

This project aims to develop a network monitoring system for IoT environments that will offer visualisation of network topology and network metrics in real-time.

Skills required:

- Knowledge of network performance metrics
- Programming skills to create a network visualisation tool
- Very good research and academic scholarship skills

ISEC13: Cloud-enabled energy-efficient personal computing

Supervisor: GEORGIA SAKELLARI and GEORGE LOUKAS (<https://group.isec>)

Description: Cloud computing has been gaining considerable ground in companies and organisations, both small and large. Yet, the uptake of cloud computing for personal use has been limited to specialised applications, such as online file storage and email and more recently online games. The bulk of personal computing applications though are still run locally on PCs. The purpose of this project is to demonstrate the practical environmental benefits of an integrated PC-cloud computing paradigm, where a user's applications are run on the PC or a cloud system depending on an energy consumption/computation profile. It will develop a software toolkit that will allow running PC applications either locally or on a private cloud environment. A series of PC applications with different energy and performance requirements will then be tested to evaluate which characteristics of an application denote where that application should run.

Skills required:

- Good analytical skills
- Good cloud computing knowledge, preferably OpenStack Cloud
- Programming skills to enable PC applications to be either cloned or migrated to a cloud environment
- Very good research and academic scholarship skills



Internet of Things and Security (ISEC) project Ideas for students

ISEC12: Network Monitoring and Management in IoT

Supervisor: GEORGIA SAKELLARI and GEORGE LOUKAS (<https://group.isec>)

Description: The constant growth of IoT devices is changing the way that networks are monitored and managed. The vast number of devices and the different communication protocols (IP, RFID, NFC, ZigBee, etc.) that they are using has created implications on how they can be managed through a single network monitoring/management software system. Adding to this, the security implication makes the task of properly monitoring and managing an IoT network a very challenging one.

This project aims to develop a network monitoring system for IoT environments that will offer visualisation of network topology and network metrics in real-time.

Skills required:

- Knowledge of network performance metrics
- Programming skills to create a network visualisation tool
- Very good research and academic scholarship skills

ISEC13: Cloud-enabled energy-efficient personal computing

Supervisor: GEORGIA SAKELLARI and GEORGE LOUKAS (<https://group.isec>)

Description: Cloud computing has been gaining considerable ground in companies and organisations, both small and large. Yet, the uptake of cloud computing for personal use has been limited to specialised applications, such as online file storage and email and more recently online games. The bulk of personal computing applications though are still run locally on PCs. The purpose of this project is to demonstrate the practical environmental benefits of an integrated PC-cloud computing paradigm, where a user's applications are run on the PC or a cloud system depending on an energy consumption/computation profile. It will develop a software toolkit that will allow running PC applications either locally or on a private cloud environment. A series of PC applications with different energy and performance requirements will then be tested to evaluate which characteristics of an application denote where that application should run.

Skills required:

- Good analytical skills
- Good cloud computing knowledge, preferably OpenStack Cloud
- Programming skills to enable PC applications to be either cloned or migrated to a cloud environment
- Very good research and academic scholarship skills

George Samakovitis

Graph Representation and Analysis of Transaction Paths in Cryptocurrency Networks

Cryptocurrency exchange networks are gaining increased popularity in the past 3-4 years, not only because of the speculative value attached to crypto-coins but also because cryptocurrency networks are seen as prime candidates for Money Laundering. Hence, there is great industry interest in understanding how money flows within, and between, these networks (Bitcoin, Ethereum, Litecoin etc).

This project will contribute to modelling and visualising such money flows through describing and visualising sequences of transactions. Using the publicly available Bitcoin Blockchain (or that of another cryptocurrency of your choice) as your dataset, you will apply Graph Analysis and Visualisation techniques to create representation models for transaction sequences in the network.

You will be using Graph Database / Graph Visualisation technologies; Complex Network Analysis techniques; machine learning approaches for community detection and clustering. The project will be supported by the StEP team.

Requirements: Good knowledge of R / Python is required; Graph Databases / Graph Analysis tools ((AllegroGraph, Neo4j, Gephi; Pajek; NetworkX; iGraph or other)

Identification of Customer Groups in Money Exchange Networks with Clustering

The project will aim to investigate how users of financial transaction networks can be categorised into segments by grouping their common transaction behaviour traits. The ultimate business goal is to allow a Payments Service Provider (PSP) to tailor service offerings to different classes of its users to accommodate different needs (for instance lower transaction fees for frequent international transactions; direct debit facilities; other Value-added Services (VAS) depending on volume and usage frequencies etc.)

You will design and deliver your algorithm for group identification and clustering and its user interface, and report on the results from your analysis.

You will use publicly available money transfer datasets (card transactions; payments; money transfers) from sources such as <https://kaggle.com> or other.

You will be using Graph Database technologies; Complex Network Analysis techniques; machine learning approaches for clustering. The project will be supported by the SEP team

Requirements: Good knowledge of R / Python is required; Graph Databases / Graph Analysis tools ((AllegroGraph, Neo4j, Gephi; Pajek; NetworkX; iGraph or other)

A Tool for Stock Portfolio Optimisation based on Investors' Risk Profiles

By definition, stock markets investors seek to maximise their returns by selecting weighted combinations of assets (a *portfolio* of shares, bonds, derivatives and other) that will help them grow their investment and secure healthy annual returns. Finance Theory suggests that since the number of combinations is infinite, investors will typically choose based on how

much they are prepared to risk losing under unfavourable conditions, which, again according to the same theoretical principles, is often correlated to higher rewards (*return on investment*, ROI).

This project will deliver an algorithm and approach with its user interface for recommending portfolio compositions to investors, based on their *risk profiles* and using related stock information, including fundamentals (EBIT; ROA; Price-Earnings Ratios; dividend returns) as well as trading prices and volumes.

You will use publicly available stock data (from Yahoo Finance or other sources) and an R/Python development environment to apply optimisation algorithms in goal-seeking scenarios. The project will be supported by the StEP team.

Requirements: Good knowledge of R / Python is required; Good mathematical / statistics background is desirable; Some understanding of basics on stock markets is desirable.

Sentiment Analysis for Predicting Stock Price Behaviour

Social media platforms provide wealthy information on views, opinions and predictions for the valuation of assets traded in stock markets; the frequency and distribution (geographical, demographic or other) of the use of language terms, traded firms' names and word sequences is often associated with expectations for stock prices, or even attempts of user groups to influence them.

You will use publicly available feeds from one or more social media platforms to build Social Behaviour Graphs and assess key graph characteristics in terms of their correlation to stock prices and / or trading volume.

The project will aim to investigate how the use frequencies and dependencies of terms are associated with prices and other trading characteristics of shares.

You will design and deliver your approach, and report on the results from your analysis. You will be required to use Graph Analysis and Visualisation technologies; Complex Network Analysis techniques; machine learning for basic Natural Language Processing. The project will be supported by the StEP team.

Requirements: Good knowledge of R / Python is required; Graph Analysis & Visualisation tools ((AllegroGraph, Neo4j, Gephi; Pajek; NetworkX; iGraph or other); Good mathematical / statistics background is desirable; Some understanding of basics on stock markets is desirable.

Pattern Recognition Approaches for Identifying and predicting Stock Market Abnormal Returns

The identification of shares that outperform market returns has long been a contentious issue, with experts taking different sides on whether and to what extent this is a feasible task. Arguably, significant effort is being placed lately in applying statistical approaches and machine learning to identify past recurring trends in price and trading volume fluctuations, as possible routes to ultimately predicting stock behaviour.

The project will aim to review and compare machine learning approaches for pattern recognition and clustering in price and trading volume changes in specific shares traded in the Stock Market. You will use publicly available trading data (retrievable from sources such as Yahoo Finance or other) and select, develop, implement and compare machine learning approaches for their performance in predicting stock behaviour. You may use any development platform – R or Python programming languages recommended. The project will be supported by the StEP team.

Requirements: Good knowledge of R / Python is required; familiarity with Machine Learning techniques is desirable; Some understanding of basics on stock markets is desirable.

Recommending Intraday Stock Trading Strategies

By definition, stock markets investors seek to maximise their returns by selecting weighted combinations of assets (a *portfolio* of shares, bonds, derivatives and other) that will help them grow their investment and secure healthy returns. Investment strategies may be *long-term* (buy and hold) or *short-term*, where investors seek to make daily profit through frequent trading. As such, to optimise their short-term (daily) return *risk-prone* investors will typically develop trading strategies, in the form of soft rules of when to sell and when to buy (and the number of shares to trade).

This project will aim to deliver an algorithm and approach with its user interface for recommending such trading strategies (rulesets) to investors, on the basis of past performance and other indicators.

You have at your disposal all available stock information, including fundamentals (EBIT; ROA; Price-Earnings Ratios; dividend returns) as well as trading prices and volumes.

You will use publicly available stock data for a predefined set of stocks (from Yahoo Finance or other sources) and an R/Python development environment to apply optimisation algorithms in goal-seeking scenarios. The project will be supported by the StEP team.

Requirements: Good knowledge of R / Python is required; Good mathematical / statistics background is desirable; Some understanding of basics on stock markets is desirable.

Validation of Factual Information in Knowledge Graph

The project is to validate statements i.e., to check if a given property holds for the subject and object entity in a medical knowledge graph. The core dataset consists of a graph of entities (drugs, diseases and products) and information linking these entities. For this task, students will create an algorithm that takes as input a given triple and returns a trust score. It is expected to provide a trust score for each of the statements (i.e., a numerical value between 0 and 1), where 0 means that they are sure that the statement is false and 1 means that they are sure the statement is true. Students can make use of structured and unstructured information from Internet sources for validating facts and they must ensure that their systems are scalable to handle large amount of input. The systems must be as fast as possible.

Skills required: Knowledge Graph, RDF format, machine learning skills, Java/python language

Dataset: Training data (25k positive and negative examples), Testing data (25k positive and negative examples). The dataset is open: https://hobbitdata.informatik.uni-leipzig.de/SWC_ISWC2019/

Mapping from Tabular Data to Knowledge Graph (KG)

Tables on the Web are the sources of highly valuable data. The addition of semantic information to Web tables may enhance a wide range of applications, such as web search, question answering, and knowledge base (KB) construction. Tabular data to Knowledge Graph (KG) matching is the process of assigning semantic tags from Knowledge Graphs (e.g., Wikidata or DBpedia) to the elements of the table. This is often difficult in practice due to metadata (e.g., table and column names) being missing, incomplete or ambiguous.

The tasks of the project include:

- 1) Assigning a semantic type (e.g., a KG class) to a column of a table;
- 2) Matching a cell to a KG entity;
- 3) Assigning a KG property to the relationship between two columns.

Skills Required: Database, Knowledge Graph, RDF format, machine learning skills, Java/python language.

Dataset: The datasets and ground truths are open: <https://doi.org/10.5281/zenodo.3518539>

Entity Recognition, Linking and Typing for Knowledge Base population

The project includes two tasks. Task 1 consists of identifying Entities in a sentence and create an OWL(W3C Web Ontology Language) individual representing it, link to a reference KB (DBpedia) when possible and assigning a type to such individual. Task 2 consists in producing rdf:type statements, given definition texts. The participants will be given a dataset of sentences, each defining an entity (known a priori).

Skills required: Database, Knowledge Graph, RDF format, machine learning skills, Java/python language

Dataset: The example dataset and ground truth dataset are open:

https://github.com/anuzzolese/oke-challenge-2016/tree/master/GoldStandard_sampleData

Mapping from Tabular Data to Knowledge Graph (KG)

Task Description: Tables on the Web are the sources of highly valuable data. The addition of semantic information to Web tables may enhance a wide range of applications, such as web search, question answering, and knowledge base (KB) construction. Tabular data to Knowledge Graph (KG) matching is the process of assigning semantic tags from Knowledge Graphs (e.g., Wikidata or DBpedia) to the elements of the table. This is often difficult in practice due to metadata (e.g., table and column names) being missing, incomplete or ambiguous.

The tasks of the project include:

- 1) Assigning a semantic type (e.g., a KG class) to a column of a table;
- 2) Matching a cell to a KG Entity;
- 3) Assigning a KG property to the relationship between two columns.

Skills required: Database, Knowledge Graph, RDF format, machine learning skills, Java/python language

Dataset: The datasets and ground truths are open: <https://doi.org/10.5281/zenodo.3518539>

Product matching

Product offers are published on the web together with some textual descriptions and are often accompanied by specification tables, i.e. HTML tables that contain specifications about the offer such as price or the country of origin. Product matching deals with identifying product offers deriving from different websites that refer to the same real-world product. However, the syntactic, structural and semantic heterogeneity among the offers makes product matching a challenging task. In this task, product matching is handled as a binary classification problem: given two product offers decide if they describe the same product (matching) or not (non-matching).

<https://ir-ischool-uos.github.io/mwpd/>

Skills required: this project requires basic proficiency in Python/Java, some knowledge of JSON, machine learning. Python/Java

Data format: The product data is in JSON format. Example data:

<https://drive.google.com/file/d/1S4SwciKmhW4xI2VQPr6fFSaxf6X4Bw48/view>

Learning outcomes: data processing and machine learning techniques, real world problem-solving skills.

Product classification

<https://ir-ischool-uos.github.io/mwpd/>

Same products are often sold on different websites, which generally organise their products into certain categorisation systems. However, such product categorisations differ significantly for different websites, even if they sell similar product ranges. This makes it difficult for product information integration services to collect and organise product offers on the Web. The Product classification task deals with assigning predefined product category labels to product instances (e.g., iPhone 12 is a 'SmartPhone', and also 'Electronics').

Skills required: this project requires basic proficiency in Python/Java, some knowledge of JSON, machine learning. Python/Java

Data format: The product data is in JSON format. Example data:

<https://drive.google.com/drive/folders/1WirDfqGvBYgly27egMx6Om9QeXO6B2UX>

Learning outcomes: data processing and machine learning techniques, real world problem-solving skills.

Cloud-based Wireless Live-Streaming Social Networks:

While a media cloud system can perform multimedia processing and storage, the challenges remain for end-users (e.g., mobile devices and PCs) to receive multimedia streaming from the cloud system with satisfied quality-of-service. To address these challenges, this project will involve an efficient multimedia distribution approach, taking advantage of live-streaming social networks to deliver the media services from the cloud to both desktop and wireless end-users.

Improving Cooperation in P2P Networks:

Maintaining cooperation in peer-to-peer (P2P) networks of selfish adaptive peers is important and challenging. This project involves an algorithm that maintains high levels of cooperation in such a network while performing the collective task of file sharing. The algorithm is relatively simple: peers do not need to store additional trust information about other nodes or to perform significant additional processing, requiring no central servers or authorities.

Task Allocation in Multi-Agent Systems:

Division of labour is a widely studied subject of collective behaviour in natural and artificial systems, contributing to the performance of the multi-agent system. This project deals with effective and distributed mechanisms of a self-organized division of labour occurs to maximising the benefit of the multi-agent system.

Social Internet-of-Things:

Social Internet of Things (SIoT), integrating the social networks and Internet of Things (IoT), leads to heterogeneous interactions of thing to thing, human to human, and human to thing, which in turn generates exploded information. How to characterize the interplay between behaviour spreading and information diffusion in SIoT is essential to predict and manage the information. This project involves the coupled modelling of social interaction and information diffusion processes in SIoT.

Identification of leisure hubs from geo-tagged social media posts

Location planners working for, e.g. a pub or cinema chains regularly need to identify and quantify new store opportunities based on existing leisure activity (“hub effect”). In this context, you will build a set of scored leisure hub polygons for the UK, based on data from location-enabled social media platforms such as Twitter. The methodology should draw from the fields of Natural Language Processing, Information Retrieval and Computational Geometry.

Identification of tourist locations from geo-tagged social media posts

This is a similar project to the project above. Tourists are often underrepresented in location planning models because, unlike residential or worker population, they’re not recorded in more traditional data sources like the census. To fill that gap, you will build a tourist intensity layer for the UK based on data from location-enabled social media platforms such as Twitter. The methodology should draw from the fields of Natural Language Processing, Information Retrieval and Computational Geometry. In the end, you can conduct the project evaluation by comparing your tourist locations with the ones from official data, e.g., the tourism data from local councils.

Insights from Uber movement data

This is the project topic allows you quite a lot of freedom for exploration. We would like to understand what the value of Uber’s open movement data is in the context of location planning or people’s travel pattern. In this project you will analyse the granularity and coverage of the Uber movement data in the UK and come up with specific case studies that can, for instance, help a retailer to better understand the travel patterns to/from a location and urban planners to better understand their cities.

Modelling road networks and pavements with smart navigation

Existing representations and models of road networks focus mainly on traffic and road usage of vehicles. This project focuses on cyclists and pedestrians who partially share road networks with vehicle users but have different travel requirement and behaviours. The project aims at modelling cycling/walking networks that can be integrated into the existing road networks with better navigation strategies for pedestrians and bike users.

Skills required:

- Programming languages such as R, Python or Scala for data manipulation
- Libraries used for processing spatial data such as geopandas (Python) or PostGIS (SQL)
- Basic understanding of unsupervised learning and statistical analysis
- Data visualisations tools for presenting spatial/temporal data
- Basic understandings of graph theory for road network representation

Unsupervised Learning for Multi-view Data

Recently, data collected from various sources or represented by different feature extractors are available in many real-world applications. For example, one document may be translated into different languages; web pages can be represented by different features based on both content and hyperlinks; an image or video can be represented by different visual descriptors, research communities are formed according to research topics as well as co-authorship links and so on. These heterogeneous features that are represented by different perspectives of data are referred to as multiple views. In this project, students will choose multi-view real datasets they are interested in and analyse them by using existing multi-view learning algorithms.

Skill required: advanced mathematical skills, knowledge of fundamental machine learning techniques, good programming skills

Dimension reduction for High-Dimensional Data

High-dimensional data, i.e., data described by a large number of attributes, pose specific two challenges to traditional clustering algorithms. One is the presence of irrelevant and noisy features can mislead the clustering algorithm. The other is in high dimensional data may be sparse due to the curse of dimensionality, making it difficult for an algorithm to find any structure in the data. To deal with these problems, it is crucial to reduce dimensions of high-dimensional data prior to clustering. In this project, several typical dimension reduction techniques will be applied to find suitable low-dimensional representations of real data, comparison and analysis of different techniques are required.

Skill required: advanced mathematical skills, knowledge of fundamental machine learning techniques, good programming skills

Semi-supervise clustering with Ordinal Priors

In reality, one may easily observe that ordinal relationships are ubiquitous among data. An orderly relationship represents that the data a is closer to b than to c . For example, an image of “apple” is often more related to that of “banana” than to that of “ball”; a frame of a video sequence is often more related to its neighbouring frames than to those far away. Such relative order naturally exists among data and it is reliable. Preserving such relative relationships enables us to obtain more accurate clustering that in line with the true data relationship. In this project, students will construct ordinal information as priors and apply existing algorithms to multiple datasets to perform semi-supervised clustering.

Skill required: advanced mathematical skills, knowledge of fundamental machine learning techniques, good programming skills

Robust nonnegative matrix factorization

Finding an optimal data representation is a fundamental problem in many data analysis tasks. A good data representation can typically reveal the latent structure of data and facilitate further data processing. Especially, nonnegative matrix factorization (NMF) technique has been demonstrated to produce superior data representation in many applications such as face recognition, motion segmentation, etc. However, the standard NMF is often afflicted with its sensitivity to outliers and noise in the data. In this project, students will apply multiple existing robust NMF on real datasets they are interested in and analyse them.

Skill required: advanced mathematical skills, knowledge of fundamental machine learning techniques, good programming skills.

Identification of leisure hubs from geo-tagged social media posts

Location planners working for, e.g. a pub or cinema chains regularly need to identify and quantify new store opportunities based on existing leisure activity (“hub effect”). In this context, you will build a set of scored leisure hub polygons for the UK, based on data from location-enabled social media platforms such as Twitter. The methodology should draw from the fields of Natural Language Processing, Information Retrieval and Computational Geometry.

Identification of tourist locations from geo-tagged social media posts

This is a similar project to the project above. Tourists are often underrepresented in location planning models because, unlike residential or worker population, they’re not recorded in more traditional data sources like the census. To fill that gap, you will build a tourist intensity layer for the UK based on data from location-enabled social media platforms such as Twitter. The methodology should draw from the fields of Natural Language Processing, Information Retrieval and Computational Geometry. In the end, you can conduct the project evaluation by comparing your tourist locations with the ones from official data, e.g., the tourism data from local councils.

Insights from Uber movement data

This is the project topic allows you quite a lot of freedom for exploration. We would like to understand what the value of Uber’s open movement data is in the context of location planning or people’s travel pattern. In this project you will analyse the granularity and coverage of the Uber movement data in the UK and come up with specific case studies that can, for instance, help a retailer to better understand the travel patterns to/from a location and urban planners to better understand their cities.

Modelling road networks and pavements with smart navigation

Existing representations and models of road networks focus mainly on traffic and road usage of vehicles. This project focuses on cyclists and pedestrians who partially share road networks with vehicle users but have different travel requirement and behaviours. The project aims at modelling cycling/walking networks that can be integrated into the existing road networks with better navigation strategies for pedestrians and bike users.

Skills required:

- Programming languages such as R, Python or Scala for data manipulation
- Libraries used for processing spatial data such as geopandas (Python) or PostGIS (SQL)
- Basic understanding of unsupervised learning and statistical analysis
- Data visualisations tools for presenting spatial/temporal data
- Basic understandings of graph theory for road network representation

John Ewer

Develop a Web-based or stand-alone application based on the Covid- 19 Aerosol Transmission Estimator and undertake Comparative Analysis of Wells-Riley models with CFD based Infection Risk Modelling.

During the Covid-19 pandemic, it has been important to try to understand the level of risk (of infection) posed to people sharing or entering indoor environments where respired aerosols containing the SARS-CoV-2 virus may exist or persist. One approach to modelling the infection risk is to use a Wells-Riley formulation (or its derivatives), which use a concept of a quanta generation rate to abstract the complexities and current unknowns of infection risk (e.g. number of virions released, decay in virion activity, number of virions needed for infection and the target person susceptibility, etc.) into a form that can be used to evaluate infection risk based on known infection spreading events.

The Covid-19 Aerosol Transmission Estimator (an Excel spreadsheet) is a good example of this. These models typically include simple ventilation rates to represent clearing of the infectivity. Conversely, sophisticated CFD models can model bulk airflows and use respired droplet tracking to give a more complete physical understanding of the droplet dispersion in a built environment. The project involves a development phase where the Aerosol Transmission Estimator is developed as a tool (either as a web based application or a stand-alone tool) and verified as correct AND an analysis phase which will consist of selecting various scenarios of interest, modelling them for infection risk and comparing how the simple Wells-Riley modelling and the sophisticated CFD modelling report the likely infection risk for the scenarios and critically analyse how and under what conditions each form of modelling is appropriate/inappropriate.

Requirements: Good programming experience (any appropriate language), project management, time management, and good analytical skills

References: <https://docs.google.com/spreadsheets/d/16K1OQkLD4BjgBdO8ePj6ytf-RpPMIJ6aXFg3PrIQBbQ/edit#gid=519189277> <https://english.elpais.com/society/2020-10-28/a-room-a-bar-and-a-class-how-the-coronavirus-is-spread-through-the-air.html>

Develop a Stand-Alone Analysis and Summary Report Writing Tool for the Fire Field Modelling Software (SMARTFIRE).

When the Fire Modelling Software (SMARTFIRE, a CFD based fire simulation environment) runs, it creates not only large quantities of field results data (spatially distributed data about the physical solution state in every cell), but also a log file (plain text) and various graph files (as CSV) to inform the user about their simulation. This project will develop a strategy and software tool to extract data from any/all that is available and produce a documented summary of key information and findings from that data. This will include analysis of information such as the status of convergence through the simulation, the physical correctness of the solution as well as visual graphs of selected information. The reporting software tool should also be able to detect and report issues such as non-physically realistic data and also the location of the issue within the simulation domain/geometry (i.e. cell number and physical coordinates). There is considerable potential for the project to be extended to big data analysis to look for issues of interest to the software user.

Requirements: Programming skills to allow file parsing, data extraction, graphing and report writing (any appropriate language), project management, time management, and good analytical skills.

Develop a Mesh Quality Checking Tool for Unstructured Meshes.

The SMARTFIRE Fire Field Modelling software, developed by the University of Greenwich Fire Safety Engineering Group, commonly uses a fully structured mesh. This is a mesh of computational cells or control volumes which have cuboid shapes such that there are 6 faces and that have adjacent faces at right angles to one another – think of a room fully stacked with cereal boxes that just fit inside the walls. In a more advanced mode of usage, the software can also use arbitrary shaped 3D cells from tetrahedral cells, pyramidal cells, hexahedral cells, up to and including cells with a completely arbitrary number of faces. The process of building a quality mesh is challenging and the meshing software may have been developed for another application area, such as finite element simulations – where the meshing rules and requirements are often quite different to those needed for a fine volume fire modelling code.

This project requires the student to investigate what constitutes a quality mesh, what issues would be problematic and develop a mesh quality checking and reporting tool (ideally with visualisation) that can evaluate the quality and correctness of a mesh that is intended for use in SMARTFIRE. Software – in the form of C++ functions – will be provided to read an unstructured mesh from an external file, although full specifications for the file format will also be provided to allow the mesh to be read in any suitable programming language.

Requirements: Good programming experience (any appropriate language), understanding of computational 3D geometry (e.g. finding and comparing normal, face areas, distances from lines to mid faces, etc.), Project management, time management, and good analytical skills.

Develop a stand-alone tool to process and extract data (e.g. monitor point values vs time; graph values) from a set of VTK/VTU field data results files from a CFD simulation.

When the SMARTFIRE fire field modelling software runs, it creates a set of large VTK/VTU files (for each saved time step) containing the spatial layout and field data (values of the physical properties of interest) in each computational cell. In order to understand the data that has been output, it is useful to be able to extract and construct graphs from the numerical data. These could be line graphs of the data between two physical coordinates or a temporal graph of how the data at a single point is changing over time. The project will be based around the (Visualization Toolkit) VTK library, which offers a C++ core, but with wrappers for using the library in Java and Python.

Requirements: Good programming experience (C++, Java or Python), understanding of physical/spatial geometry, interpolation and data visualisation project management, time management, and good analytical skills.

References: <https://vtk.org/>

Develop a design checking and advisory support utility to help users to create fire field modelling scenarios.

When a user creates a fire model case using the SMARTFIRE fire field modelling software, the Case Specification Environment creates and saves human readable scenario definition files. These scenario files contain all the commands, instructions, settings and geometry data that completely describe how the scenario is constructed, how it is configured and how it will run and even how it has been meshed. The intention of this project is to analyse the scenario to report meaningful and useful information to the software user. This will include warnings about poorly/incorrectly configured scenario features. A particularly useful facility would be to perform a “connected space” analysis. This would use the known geometry, and the air-filled spaces defined by it, to map out the geometry block connectivity, to identify (and, ideally, to display) rooms and to check if any disconnected spaces were intentional (e.g. an off-corridor room has a door that has no leaks and is always closed and hence that room will never experience any of the effects of the fire analysis). Source code (in C++) for the scenario file parser will be made available to the project student, although the student can use any suitable programming language to implement the utility. This is a very open-ended project that the interested student could take in a number of directions, e.g. developing better support for meshing, using actual course mesh simulations to generate an understanding of the scenario, feature recognition and tailored guidance, etc.

Requirements: Understanding of physical/spatial geometry and 3D geometry visualisation, suitable programming skills (any appropriate language) to parse the scenario script files, project management, time management, and good analytical skills.



A Tool to Advise CISOs

Supervisor: MANOS PANAOUSIS (<https://www.cyberrisklab.co.uk/>)

Description: This project will investigate different ways to support decisions of Chief Information Security Officers. These are key people within an organisation creating the strategy on cyber and executing it based on their available budgets.

Development: Create a front-end (dashboard) to visualise information relevant to CISOs activities. Back-end will have to support optimisation of their activities.

Indicative references:

- Hooper, V., & McKissack, J. (2016). The emerging role of the CISO. *Business Horizons*, 59(6), 585-591.
- Fitzgerald, T. (2007). Clarifying the roles of information security: 13 questions the CEO, CIO, and CISO must ask each other. *Information Systems Security*, 16(5), 257-263.
- Bradbury, D. (2011). A Day in the Life of a CISO. *Infosecurity*, 8(3), 24-27.
- Kark, K., Penn, J., & Dill, A. (2009). 2008 CISO Priorities: The Right Objectives but the Wrong Focus. *Le Magazine de la Sécurité Informatique*.
- Dawson, M., Burrell, D. N., Rahim, E., & Brewster, S. (2010). EXAMINING THE ROLE OF THE CHIEF INFORMATION SECURITY OFFICER (CISO) & SECURITY PLAN. *Journal of Information Systems Technology & Planning*, 3(6).

Adversarial machine learning against IoT devices

Supervisor: MANOS PANAOUSIS (<https://www.cyberrisklab.co.uk/>)

This project will implement different attacks against functionalities of IoT devices that use machine learning. Such a device could be, for example, a smart camera or an Amazon echo. The student will have to invent ways to compromise these devices following attacks published by the MITRE Association as the ATLAS - Adversarial Threat Landscape for Artificial-Intelligence Systems

<https://github.com/mitre/advmlthreatmatrix>

- Pitropakis, N., Panaousis, E., Giannetsos, T., Anastasiadis, E., & Loukas, G. (2019). A taxonomy and survey of attacks against machine learning. *Computer Science Review*, 34, 100199.
- Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, 0022242920913236.

Cyber forensics visualisation tool

Supervisor: MANOS PANAOUSIS (<https://www.cyberrisklab.co.uk/>)

Description: The purpose of this tool is to create an application that offers advice to cyber forensics investigations. The back end of the application can be purely based on published work of the supervisor while the front-end will have to be developed as part of this project. The related publication is here

<https://ieeexplore.ieee.org/abstract/document/9339971/>

Risk assessment tool using Caldera

Supervisor: MANOS PANAOUSIS (<https://www.cyberrisklab.co.uk/>)

description: The purpose of this tool is to use the Caldera project in combination with an original front end that the student will develop to derive and visualise the risk of any available infrastructure as analysed by the caldera software available here <https://github.com/mitre/caldera>.

Threat intelligence tool

Supervisor: MANOS PANAOUSIS (<https://www.cyberrisklab.co.uk/>)

Description: This project will aim to use and extend the OpenCTI platform in ways that will be discussed with the supervisor. One of the ideas is to use the platform to gather threat intelligence from online resources and classify them so that they can be used to inform any risk assessment methods developed for working from environments. The platform is available here

<https://github.com/OpenCTI-Platform/opencti>.

Cyber insurance tool

Supervisor: MANOS PANAOUSIS (<https://www.cyberrisklab.co.uk/>)

Description: this tool will bring together research from the cyber economics field to develop a tool that enables clients and cyber insurers to communicate with each other prior to underwriting as well as during the contract lifetime, especially in the case of a cyber incident.

Mohammad Majid Al-Rifaie

I. Evolving Deep Neural Networks

The use of deep neural network (DNN) in many real-world applications has seen a significant surge in recent years. One of the key challenges is the design of a suitable neural network that is capable of solving a certain set of problem. In this project, interested student/s will be supervised and guided throughout their project to offer a nature-inspired solution to this problem. This project uses a simple swarm intelligence algorithm, Dispersive Flies Optimisation (DFO), as an efficient gradient-free approach to (1) design a neural network (number of layers, number of neurons in each later, etc), and(if brave)/or (2) train the neural network for an existing dataset. This project could potentially evolve CNN architecture and parameters for text classification purposes.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required: knowledge and use of neural network (e.g. CNN) and existing packages; advanced programming skills; good analytical skills; ability and willingness to read academic papers

Resources: source code for DFO, dataset (e.g. IMDB review polarity dataset with 25,000 highly polar movie reviews), tool for weeding out unsuitable architecture (see below).

Reading:

Stanley, K.O., Clune, J., Lehman, J. and Miikkulainen, R., 2019. Designing neural networks through neuroevolution. *Nature Machine Intelligence*, 1(1), pp.24-35.

Hooman, O.M., al-Rifaie, M.M. and Nicolaou, M.A., 2018, September. Deep Neuroevolution: Training Deep Neural Networks for False Alarm Detection in Intensive Care Units. In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 1157-1161). IEEE.

Weed out the bad architecture by using:

Mellor, J, Turner, J., Storkey, A., and Crowley, E.J., 2020. Neural Architecture Search without Training. arXiv, cs.LG [2006.04647] Link to paper and code and an explanatory video.

II. Protein Folding

Protein structure prediction is an important problem in the field of molecular biology. By understanding the structure of proteins, we get closer to deducing its function. By knowing the structure of proteins, we can accelerate our research into designing new drugs combat major diseases such as Alzheimer, Dementia and Cancer.

In this project, student/s will apply a simple and efficient swarm intelligence algorithm, Dispersive Flies Optimisation (DFO), to the NP-hard protein structure prediction problem using the simplified toy model. This research will shed light into ways to improve the prediction power of the algorithms and then applies the knowledge gained on more complex models.

Using three dimensional Hydrophobic-Polar with Side-Chains model (**3D-HP-SC**). "It was implemented in the Python programming language together with the Biopython 3 framework." <http://biopython.org/> This paper uses 17 real protein sequences extracted from the Protein Data Bank for a benchmark to the PSP problem using the three dimensional Hydrophobic-Polar with Side-Chains model (3D-HP-SC). [ref: Hattori] **LONG HP SEQUENCES:** Long sequences with DFO: "Long sequences of the HP model of protein folding are studied using replica exchange Wang–Landau sampling. We find lower ground state energies than found from earlier simulations using PERM and replica exchange Monte Carlo and, for the first time, extract specific heat curves." [ref: Farris]

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers.

Resources: source code for DFO, dataset (e.g. IMDB review polarity dataset with 25,000 highly polar movie reviews), tool for weeding out unsuitable architecture (see below).

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Reading:

Baker, D. and Sali, A., 2001. Protein structure prediction and structural genomics. *Science*, 294(5540), pp.93-96.

Bošković, B. and Brest, J., 2018. Protein folding optimization using differential evolution extended with local search and component reinitialization. *Information Sciences*, 454, pp.178-199.

Chen, X., Lv, M., Zhao, L. and Zhang, X., 2011. An improved particle swarm optimization for protein folding prediction. *International Journal of Information Engineering and Electronic Business*, 3(1), p.1.

Boiani, M. and Parpinelli, R.S., 2020. A GPU-Based Hybrid jDE Algorithm Applied to the 3D-AB Protein Structure Prediction. *Swarm and Evolutionary Computation*, p.100711.

Hattori, L.T., Gutoski, M., Benítez, C.M.V., Nunes, L.F. and Lopes, H.S., 2020. A benchmark of optimally folded protein structures using integer programming and the 3D-HP-SC model. *Computational biology and chemistry*, 84, p.107192.

Farris, A.C. and Landau, D.P., 2021. Replica exchange Wang–Landau sampling of long HP model sequences. *Physica A: Statistical Mechanics and its Applications*, 569, p.125778.

III. Evolving Deep Neural Networks

Feature selection is a technique applicable to datasets with high dimensionality. It aims to select a subset of features, which enables a classifier to maximise its performance. It is a vital step for a number of classification tasks. As a result, effective feature selection methods are needed to speed up the search and improve the predictive accuracy.

In this project, the student will apply a simple swarm intelligence algorithm, Dispersive Flies Optimisation (DFO) on a selected high-dimensional dataset with the goal of (1) reducing its dimensionality while (2) improving the classification accuracy.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required knowledge and use of classification packages (e.g. sklearn SVM); advanced programming skills; good analytical skills; ability and willingness to read academic papers

Resources: source code for DFO, high-dimensional datasets.

Reading:

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J. and Liu, H., 2017. Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6), pp.1-45.

Alhakbani, H. and al-Rifaie, M.M., 2017, July. Feature selection using stochastic diffusion search. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 385-392).

IV. Solving the Single Row Facility Layout Problem

Single row facility layout problem is an NP-hard permutation problem often found in facility design, factory construction, production optimisation, and other areas. Facility layout problems establish a family of challenging optimisation tasks related to the design and planning of complex production facilities. In general, they look for an optimal arrangement of elements (work and/or repair stations, production support facilities, equipment, machines, robots, etc.) to reduce total production costs. Single row facility layout problem (SRFLP) is a linear placement problem that

seeks optimal linear ordering of facility elements to minimise the sum of distances between each pair of them.

The SRFLP is an appealing problem with a number of practical applications in facility and building design, operations research, circuit design, computer engineering, physical data storage, and many other areas.

The goal of this project is to use a simple nature-inspired algorithm, Dispersive Flies Optimisation (DFO), and find solution to the existing SRFLP benchmark problems.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers.

Resources: source code for DFO, benchmark SRFLP problems .

Reading:

Heragu, S.S., 2008. Facilities design. CRC Press.

Kothari, R. and Ghosh, D., 2012. The single row facility layout problem: state of the art. *Opsearch*, 49(4), pp.442-462.

Di Bari, G., Baioletti, M. and Santucci, V., 2020, July. An experimental evaluation of the algebraic differential evolution algorithm on the single row facility layout problem. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion* (pp. 1678-1684).

V. Solving Bin Packing Problems

The bin packing problem (BPP) can be informally defined in a simple way. We are given n items, each having an integer weight w_j ($j = 1, \dots, n$), and an unlimited number of identical bins of integer capacity c . The objective is to pack all the items into the minimum number of bins so that the total weight packed in any bin does not exceed the capacity.

Bin packing problem has a wide ranging applications which includes the general filling up of containers, creating backup files in media, loading trucks with some weight capacity constraints and more recently the design of field-programmable gate array semiconductor chip.

In this project, student works on using existing set of benchmarks (1D and 2D and/or 3D) and apply a simple and efficient nature-inspired algorithm, Dispersive Flies Optimisation (DFO), as well as Simulated Annealing to find an optimal bin packing approach.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers.

Resources: source code for DFO, benchmark bin packing problems .

Reading:

Oliveira, J.F., Neuenfeldt Júnior, A., Silva, E. and Carravilla, M.A., 2016. A survey on heuristics for the two-dimensional rectangular strip packing problem. *Pesquisa Operacional*, 36(2), pp.197-226.

Delorme, M., Iori, M. and Martello, S., 2018. BPPLIB: a library for bin packing and cutting stock problems. *Optimization Letters*, 12(2), pp.235-250.

VI. Solving Travelling Salesman Problem

Travelling Salesman Problem (TSP) is well-known problem which refers simply to route a salesman through a number of cities where the locations are known. The criterion here is to construct this route by passing every given city once and reaching the starting point by ensuring that the route is

the shortest path. In TSP, there are $(N - 1)!/2$ single solutions where N is the number of cities. At least one of these solutions is the best one with respect to its total distance. The optimum solution of a TSP can be found by a brute-force search algorithm but having a large scale problem may fail the search within the given time constraint and the best solution found in a predetermined time interval might not be a feasible one. In addition, the searching increases exponentially as the problem domain gets bigger, which makes the problem near to impossible to solve by a full search method. Considering such limitations, using the optimisation algorithms focused on the problems such as TSP can be seen as a more effective way to study on large scale domains.

Among the algorithm addressing this problem is Ant Colony Optimisation, which is used for path finding. However, in this project, the research focuses on using another swarm intelligence algorithm (Dispersive Flies Optimisation or DFO) which is used primarily for optimisation. Therefore, this project involves adapting the algorithm to deal with this problem.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers.

Resources: source code for DFO, benchmark TSP libraries from TSPLIB.

Reading:

Varadarajan, S., Whitley, D. and Ochoa, G., 2020, June. Why many travelling salesman problem instances are easier than you think. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference (pp. 254-262).

Chaudhari, K. and Thakkar, A., 2019. Travelling salesman problem: An empirical comparison between aco, pso, abc, fa and ga. In Emerging Research in Computing, Information, Communication and Applications (pp. 397-405). Springer, Singapore.

Chen, J., Ye, F. and Li, Y., 2017, November. Travelling salesman problem for UAV path planning with two parallel optimization algorithms. In 2017 progress in electromagnetics research symposium-fall (PIERS-FALL) (pp. 832-837). Ieee.

VII. FPGA Implementation of a Nature-Inspired Algorithm

The aim of this project is to implement a simple and efficient nature-inspired algorithm, Dispersive Flies Optimisation (DFO) on Field Programmable Gate Arrays (FPGAs).

As an integrated circuit, FPGA is designed to be configured and “programmed” by a programmer or a designer depending on their needs. FPGAs contain an array of programmable logic blocks, where the functionalities are implemented.

When used in embedded systems and their development, they exhibit a remarkable level of flexibility and ability in carrying out tasks.

FPGAs are particularly efficient when dealing with a problem that can be parallelised and given nature-inspired algorithms are exactly that, this project would invite interested students to embark on coding a simple nature-inspired algorithm, Dispersive Flies Optimisation or DFO, on these programmable logic blocks to solve a particular problem.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Starting points:

What’s FPGA programming?

Video: Lesson #1

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers, interest in hardware programming.

Resources: source code for DFO.

Reading:

Da Costa, A.L., Silva, C.A., Torquato, M.F. and Fernandes, M.A., 2019. Parallel implementation of particle swarm optimization on FPGA. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 66(11), pp.1875-1879.

Ettouil, M., Smei, H. and Jemai, A., 2018, December. Particle swarm optimization on FPGA. In *2018 30th International Conference on Microelectronics (ICM)* (pp. 32-35). IEEE.

VIII. GPU and Multi-Processing Implementation of a Swarm Intelligence Algorithm

One of the key advantages of swarm intelligence and nature-inspired computation techniques is their ability to be massively parallel. To help with this, the interested student will be tasked to work on the parallelising the performance of a simple nature-inspired algorithm Dispersive Flies Optimisation or DFO. This process can be done in the following two methods: 1) GPU-assisted parallelisation: using TensorFlow and/or PyTorch, 2) CPU multi-threading using existing libraries (e.g. Uber AI lab's Fiber). The final implementation will be evaluated by comparing different architecture and their associated run-time on a set of standard benchmarks.

Note: Suitable for the strong students taking COMP 1805 Natural Computing

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers.

Resources: source code for DFO, Uber AI lab platform for distributed computing.

Reading:

Introduction to Fiber <https://uber.github.io/fiber>

Boiani, M. and Parpinelli, R.S., 2020. A GPU-Based Hybrid jDE Algorithm Applied to the 3D-AB Protein Structure Prediction. *Swarm and Evolutionary Computation*, p.100711.

Ding, K., Zheng, S. and Tan, Y., 2013, July. A gpu-based parallel fireworks algorithm for optimization. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation* (pp. 9-16).

IX. Network Alignment with Swarm Intelligence

The Network Alignment (NA) problem is a hard combinatorial optimisation problem with a wide range of applications, especially in computational biology. Given two (or more) networks, the goal is to find a mapping between their respective nodes that preserves

the topological and functional structure of the networks. This project attempts to deal with the problem using a simple nature-inspired algorithm Dispersive Flies Optimisation or DFO. There are various resources available for this project. The Implementation of some non-swarm algorithms are available in C++ which can be used in the evaluation stage (e.g. NETAL, L-GRAAL). A high-quality set of networks are available as benchmarks.

Note: Suitable for the strong students taking COMP 1805 Natural Computing. Recommended language: C++

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers.

Resources: source code for DFO, Implementations of other algorithms (see below), benchmark networks (see below).

Protein-Protein Interaction (PPI) networks, representing pairwise biophysical protein interactions of organisms. PPI networks from HINT are accessible from:

Das, J., & Yu, H. (2012). HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC systems biology*, 6(1), 1-12.

Reading:

AntNetAlign: Ant Colony Optimization for Network Alignment:

<https://digital.csic.es/handle/10261/257837>

Hayes, W. B. (2020). An introductory guide to aligning networks using sana, the simulated annealing network aligner. In *Protein-Protein Interaction Networks* (pp. 263-284). Humana, New York, NY.

URL (paper): <https://arxiv.org/pdf/1911.10133.pdf>

URL (code): <https://github.com/waynebhayes/SANA>

Neyshabur, B., Khadem, A., Hashemifar, S., & Arab, S. S. (2013). NETAL: a new graph-based method for global alignment of protein–protein interaction networks. *Bioinformatics*, 29(13), 1654-1662.

URL (paper)

Malod-Dognin, N., & Pržulj, N. (2015). L-GRAAL: Lagrangian graphlet-based network aligner. *Bioinformatics*, 31(13), 2182-2189.

URL (paper): <https://academic.oup.com/bioinformatics/article/31/13/2182/196533>

X. Cuneiform Sign Detection and Reading

Reading and understanding cuneiform script enable us to enter and analyse our ancient past. However, there is a cost: reading age-old clay tablets is time-consuming and requires years of training. According to the *New Scientist*, 3 August 2022, only ~75 people can fluently read cuneiform scripts. In this project, students are challenged with, first reproducing other researchers' results, and then improving it. The first research (see item 1 in the reading list) uses a deep-learning based sign detector that locates and classifies cuneiform signs in images of clay tablets. The second work (see item 2 in the reading list) is responsible for the automatic transliteration and segmentation of Unicode cuneiform glyphs using Natural Language Processing (NLP) techniques.

Students can pick either strands on the project. They can then either improve the work using various machine learning techniques or use swarm intelligence methods to fine-tune some aspects of the work and/or replace a number of the existing building blocks to improve performance.

Note: This is a hard project. Suitable for the strong students taking COMP 1805 Natural Computing.

Skill required: advanced programming skills; good analytical skills; ability and willingness to read academic papers (some beyond the computing world).

Resources: datasets, seed papers, some of the algorithms (see the reading list).

Reading:

Dencker, T., Klinkisch, P., Maul, S.M. and Ommer, B., 2020. Deep learning of cuneiform sign detection with weak supervision using transliteration alignment. *Plos one*, 15(12), p.e0243039.

URL: <https://doi.org/10.1371/journal.pone.0243039>

Gordin, S., Gutherz, G., Elazary, A., Romach, A., Jiménez, E., Berant, J. and Cohen, Y., 2020. Reading Akkadian cuneiform using natural language processing. *PloS one*, 15(10), p.e0240511.

URL: <https://doi.org/10.1371/journal.pone.0240511>

Pagé-Perron, E., Sukhareva, M., Khait, I. and Chiarcos, C., 2017, August. Machine translation and automated analysis of the Sumerian language. In *Proceedings of the Joint SIGHUM Workshop on*

Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (pp. 10-16).

URL: <https://aclanthology.org/W17-2202.pdf>

A gentle introduction to cuneiform: video [Irving Finkel, filmed in the Royal Institute on 18 January 2019]URL (paper)

Malod-Dognin, N., & Pržulj, N. (2015). L-GRAAL: Lagrangian graphlet-based network aligner. *Bioinformatics*, 31(13), 2182-2189.

URL (paper): <https://academic.oup.com/bioinformatics/article/31/13/2182/196533>

Other Projects:

Search for a Complex Problems to Solve with

- Swarm Intelligence
- Evolutionary Computation Methods

Sample problems:

- Problem with multiple optimal solutions: multi-objective optimisation
- Fine-tuning parameters of an algorithm / problem with many parameters
- Solving fundamental problems in mathematics (sum of three cubes, Diophantine equations), physics (in electronics), chemistry (e.g. food optimisation)

Choice of programming languages:

Students are free to choose their strongest programming language. The source-code for the nature-inspired algorithm is available in the following languages:

- Python
- R
- Java
- MATLAB
- C++
- Verilog / VHDL

For projects related to machine learning and deep neural networks (i.e. Projects I, III, VIII, IX, XI), students are encouraged to use Python due to the availability of several relevant libraries. Otherwise, the choice of the strongest language is the overarching advice.

Muhammad Taimoor Khan

Run-time security monitoring of web application

Description: The goal of this project is to use ARMET to monitor web applications at run-time. The application needs to be modelled in ARMET, which will then monitor the application at run-time to detect known and unknown threats at run-time.

Skills required: Interest in programming and logic

Secure-by-design web application

Description: The goal of this project is to develop a translator that translates a given web application into Ur/Web language. The translation will identify those parts that are not protected by design and thus are security risks, which needs to be handled either at design-time or run-time. The UrWeb language is proven to be secure by design against certain classes of security attacks, e.g.,

- Suffer from any kinds of code-injection attacks
- Return invalid HTML
- Contain dead intra-application links
- Have mismatches between HTML forms and the fields expected by their handlers
- Include client-side code that makes incorrect assumptions about the "AJAX"-style services that the remote web server provides
- Attempt invalid SQL queries
- Use improper marshaling or unmarshaling in communication with SQL databases or between browsers and web servers.

Skills required: Interest in programming and logic

Model based security risk assessment of machine learning applications

Description: The goal of this project is to develop risk assessment technique that is capable of detecting known security risks to systems. In fact, we plan to use Jeeves to analyze given APIs of a system implementation and perform formal risk assessment for the API to detect any undesired data flows that may threaten the security of the system. We target APIs written for various machine learning algorithms to test our approach.

Skills required: Interest in programming and logic

Modelling classes of security threats

Description: The goal of this project is to develop a library of classes of security threats that is amenable for verification. The project involves investigation of various classes of attacks based on their causes and effects. An adequate model of classes will be constructed based on their corresponding cause-effect relations. Then, a program will be annotated with the developed models. Based on the annotations, the tool will translate the annotated program into WhyML (<http://why3.lri.fr>) that will semi-automatically prove or disprove, if a given program is free from the model classes of threats. Furthermore, the project will result in

- Library of tactics that are required to prove a program is free from certain security threat classes automatically
- Library of formalized classes of threats that are amenable for verification.

The project aims to develop library of threats based on the attacks described in CAPEC repository that is available at <https://capec.mitre.org>.

Skills required: Interest in programming and logic

Model based intrusion detection

Description: The goal of this project is to develop intrusion detection technique that is capable of detecting known and hypothetical threats based on their models. First, a mapping needs to be established between threat models and their corresponding data representation. Based on the representation, the technique will automatically detect modelled threats without any false alarms. Importantly, the models will include some hypothetical threats that will be exploited by the technique at runtime to detect unknown threats. Furthermore, the project will result in

- Library of formalized known threats that is amenable for verification and
- Library of formalized hypothetical threats that is amenable for verification.

Skills required: Interest in programming and logic

Governance-by-design

Description: The goal of this project is to ensure compliance of a software against various governance ordinance and directives by design. First, a program needs to be annotated with governance ordinance rules. Based on the annotations, the tool will translate the annotated program into WhyML (<http://why3.lri.fr>) that will semi-automatically prove or disprove, if a given program adheres to the annotated governance rules. Furthermore, the project will result in

- Library of tactics that are required to prove governance laws automatically
- Library of formalized governance laws that are amenable for verification.

The project is a motivation from online safety laws in UK and will help to demonstrate such laws. For details, please read <https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws>.

Skills required: Interest in programming and logic

Ethics-by-design

Description: The goal of this project is to ensure compliance of a software against ethical properties by design. First, a program needs to be annotated with ethical properties. Based on the annotations, the tool will translate the annotated program into WhyML (<http://why3.lri.fr>) that will semi-automatically prove or disprove, if a given program adheres to the annotated ethics. Furthermore, the project will result in

- Library of tactics that are required to prove ethical properties automatically
- Library of formalized ethical properties that are amenable for verification.

Skills required: Interest in programming and logic



Internet of Things and Security (ISEC) project Ideas for students

ISEC32: Model based intrusion detection

Supervisor: M TAIMOOR KHAN (<https://group.isec>)

Description: The goal of this project is to develop intrusion detection technique that is capable of detecting known and hypothetical threats based on their models. First, a mapping needs to be established between threat models and their corresponding data representation. Based on the representation, the technique will automatically detect modelled threats without any false alarms. Importantly, the models will include some hypothetical threats that will be exploited by the technique at runtime to detect unknown threats. Furthermore, the project will result in

- Library of formalized known threats that is amenable for verification and
- Library of formalized hypothetical threats that is amenable for verification.

Skills required:

- Interest in programming and logic

ISEC33: Governance-by-design

Supervisor: M TAIMOOR KHAN (<https://group.isec>)

Description: The goal of this project is to ensure compliance of a software against various governance ordinance and directives by design. First, a program needs to be annotated with governance ordinance rules. Based on the annotations, the tool will translate the annotated program into WhyML (<http://why3.lri.fr>) that will semi-automatically prove or disprove, if a given program adheres to the annotated governance rules. Furthermore, the project will result in

- Library of tactics that are required to prove governance laws automatically
- Library of formalized governance laws that are amenable for verification.

The project is a motivation from online safety laws in UK and will help to demonstrate such laws. For details, please read <https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws>.

Skills required:

- Interest in programming and logic

ISEC34: Ethics-by-design

Supervisor: M TAIMOOR KHAN (<https://group.isec>)

Description: The goal of this project is to ensure compliance of a software against ethical properties by design. First, a program needs to be annotated with ethical properties. Based on the annotations, the tool will translate the annotated program into WhyML (<http://why3.lri.fr>) that will semi-automatically prove or disprove, if a given program adheres to the annotated ethics. Furthermore, the project will result in

- Library of tactics that are required to prove ethical properties automatically
- Library of formalized ethical properties that are amenable for verification.

Skills required:

- Interest in programming and logic

Markus Wolf

Maze Solving App

A number of algorithms exist for solving mazes. Some of these are meant to solve mazes where the entire maze is known (seen from above) and others are designed for someone who is inside the maze, which is what this app focuses on. Tremaux's algorithm requires the traveller to mark paths and this app would use location information to track a traveller's movements. As a traveller moves through the maze, the app visually draws the maze and allows the user to mark paths. The app should direct the user through the maze and also advise how to react when a previously placed mark is reached.

Holiday Scheduling System

Everybody deserves holiday, but most organisations need to continue functioning when staff is on holiday leave. This is particularly difficult in organisations which rely on a variety of skills in their day-to-day operations. Most workers also tend to go on leave at similar times of the year, so how can an organisation ensure that it can still operate at these times? An automated holiday scheduling system would capture constraints, such as holiday entitlement, minimum number of staff levels in particular areas and balancing peak times. Constraints would be enforced when approving holiday requests. An algorithm should be devised which generates a possible holiday schedule based on all constraints as well as verifying whether particular requests can be accommodated and, where they can't, suggest alternatives.

Infrared Motion Detector Using Arduino

Using Arduino boards enables us to build electronics projects with relative ease. This project uses infrared sensors to detect motion. However, you will not only detect motion, but employ machine learning to try to distinguish between a number of objects and recognise commonly detected objects by establishing a set of infrared signatures. For example, distinguish between a cat and a person or distinguish

Image Recognition to Aid Fire Evacuations

When a University building has to be evacuated, teaching staff has to report which room they were in and how many students were there and were evacuated. The idea for this project is that image recognition could be employed to count the number of people who are present in a classroom in real-time. When a building is evacuated it is possible to know how many people were present in each room. The system would know whether everyone was evacuated and could provide an all clear for each monitored room.

Machine learning algorithms for Surrogate models and uncertainty optimisation

Multiphysics modelling is widely utilised in microelectronic/power electronic in order to capture the underlying physical process. The need for the identifying the suitable blackbox function capturing relationship between the physical process with some physical quantitative and qualitative dependent variables is vital. Surrogate modelling is very often used to approximate the relationship between physical process and the quantitative and qualitative variables. Furthermore, the uncertainty of the quantitative dependent variable further exacerbates the complexity. If the optimising the quantitative and qualitative variables is required for certain constrains and some uncertainty, then computational optimisation often produce poor optimum for poorly chosen surrogate models. Machine learning models such 'Random forests', 'Naïve Bayes', 'Nearest neighbor', and Artificial neural networks, have tremendous advantages over widely used response surface-based surrogate models. This project aims to utilise the machine learning surrogate models for capturing the relationship between the underlying physical process and the quantitative and qualitative dependent variables.

This project will deliver an algorithm and methodology with its user interface for given data, it automatically generates the appropriate model. With user defined optimisation task, the algorithm will find the optimum.

You will use either use python or R. Python development environment (scikit-learn) can be used to generate the surrogate models and then apply the SciPy for optimisation tasks.

Requirements: Good knowledge of R / Python is required; Good mathematical background is desirable; Some understanding on surrogate modelling is desirable.

Reducing computational complexity in Finite Element modelling by fast matrix to matrix algorithm.

The numerical approximation of partial differential equations often requires numerical discretisation of the domain of interest, hence finer discretization increases the dimensionality of the resulting system. Various numerical discretisation approaches have been proposed in the literature such as finite element method (FEM), finite volume method (FVM), finite difference method (FDM), and boundary element method (BEM). For time dependent FEM analysis, a time invariant system (TIS) arises which is computationally expensive to solve for time varying load and for the larger degree of freedom. Generally, a model order reduction (MOR) technique is widely utilised with a given tolerance for generating a reduced order system. Unfortunately, the resulting reduced system is generally dense in contrast to sparse full order system, hence direct system solver is employed to solve the reduced order system. The computational complexity of direct solver is of $O(n^3)$ in contrast to iterative solver's computational complexity of $O(n^2)$ since in direct solver, matrix $(n \times n)$ to matrix $(n \times n)$ multiplication is required. The idea is to use the fast matrix to matrix multiplication techniques to reduce to computational complexity. The recent developments of fast matrix multiplication algorithms such as Refined laser method (evolved from Strassen's laser method) and

Coppersmith and Winograd's method will reduce the complexity. A gentle introduction on this topic, see [1]

[1] Kevin Hartnett, Matrix Multiplication Inches Closer to Mythic Goal, Quanta magazine, 23 march 2021

The student will develop a Python/R/MatLab code to implement the 'Laser method' and Coppersmith and Winograd's method for matrix-to-matrix multiplication. The data of matrices can be extracted from matrix market (<https://math.nist.gov/MatrixMarket/>)

Requirements: Good knowledge of R / Python is required; Good mathematical background is desirable; Some understanding linear/nonlinear system is desirable

Machine learning algorithms for medical image processing

Image analysis consists of image segmentation, image registration to structure from motion, object recognition and scene understanding. Medical image analysis is a growing field of research for diagnosis of medical image such as x-ray, and CT scan. The medical image detection and recognition deals with the problem of detecting a certain element in the image. Medical image segmentation aims to determine the outline of an organ or an anatomical structure as accurately as possible. A gentle introduction to this topic can be found in the article by Andreas et al [1]. The collection of open access medical images can be extracted from the repositories (<http://Aylward.org>, <https://www.ucl.ac.uk/child-health/about-us/support-services/library/resources-z/medical-image-repositories>, <https://www.smir.ch>). The deep artificial neural network models enable computers to discover complicated patterns in an image [2].

This project will deliver an algorithm and methodology for image recognition and identification of an image.

The student will use either Python or R. Python development environment (scikit-learn) can be used to generate machine learning models. The student can generate the deep learning neural network model in Python or in R for pattern recognition or classification of a medical image data.

[1] Andreas Maier, Christopher Syben, Tobias Lasser, Christian Riess, A gentle introduction to deep learning in medical image processing, Z.Med Phys, 29, 2019, pp 86-101

[2] Alexander Selvikvag Lundervold, and Arvid Lundervold, An overview of deep learning in medical imaging focusing on MRI, Z.Med Phys, 29(2), 2019, pp 102-127

Rafael Martinez Torres

Big Data and London Traffic flow

Currently companies collect big amounts of data concerning the main activities by individuals. One of the most important issues is about traffic flow. By using big data techniques, models can be inferred to detect when and where a bottleneck is to appear so the drivers can be alerted.

Skills: This project requires from you some skills in Python as well as your willingness to learn about data format conversion, statistics...

Resources: Some of sources are available for further processing at <https://www.gov.uk/government/statistical-data-sets/road-traffic-statistics-tra>. Others may require an explicit authorization by competent supervisor.

Music Training-Session Scheduling.

Music associations use to coordinate themselves by exchanging messages on a general-purpose social network like WhatsApp. Consequently, mobile may result overflowed with distracting messages. The challenge is to support these communities with a new specialized utility providing services like: a calendar where participants must adhere to, very much like Doodle systems do; keep a user's profile describing its speciality (drums, flute, clarinet, trumpet...); etc.

Skills: The user interface as well as underlying data technology may vary. A traditional approach should address it into a web /relational database, while a current trend would suggest Mobile platforms apps as well as service providers like Google. In both cases you are required to learn appropriate frameworks on your own.

Microsoft 10X Utility

Microsoft has just released (2019-20) Microsoft 10X as new product line optimized for dual screen devices, so there is an emerging platform waiting for eager developers to write software for.

The domain object is still open at your choice, but a starting point can be a sort of utility tool for sorting pictures according to date, location...

Skills: You are required to be willing to learn both C# language as well as Windows 10X framework on your own. That means that a big amount of your time will be devoted to self-formation before starting to code.

Resources

- <https://developer.microsoft.com/en-us/windows/get-started-windows-10/>
- <https://docs.microsoft.com/en-us/windows/apps/10x/faq>

Machine Learning and Trade Markets.

The current high volatility on Stock Markets makes very difficult to adopt an appropriate strategy. You can first describe a market based on a set of rules of increasing complexity, and then, supported by Machine Learning Techniques, infer a mathematical model in terms of an automaton or a Mealy machine.

Skills: There is Java technology available not to start from zero. You are required to have skills in Java language, and willing to learn on your own IntelliJ IDEA as well as Maven building system. Final documentation scores as much as the final product itself.



ISEC14: Visualising the Internet of Things through the type of data collected (Energy)

Supervisors: RALPH BARTHEL and GEORGIA SAKELLARI (<https://group.isec>)

Description: An increasing number of everyday objects are networked and contain sensors and actuators which enables them to send and receive data. There is little doubt that the emergence of the Internet of Things leads to a number of interesting products with many potentially benefits e.g. energy saving in smart home environments or traffic reduction in smart cities. However, the Internet of Things is also generating a lot of data about our interactions in smart environments (e.g. in the personal home) that we are not necessarily aware of. Frequently this data is created in a passive way e.g. without the people being explicitly aware what data is being collected and how it can be reviewed. The aim of the project is to help making data IoT data visible through a dashboard that visualises what IoT data is being collected in testbed environment (e.g. a kitchen) over a period of time (e.g. a day, a week) and what happens with the data. This can be done through identifying the energy consumption of the IoT devices. This is linked to energy efficiency and to identifying human behaviour through their energy profile.

Skills required:

- Good analytical skills
- Good programming knowledge to create an interactive prototype of a dashboard that supports the research and its concepts
- Good research and academic scholarship skills

ISEC15: Visualising Internet of Things data (Privacy)

Supervisor: RALPH BARTHEL (<https://group.isec>)

Description: An increasing number of everyday objects are networked and contain sensors and actuators which enables them to send and receive data. There is little doubt that the emergence of the Internet of Things leads to a number of interesting products with many potentially benefits e.g. energy saving in smart home environments or traffic reduction in smart cities. However, the Internet of Things with its sensors is also generating a lot of data about our interactions that we are not necessarily aware of. Frequently this data is created in a passive way e.g. without the people being explicitly aware what data is being collected and how it can be reviewed. The aim of the project is to help making data IoT data visible through a dashboard that visualises what IoT data is being collected in testbed environment (e.g. one or more rooms that contain several sensors such as the living lab) over a period of time (e.g. a day, a week). This can be done through identifying the type of data collected, but this is linked to privacy concerns and the user's control over their generated data.

Skills required:

- Good analytical skills
- Good programming knowledge to create an interactive prototype of a dashboard that supports the research and its concepts
- Good research and academic scholarship skills

ISEC16: Monitoring and monetising the use of personal digital data using blockchain technologies

Supervisor: RALPH BARTHEL (<https://group.isec>)

Description: Current models of managing access permissions to personal digital data are frequently very simplistic. The recent example of changes to the EU General Data Protection Regulation (GDPR) demonstrates that in many cases permissions are given on a binary basis (opt-in, opt-out) and there is an underlying assumption that people do not engage with permissions to access their personal data outside of this simple binary (one-time) choice. Recent examples that were discussed extensively in the media such as the the Facebook–Cambridge Analytica data scandal, in which personally identifiable information from Facebook was used to influence voter opinions, show that current models of managing and monitoring access to personal digital data are insufficient. This project explores how blockchain technologies can be used to enable people to continually engage with their personal digital data and give revocable permission to third parties to access the data. How could the sharing of personal digital data be monetised as a commodity? In what ways can transparency be assured how the data is being used? These questions should be explored and evidenced through a prototype of a suitable product using blockchain technologies.

Skills required:

- Very good analytical skills
- Advanced programming knowledge to create an interactive prototype of a blockchain application that would be suitable to monetise access to personal digital data
- Excellent research and academic scholarship skills

ISEC17: A Computer Supported Collaborative Learning (CSCL) tool for Individual and Socially Shared Regulation of Learning (SSRL) during group work activities

Supervisor: RALPH BARTHEL (<https://group.isec>)

Description: In Higher Education settings group work is frequently used as one way to assess student performance. Given the right circumstances and assessment aims group work can be a very effective collaborative learning and assessment activity. There is growing evidence in research literature that socially shared regulation of learning (SSRL) is important for effective and successful collaborative learning activities. SSRL includes aspects such as learners' being aware of their own and others' learning process, the support of the externalisation and sharing of one's own and other's learning processes and prompts for the activation of regulatory processes at different stages during collaborative learning activities. The product of the project will be a technology tool using a suitable implementation technology that can be used to support students' SSRL in collaborative learning activities such as assessments in higher education. The product should be a functional prototype that will be evaluated in an empirical study with an analysis and presentation of the main study results in the project report.

Skills required:

- Very good analytical skills
- Advanced programming knowledge to create an product that is suitable to support SSRL in collaborative learning activities
- Very good research and academic scholarship skills

Blockchain–powered Identity Management in Internet of Things

As individuals disclose personal information before accessing online services in the Internet of Things (IoT), this information are usually exposed to a lot of security issues such as theft. In many IoT environments, databases that store this information also exposed the information without the knowledge of the owners, and third parties can gain access to the information without the owners' consents. The current identity management systems are far from being withstanding these issues. By utilizing blockchain technology, it is possible to design a decentralized and distributed application that makes it possible to protect individuals' identities from theft and fraudulent activities in the IoT. The key idea behind blockchain-power identity management in the IoT is to exploit the characteristics of the blockchain to provide secure identity management.

Keywords: blockchain technology, identity management, internet of things

Skills: Basic knowledge of blockchain technology and internet of things; Basic understanding of identity management

Ransomware Attack Mitigation in Internet of Things: Control and Security

In the Internet of Things (IoT), ransomware attack has recently attracted significant attention due to its successful impacts on IoT infrastructure such as pipeline systems, compromising system security, and disrupting critical processes in the IoT. Hence, mitigating ransomware attack is expected to bolster the security of IoT. To implement a ransomware attack mitigation solution, a distributed control mechanism is required to provide autonomous security capability across the IoT. In this project, we focus on the control and security aspects of ransomware attack mitigation in IoT applications such as smart campuses, smart transportation, smart grids, and smart healthcare.

Keywords: ransomware attack, internet of things, control, security

Skills: Basic understanding of cybersecurity, ransomware attacks, and internet of things

Decentralized Trust and Safety-enabled Supply Chain Platform using Blockchain Technology

Blockchain technology is gradually replacing existing centralized technologies used in supply chain across many Internet of Things (IoT) supported sectors such as logistics due to technical concerns such as user trust issues and inadequate safety of the supply chain. Other concerns include, but not limited to, difficulties to perform door-to-door deliveries of packages and heavily associated costs of delivery. Blockchain technology provides the possibilities of decentralizing and enhancing trust and safety of supply chain for increased transparency, faster processing, secure tracking, reliable shipment and delivery of packages, ease of supply chain auditing, and enhanced fairness and dedication. In this project, we focus on utilising blockchain technology to develop a supply chain platform that satisfies the IoT trust and safety requirements across IoT supported sectors.

Keywords: supply chain, blockchain technology, internet of things, trust, safety

Skills: Basic knowledge of blockchain technology, supply chain, and internet of things

Privacy-preserving Energy Trading and Storage Systems

In this project, the privacy problem for energy trading and storage in energy grids is considered by taking into consideration the energy trading and storage unique features such as identity, costs, and protection nature. With the significant increase in the integration of the internet and energy grids, we aim to develop privacy-preserving techniques for real-world energy trading and storage systems.

Keywords: privacy, energy trading and storage, internet, identity, costs, protection

Skills: Basic understanding of knowledge of energy trading and/or storage; Basic knowledge of privacy and user identity

Collaborative Key Management Techniques for Internet of Things Cryptocurrencies

The distributed nature of the Internet of Things (IoT) cryptocurrencies provides unique opportunities for collaborative key management techniques that can potentially lead to significant mitigation of cyber-attacks such as replay attack, funds stealing, denial of service attack, relay attack, and compromised attack. Collaborative key management design strategies are key issues in the design of secure key management solutions for IoT cryptocurrencies. This project aims to develop collaborative key management techniques for cyber-attacks mitigation, transactions, consensus, incentives, system stability, including optimisation of privacy and performance. The developed techniques will be applied in real-world IoT cryptocurrencies.

Key words: collaborative key management, internet of things cryptocurrencies, transactions, consensus, incentives, system stability, privacy and performance optimisation

Skills: Basic understanding of key management and internet of things cryptocurrencies; Basic knowledge of internet of things and privacy

Blockchain-based Documents Certification Platform

Documents such as certificates and licenses are exposed to forgery and counterfeiting due to mainly lack of high-end verification systems. The tasks of documents verifications in traditional systems make documents certifications and preventing forgery and counterfeiting of certified documents very difficult. Blockchain technology offers promising solutions to help with documents verifications and certifications with trust and transparency. It presents huge benefits to several industries such as banks, healthcare, and educational institutions to verify and certify documents in very less time, effort, and cost. In this project, we focus on providing distributed, secure, and privacy-preserving document certification solutions.

Keywords: documents certification, blockchain technology, security, privacy

Skills: Basic knowledge of blockchain technology, security, and privacy; Basic understanding of document certification

Design of Trust Techniques for Cybersecurity and Operations in Internet of Things

The gradual transformation of Internet of Things (IoT) objects as information sources are revolutionizing and enhancing innovations across the IoT industry. In this project, we consider the trust design problem – of IoT objects – for cybersecurity and operations in the IoT by considering the unique characteristics of IoT, i.e., communication, data, and identity. We aim to develop practical coding

strategies for (I) understanding the rapidly changing IoT attack surface, and (II) optimising the trust-based schemes and solutions for cybersecurity and operations in the IoT.

Keywords: trust, cybersecurity, operations, internet of things, communication, data, identity, attack surface

Skills: Basic knowledge of internet of things, cybersecurity, and data communications

Blockchain-based Healthcare Platform

Blockchain technology has emerged as a new decentralized and distributed paradigm that allows many healthcare applications to directly communicate with each other with little or no human intervention. Functionalities such as patient data management, identity verification, drug traceability, supply chain management, and clinical trials, to name a few, can be integrated to enhance the present healthcare ecosystem. The outcomes of this project will result in significant transformation in the healthcare industry.

Keywords: blockchain technology, healthcare, data management, identity verification, drug traceability, supply chain management, clinical trials

Skills: Basic knowledge of blockchain technology and healthcare supply chain

Peer-to-Peer Secure Situational Awareness and Decision-Making System for the Internet of Things

Providing real-time and secure situational awareness and decision-making has become a huge concern in the Internet of Things (IoT). In peer-to-peer systems, Distributed Ledger Technologies (DLTs) such as Blockchain and Tangle are considered above many other systems to mitigate these concerns due to their distributed and decentralized characteristics. Developing a peer-to-peer system makes it possible to offer secure and reliable situational awareness and decision-making in the IoT. Hence, DLTs are expected to play a crucial role in enhancing today's situational awareness and decision-making among nodes in the IoT. In this project, we focus on designing the security and reliability of situational awareness and decision-making system in the IoT using DLTs.

Keywords: distributed ledger technologies, situational awareness, decision-making, internet of things, security, reliability

Skills: Basic knowledge of distributed ledger technologies and internet of things; Basic understanding of situational awareness and cybersecurity

This is a growing list of ideas for year-end projects and topics of general interest. Feel free to contact me (s.garasto@greenwich.ac.uk) to know more if you are interested in doing any of them.

Note that the list of skills involved does not equal 'skills one needs to have fully developed before starting the project': they can be developed during the project. The only pre-requisite is some familiarity with coding in python.

General topics of interest

- Natural language processing and machine learning
- Data science for social good
- Ethics of AI, machine learning and data science

Prediction analytics and algorithmic fairness

Machine learning algorithms are often used to make predictions about someone's behaviour. For example in the workplace a company might want to predict employee retention after their participation in a training programme. These algorithms can be challenging both from a technical and an ethical point of view. This project has two main objectives: a) compare classification algorithms and apply several feature importance techniques to estimate which are the most relevant features; b) assess and improve the fairness of the algorithm with respect to the protected characteristic (i.e. binary gender) included in the dataset. There are several relevant datasets, such as this one that can be found [on Kaggle](#). Other example datasets that could be used for this task are these ones about students' [engagement](#) and [performance](#).

Skills required: Supervised machine learning, handling unbalanced datasets and missing values, ethics and fairness in machine learning, feature importance/feature selection techniques. The only pre-requisite is some familiarity with coding in python.

Resources: some resources that may be useful for the project are the following:

- [Fairlearn](#): a python toolkit to improve fairness of AI systems
- A good resource is this tutorial on [21 fairness definitions and their politics](#) (FAccT 2018) (the link is also in fair ML book)
- [On the trade-off between fairness and accuracy](#)
- [Some options to deal with missing data](#) (which may be useful for some datasets)

How to start:

- Read the propublica [article](#) and [technical explanation](#) on bias in datasets
- Watch the [21 fairness definitions and their politics](#)
- Explore the potential datasets ([Employee retention](#), students' [engagement](#) and [performance, or others](#)), [the goal is to choose the one you'll be using for the project.](#)
- Load the chosen dataset in python
- Write down your project objectives

Machine learning for Healthcare

There are numerous dataset on Kaggle that could be used to apply machine learning algorithm to healthcare related datasets. For example, you can have a look at [ECG classification](#) or the [non-invasive blood pressure estimation](#). Aside from the prediction task, one could look at the

explainability of the model, which is important in an healthcare setting, for example by identifying which input variables affect the prediction the most.

Skills involved: Supervised machine learning (classification and regression techniques), data visualisation and cleaning, feature importance/selection. The only pre-requisite is some familiarity with coding in python.

How to start:

- Choose the dataset you want to be working with. Explore a few of them and write down the following: what is the input? What is the output? Do I want to study this problem? Options are:
 - [ECG classification \(relevant papers: <http://ecg.mit.edu/george/publications/mitdb-embs-2001.pdf> and <https://ieeexplore.ieee.org/abstract/document/9344639>\)](http://ecg.mit.edu/george/publications/mitdb-embs-2001.pdf)
 - [non-invasive blood pressure estimation](#)
 - [Many other datasets can be found on paperswithcode \(each with relevant papers at the bottom of the page\)](#)
- Once you have chosen the dataset, read the associate relevant papers and look for at least another relevant paper
- Load the chosen dataset in python
- Write down your project objectives

Detection of idioms usage within text

Large natural language processing models have gathered much attention in these past few years. They have been applied to multiple tasks, like machine translation, online searches, and even [writing articles for the Guardian](#), with varying degrees of success. However, these models are still far from perfect. One task that they still struggle with is identifying and understanding idiomatic expressions or, more generally, multi-word expressions, like “gold mine” or “cutting edge”. In this project, you will develop a natural language processing algorithm (most likely, using deep neural network) to detect and understand multi-word expressions in sentences. More background and the dataset to use can be found in [this paper](#) and this [github repo](#).

Skills involved: supervised machine learning, natural language processing, deep neural network. The pre-requisites are some familiarity with coding in python and with neural networks.

How to start:

- Read the [paper](#) that released the dataset and the [problem to solve](#) (task 1 in the github repo)
- Download the [relevant dataset](#) and load it in python
- Read about [token classification](#)
- Write the project objectives

Forecasting future frames in a video

Prediction of future events is a challenging task in artificial intelligence and is relevant to many different fields, like autonomous driving, robotics or weather predictions. One particular instance of this task is that of ‘next-frame prediction’: given a sequence of previous frames in a video, can we predict what the next frame is going to be? Being able to perform such a task suggests that an algorithm has successfully built a representation of what is going on in the video. The goal of this project would be to apply deep neural networks to the ‘next-frame prediction’ task. Networks to use include convolutional and recurrent neural networks. There are several datasets available, for example datasets of [people performing pre-defined actions](#), or of [robots performing actions](#). Reflecting on the potential ethical risks of doing these predictions will also be an important part of this project.

Skills involved: supervised machine learning, deep neural networks (e.g. convolutional or recurrent neural networks), time series. The only pre-requisite is some familiarity with coding in python.

How to start:

- Have a look at the datasets ([people performing pre-defined actions](#), or of [robots performing actions](#)) and choose one that you like most.
- Read some related papers (for example, [Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning](#) and [A review on Deep Learning Techniques for Video Prediction](#)): what techniques can you use to solve this problem?
- Load the chosen dataset in python
- Write down the project objectives

Tuan Nguyen

Face liveness detection

Facial authentication is a popular mechanism on smartphones or laptops because of its convenience and achieving good performance. However, it is vulnerable to various spoofing attacks. Thus, liveness detection in biometrics is required to prevent fraud.

Your task for this project is to develop a machine learning based method to recognise a live human face or a photograph or video which allows quick and accurate spoof detection.

Skill required: programming with python

Object detection from images captured from drones

Our human can perform object detection very easily and effortlessly, but it is difficult for machine to achieve this task. This task becomes more complex from images captured from drones.

The aim of this project is to develop a machine learning based method to detect objects from drones in real time. This project will use public image dataset.

Skill required: programming with python

Fish recognition

The automated recognition of fish in aquarium has many uses, including tourist attraction and educational experience. Traditional methods of obtaining information about fish is asking an expert or by scanning through documentation in the aquarium. However, these methods are time-consuming and limiting factor on the amount of data which can be obtained.

In this project the student will engage with expert team in the field of computer vision to benefit from excellent support to develop an application for fish recognition that provides the user with instant information about a specific fish. The system will capture an image of the fish and display it on the screen. Then, the system recognises the fish species and displays the recognition result to the user.

Skill required: programming with python

Object/Person re-identification

“Given an image/video of a person taken from one camera, re-identification is the process of identifying the person from images/videos taken from a different camera with non-overlapping fields of views. Re-identification is indispensable in establishing consistent labeling across multiple cameras or even within the same camera to re-establish disconnected or lost tracks.” [1]

In this project, you will investigate existing machine learning techniques for Object/Person re-identification, then implement one algorithm for occluded object/person and try it on public data sets.

Skill required: programming with python

[1] Gala, Apurva and Shishir K. Shah. “A survey of approaches and trends in person re-identification.” *Image Vision Comput.* 32 (2014): 270–286.

Auto generating captions for images

One of the most interesting uses of computer vision, from an AI point of view, is image recognition, which gives a machine the ability to interpret the input received through computer vision and categorise what it “sees.”

Question: Can computers understand the context of images?

In this project, you will investigate existing machine learning techniques which allows computers understand images and implement one algorithm and try it on public data sets.

Skill required: programming with python

Colourising images or videos

Colorising black and white images or movies is an old idea dating back to 1902. Many algorithms have been developed to serve this purpose. The main idea of this project will be to develop a tool/software which automatically generates realistic colorizations of Black & White (B&W) photos or B&W videos. An advanced feature allows the users to highlight area with the different colour of their choices.

Skill required: programming with python

Automatic detection of cyberbullying

Social media offer great communication opportunities to everyone, especially for young people. However, they are now considerably affected by cyberbullying. In this project, you will identify a suitable method to detect potentially harmful messages.

Skill required: programming with python

Building a movie/book recommender system to predict the rating score of a movie/book

A recommender system can suggest relevant items to users such as youtube, Netflix basing on what people viewed or liked.

In this project, you will investigate several methods used in recommender systems and identify a suitable method to predict what the rating score that users will give to a movie/book. The movielens dataset (<https://grouplens.org/datasets/movielens/>) or Amazon book dataset will be used for this project.

Skill required: programming with python

Detecting if the provided pairs of questions contain the same meaning

Each month, many people visit and raise questions at StackOverflow or Quora. Therefore, many people can ask similarly worded questions. Multiple questions with the same intent can cause seekers to spend more time finding the best answer to their question, and make writers feel they need to answer multiple versions of the same question.

In this project, you will identify suitable algorithms to detect if there are multiple questions which have the similar meaning.

Skill required: programming with python

Text summarisation

Text summarization is the task producing a concise and fluent summary while preserving key information content and overall meaning. In this project, you will create a tool which allows users to get the summary of an article.

Skill required: programming with python



TUAN VUONG

ISEC06: Data collection for driver training enhancement

Supervisor: TUAN VUONG (<https://group.isec>)

Description: Driver training has always been based on the traditional instructor-trainee model where feedback to the driver is provided only through the observations of the instructor. While Internet of Things and data analytics technologies have already been introduced for drivers, especially for insurance (e.g., the car insurance black boxes), the actual process of learning how to drive does not benefit at all from such technologies.

In this project, the student will develop a mobile application for gathering a large variety of data (speed variations, sudden brakes etc.) that may or may not be indicative to quality of driving. The app will be given to a driver instructor who will use it with the consent of trainees when conducting mock driver's license tests. The output of the project will be a dataset of sensor data labelled by the instructor's scores in each mock test.

Skills required: Mobile app development

ISEC07: Machine learning and data analytics for driver training enhancement

Supervisor: TUAN VUONG (<https://group.isec>)

Description: Driver training has always been based on the traditional instructor-trainee model where feedback to the driver is provided only through the observations of the instructor. While Internet of Things and data analytics technologies have already been introduced for drivers, especially for insurance (e.g., the car insurance black boxes), the actual process of learning how to drive does not benefit at all from such technologies.

In this project, the student will identify the appropriate publicly available datasets, e.g., for telematics datasets or insurance claims, and will develop data analytics that can identify the characteristics of unsafe driving, e.g. for predicting accidents, helping instructors in driver training and providing feedback to trainees.

Skills required: Strong analytical skills and interest in machine learning and statistics

ISEC08: Machine learning in intrusion detection

Supervisor: TUAN VUONG and GEORGE LOUKAS (<https://group.isec>)

Description: Intrusion detection has long benefitted from machine learning. However, which method is the most appropriate is difficult to tell because of the many parameters involved in different cyber attacks and the availability of different features. This project is a performance evaluation study, whereby you will collect a very large number of publicly available cyber attack datasets, possibly complemented by datasets you will create in real or laboratory conditions, and apply on them an equally diverse range of machine learning algorithms and their configurations. The output will be a thorough investigation of what works best in what cyber attack context. The machine learning libraries

can be from educational packages or open source software such as MATLAB, scikit-learn, R, Torch, and OpenAI.

Skills required:

- Interest in machine learning
- Moderate programming skills (e.g., in Python or R)

ISEC09: Reinforcement learning in the Internet of Things

Supervisor: TUAN VUONG (<https://group.isec>)

Description: Artificial intelligence is already leading to new Internet of Things (IoT) services and applications. In IoT scenarios, it may be a machine that learns on its own from the environment or it may be users that help it learn by labelling data with their own feedback. This project is about exploring the applicability of reinforcement learning in IoT applications, with the purpose of introducing a new IoT service that would meaningfully benefit from it. You can start with OpenAI's Proximal Policy Optimization algorithms which have very high performance without being too complex to implement and tune.

Skills required:

- Strong interest in machine learning
- Moderate programming skills

ISEC10: A software and hardware upgrade kit that makes a robot autonomous

Supervisor: TUAN VUONG (<https://group.isec>)

Description: A non-autonomous mobile robot moves based on commands from a user. An autonomous one, instead has high-level goals set to it by a user, but learns on its own how to meet these goals, for example by learning what is a hazard and learning how to avoid it. This project will identify what makes a particular type of robot autonomous (based on size, function and capabilities) and develop and evaluate practically a methodology that achieves this, including the addition of sensors and AI.

Skills required:

- Embedded systems or Python/Java programming skills depending on type of robot chosen.
- Experience in developing algorithms
- Interest in AI

ISEC11: Using AI to compose music

Supervisor: TUAN VUONG (<https://group.isec>)

Description: Composing music can be one of the multiple applications of AI. For example, a previous project has involved creating a deep neural network (benefitting from Long Short Term Memory) that produces pleasant music of a particular style by training it with the music of that style. This project is about exploring the applicability of AI in music.

Skills required:

- Familiarisation with neural networks.
 - Knowledge of music notation, such as ABC Notation

ISEC 45: Automation in self driving car.

Supervisor: TUAN VUONG

Automation in driving has been advanced with AI and machine learning; fully automated or self-driving vehicles have begun to be tested on public roads. Test drivers have clocked over 1.5 million miles in Google's self-driving cars. To automate a component such as braking, the vehicle must generally be aware of what objects surround it and their proximity as well as the vehicles speed. On the other hand, to automate the entire driving process, the vehicle must be able to respond to many challenging factors including road signs and traffic lights (esp. with road maintenance, complex roundabout/crossroad), changes in the road (e.g lane marks, debris) and weather conditions (heavy rain, snow, storm..).

The project will address some aspect of self-driving cars technologies on a simulation platform and software provided by the supervisor.

Skills required:

- Ubuntu
- Programming for ROS (Robot Operating System)
- OpenCV for computer vision.

ISEC 46: Smart home sensor system for intruder detection & energy control

Supervisor: TUAN VUONG

Description: The smart home market is growing in both values and the variety of different IoT devices. In this project, the student will identify a set of smart IoT device including phone, sensors, cameras, appliances to build a solution to collect a data of usage. The dataset can be used for identifying the number of users within the smart home, enhance the intruder detection or energy setting of the system.

Skills required:

- IoT devices
- Strong analytical skills and interest in machine learning and statistics
- Machine learning

ISEC31: Modelling classes of security threats

Supervisor: M TAIMOOR KHAN (<https://group.isec>)

Description: The goal of this project is to develop a library of classes of security threats that is amenable for verification. The project involves investigation of various classes of attacks based on their causes and effects. An adequate model of classes will be constructed based on their corresponding cause-effect relations. Then, a program will be annotated with the developed models. Based on the annotations, the tool will translate the annotated program into WhyML (<http://why3.lri.fr>) that will semi-automatically prove or disprove, if a given program is free from the model classes of threats. Furthermore, the project will result in

- Library of tactics that are required to prove a program is free from certain security threat classes automatically
- Library of formalized classes of threats that are amenable for verification.

The project aims to develop library of threats based on the attacks described in CAPEC repository that is available at <https://capec.mitre.org>.

Skills required:

- Interest in programming and logic

Yasmine Arafa

Analysing Patterns of Affect in Facial Expressions

Generative modelling is an unsupervised machine learning mechanism that is capable of discovering patterns in input data and can be used for detecting such patterns in image data too. Generative Adversarial Networks (GANs) have emerged as a clever way for training generative models using a supervised learning model approach that involves two sub-models: the *generator* model that is trained to generate new samples of data, and the *discriminator* model that tries to classify the samples as either real (originating from the dataset) or fake (generated).

In this project, the aim is to detect the emotional patterns expressed when specific affects are elicited. The project will involve processing video recording of participants reacting to given stimuli and would require an investigation into the different models capable of emotional expression patterns recognition in videos. A possible candidate model to use is a FEV-GAN which (Facial Expression Video GAN), is based on Spatio-temporal Convolutional GANs, that are known to model both expression content and motion in the same network. The goal is to extend the *generator's* functionality to include classification capabilities, training the GAN to empower the generator to function as a classifier of distinct facial motion patterns.

Skill required: Competence in Python, basic knowledge of deep learning and neural networks or willingness to learn ML techniques and ML development frameworks and packages (e.g. Tensorflow, PyTorch).

Resources:

- Facial expression video generation based-on spatio-temporal convolutional GAN: FEV-GAN. <https://www.sciencedirect.com/science/article/pii/S266730532200076X>
- Using computer-vision and machine learning to automate facial coding of positive and negative affect intensity. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6363175/>
- Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms. <https://arxiv.org/pdf/2201.12728.pdf>
- Training machine learning algorithms for automatic facial coding: The role of emotional facial expressions' prototypicality. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0281309>
- Datasets: FERV39k: A Large-Scale Multi-Scene Dataset for Facial Expression Recognition in Videos <https://arxiv.org/abs/2203.09463>, AffectNet <https://paperswithcode.com/dataset/affectnet>
- GAN Tutorial: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans>

Detecting the Golden Ratio in 2D Logo Images

The golden ratio is a design technique to create strong visuals (e.g. company logos) through balance and proportion of the elements in an image. It is used by logo designers as a tool to create aesthetically pleasing compositions that are visually captivating. The ratio is the relationship between 2 or more elements. The Golden ratio will form if the proportion between the elements is 1:1.618 (see resources below for more detail).

In this project, the aim is to detect the existence of the golden ratio in 2D logos. The project will involve processing 2D logo images and classifying those that apply the ratio or not. The implementation may use image processing to detect the existence of the ratio or alternatively use machine learning techniques.

A possible candidate machine learning algorithm is StyleGAN, which is an extension of progressive Generative Adversarial Network (GAN) that is capable of generating high-quality images. StyleGAN can be trained on a dataset that has includes 2D logos that apply the golden ratio and generate images with the ratio.

Generative Adversarial Networks (GANs) have emerged as a clever way for training generative models using a supervised learning model approach that involves two sub-models: the *generator* model that is trained to generate new samples of data, and the *discriminator* model that tries to classify the samples as either real (originating from the dataset) or fake (generated). In this project, the GAN can be trained such that the generator learns to be used as a classifier too.

Skill required: Competence in Python, basic knowledge of machine learning techniques or willingness to learn ML techniques and ML development frameworks and packages (e.g. Tensorflow).

Resources:

- Using The Golden Ratio in Logo Design <https://gingersauce.co/using-the-golden-ratio-in-logo-design-why-how/>
- Nguyen, A., 2023. StyleGAN: In Depth Explanation. <https://pub.towardsai.net/introduction-to-stylegan-ec0a6b0706c>
- GAN Tutorial: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans>