In[19]:=
```
(*---1. Load and Clean Data---*)dsraw = Dataset[Get@Databin["1Ah7XTWPi"]];
dsclean = DeleteDuplicates[
    dsraw[[Complement[Range[Length[Normal[dsraw]]], {1, 3, 5, 9}]]][All, "Data"];
ds =
  dsclean[All, {"title", "proposer", "emailproposer", "descriptionofproject", "comments"}];

(*Extract data for the two different tables*)
allData = Normal[ds[All, Values]];
descData = Normal[ds[All, {"title", "descriptionofproject"}] // Values];


(*---2. Table 1:EVERYTHING (Full Data)---*)
(*We use narrower widths here to fit all 5 columns on one page*)
formattedAll =
  Map[{Pane[#[[1]], ImageSize → 100], (*Title*)Pane[#[[2]], ImageSize → 80], (*Proposer*)
     Pane[#[[3]], ImageSize → 100], (*Email*)Pane[#[[4]], ImageSize → 180], (*Description-
      tighter to fit table*)Pane[#[[5]], ImageSize → 90]  (*Comments*)} &, allData];

Print[Style["Table 1: Full Project Data", "Subtitle"]];
Print[
  Grid[Prepend[formattedAll, {"Title", "Proposer", "Email", "Description", "Comments"}],
   Frame → All, Alignment → {Left, Top}, Spacings → {1, 1}]];


(*---3. Table 2:Project Name& Description Only---*)
(*We use wide widths here so the description is easy to read*)
formattedDesc = Map[{Pane[#[[1]], ImageSize → 120], (*Project Name*)
     Pane[#[[2]], ImageSize → 430] (*Description-Wide*)} &, descData];

Print[Style["\nTable 2: Project Descriptions", "Subtitle"]];
Print[Grid[Prepend[formattedDesc, {"Project Name", "Description"}],
   Frame → All, Alignment → {Left, Top}, Spacings → {1, 1}]];
```

Table 1: Full Project Data

Title	Proposer	Email	Description	Comments
The Science behind Football	Murilo Baptista & Nicolas Rubido	murilo.baptista@abdn.ac.uk	In recent years, football analysis has seen rapid growth and has greatly impacted the evaluation of football performance [1]â[5]. Several clubs are using the insights from the analysis of data of played games to improve on their performance in future matches. Football is a team sport, and the game has 2 teams, with players trying to score a goal. To analyse a match, it is important to understand the dynamics of passes, the positioning	no comments

Machine-learning guided discovery of novel naturally occurring antimicrobials	HAI DENG	h.deng@abdn.ac.uk	Bacterial AMR is now the 3rd-leading cause of death globally, behind ischemic heart disease and stroke. The WHO has classified life-threatening community- and hospital-acquired infections due to Gram-negative bacteria, highlighting carbapenem-resistant Enterobacteriaceae (CRE) (including K. pneumoniae and E. coli) and Acinetobacter baumannii (CRAB) as the most critical AMR global health threats. Recent progress in new antibiotic development has been dominated by incremental improvements in well-established drug classes, such as Î²-lactam antibiotics, already impacted by resistance. There are few drugs in development that target a new mode of action, which is essential for combatting the ongoing emergence of bacteria resistant to existing classes of antibiotics. Antimicrobial natural products, compounds discovered in nature, have been playing critical roles of human medicines and are critical for the development of new therapeutics against multidrug-resistant pathogens. Recent discoveries of antibiotics, including repurposed compounds like hygromycin A,1 and novel agents such as lolamicin 2 and abaucin,3 suggest that many novel antibiotics remain undiscovered.	I am a chemistry academic and will need a colleague from Data Science to help cosupervision.

Advanced Techniques for Time Series Analysis	Paco Perez-Reche	fperez-reche@abdn.ac.uk	With the increasing ability to store and process large amounts of data, it is now possible to track a wide range of processes over time. As a result, there is an abundance of data in the form of time seriesâfrom epidemiological records (e.g. COVID-19 case counts or mortality rates), to physical measurements (e.g. seismic activity), to behavioural data (e.g. the motion of insects). While time series can be visualised relatively easily, identifying patterns, trends, or anomalies in large and complex datasets remains a significant challenge. Advanced techniquesâincluding machine learning methodsâcan automate aspects of this analysis, providing faster and deeper insights. Additionally, alternative representations of time series, such as networks, can reveal structures and properties that may remain hidden in traditional approaches. This project will explore modern, data-driven methods for analysing time series. Depending on your interests, you may investigate techniques such as classification, clustering, or forecasting for time series, or explore novel representations like time series networks and their associated metrics.	No

Developing a Retrieval-Augmented Generation (RAG) System for Extraction and Consolidation of Data from Public Sector PDF Documents	Paco Perez-Reche	fperez-reche@abdn.ac.uk	Government and public sector bodies (e.g., Scottish Health Boards, Local Authorities) routinely publish critical operational and financial data within lengthy, unstructured PDF reports. Analysts are frequently required to synthesize key information across multiple documents to answer specific policy or financial questions. This project proposes the development of a fully automated workflow that combines document scraping and state-of-the-art Artificial Intelligence (AI) techniques to build a robust Retrieval-Augmented Generation (RAG) system, drastically reducing manual analysis time and improving data consistency. The primary goal is to create a functional, scalable prototype that can efficiently answer complex, cross-document queries based solely on official PDF documents. More specifically, you will: 1. Develop a Multi-Document RAG Pipeline: Implement a pipeline capable of processing multiple PDFs from identified domains and generating consolidated answers. 2. Ensure Data Grounding: Guarantee that all extracted summaries and answers are directly traceable and grounded to specific pages and documents within	No.

Statistical and Non-linear Analysis of Electroenceph-alographic signals	NicolÃ¡s Rubido & Murilo S. Baptista	nicolas.rubidoobrer@abdn.ac.uk	Electroencephalography (EEG) is a method to record the spontaneous electrical activity of the brain, which shows a broad range of complex signal behaviours. To understand cortical function, the EEG signal is typically divided into bands by frequency (delta, theta, alpha, beta, and gamma), isolating the rhythmic activity. These bands are related to different physiological functions and produced by different neuronal circuits. However, the bands show different characteristics for different cortical locations and states of mind, such as during the sleep-wake cycle states, which includes the states of wakefulness, rapid-eye movement, and non-rapid-eye movement sleep. Since considerable motivation exists for the development of an adequate statistical model for spontaneous EEG activity, we propose to study the main EEG bands recorded during the sleep-wake state using methods from Statistics and Chaos theory. Those taking on this project will be provided with an EEG dataset, that they will need to filter into frequency bands (pre-processing), obtain the main statistical descriptors for these bands (feature extraction), create surrogate signals for statistical hypothesis testing (statistical modelling), and analyse the phase portraits created by combining some bands (data exploration), which can lead to a	For enquiries, please email nicolas.rubidoobrer@abdn.ac.uk or murilo.baptista@abdn.ac.uk

Artificial Intelligence in Biotechnology	Dr Andrew Angel	andrew.angel@abdn.ac.uk	Background: Biotechnology includes the engineering of biological elements to develop products that can improve human life and the application of machine learning is revolutionizing the field. Inspired by a recent industrial collaboration, one potential project is to use machine learning to accelerate the development of high-temperature RNA thermoswitches. RNA thermoswitches are RNA sequences that can control expression – for example, by only becoming active once a certain temperature has been reached – often via temperature-dependent structural changes . Naturally occurring thermoswitches are optimised for temperatures around 37Â°C and often show undesired basal ('leaky') gene expression at lower temperatures. Owing to their customisable length and base compositions, the ability to design synthetic RNA thermoswitches responsive to higher temperatures (>60Â°C) with reduced leakiness at lower temperatures has significant industrial applications. This project aims to computationally design and evaluate novel RNA thermoswitches specifically optimised for use in the bacterium Bacillus subtilis, which is widely used in industrial biotechnology.	There is also scope for projects in other areas of biotechnology , for example in drug design.

Waves in the Martian atmosphere	Roland Young	roland.young@abdn.ac.uk	There are many different types of atmospheric wave, which are periodic disturbances in atmospheric fields such as temperature or density. These range from sound waves, with wavelengths measured in metres or less, to planetary-scale Rossby waves, with wavelengths measured in thousands of kilometres. Intermediate are gravity waves; these have gravity as the restoring force, similar to a wave on the surface of the ocean. Waves are an important part of the climate system as they transport momentum and energy from place to place without bulk motion of the air. Waves occur in other planetsâ atmospheres as well as on Earth. On Mars, the motion of the atmosphere is a major unknown in our understanding of that planetâs climate, because it is difficult to measure remotely. Identifying waves can tell us something about how the atmosphere transports momentum and energy, linking remote parts of the atmosphere through the general circulation. We can identify waves by studying images of clouds taken in visible or ultraviolet wavelengths, filtered using a high-pass filter to remove large-scale structures.	No

Evaluating Systemic Inequality in Access to debt advice in Scotland	Your Financial Wellbeing (YFW)	m.romano@abdn.ac.uk	About Your Financial Wellbeing (YFW) Your Financial Wellbeing is a social enterprise working to change how financial wellbeing support is offered to people who are struggling with money. We understand that financial stress affects mental, physical and social wellbeing, and can place huge pressure on individuals, families and communities. This is why we design services that are more accessible, reduce barriers, and help people get support sooner to prevent finances spiralling out of control. We provide money advice (omni channel), financial wellbeing workshops, and develop ethical, AI tools that support both clients as well as the advisors delivering frontline services. Project Overview This project will use data-science methods to explore who is and, crucially, who is not accessing money and debt-advice services across Scotland when they need it. By analysing public datasets (e.g., the Financial Conduct Authorityâs Financial Lives Survey, the Scottish Index of Multiple Deprivation, local-authority crisis-fund data, and Office for National Statistics wellbeing datasets), students will	No.

Machine Learning Analysis of Well-Being and Behaviour Data	Hey Joe (company)	m.romano@abdn.ac.uk	Company Background Hey Joey is a digital companion app designed to support parents of neurodivergent children. The app helps parents record daily mood and sensory data, identify patterns, and receive personalised insights to make everyday life easier. Currently in beta testing, Hey Joey aims to launch a scalable version of the platform in early 2026. We are now exploring the use of conversational AI to enhance user experience and engagement. Roles and responsibilities â¢ Clean and prepare anonymised synthetic datasets for analysis. â¢ Apply supervised and/or unsupervised learning techniques to detect correlations and trends (e.g. daily patterns, triggers, or positive periods). â¢ Develop a prototype model or dashboard to visualise findings. â¢ Evaluate ethical and privacy implications of predictive features in family wellbeing contexts. â¢ Present a final report outlining findings, accuracy, and recommendations for next steps.	No

Synthetic Fraud Pattern Modelling & Anomaly Detection for Digital Music Uploads	XYNQ / REDFLAG	m.romano@abdn.ac.uk	Brief Abstract XYNQ Ltd is developing REDFLAG, an AI-driven verification engine designed to detect fraudulent or manipulated music uploads before they reach distributors or streaming platforms. One of the core challenges in early development is the lack of publicly available fraud-labelled datasets, which limits experimentation with anomaly detection techniques. This 12-week MSc project will focus on generating synthetic fraud datasets, designing feature representations for music metadata and behavioural patterns, and prototyping anomaly detection methods (e.g., statistical baselines, clustering, or ML approaches). The outputs will directly support the feasibility stage of REDFLAG and contribute valuable groundwork for our collaborative Interface research project. The work is highly aligned with applied data science and gives the student exposure to real-world anomaly detection and synthetic data engineering challenges. Expected Outcomes	No

Spatial Characterisat-ion of Protein Distribution in Muscle Spindles	Ekkehard Ullner, Mamen Romano and Guy Bewick	e.ullner@abdn.ac.uk	Project Overview Understanding the structure and function of muscle fibres is critical for advancing our knowledge of motor control and neuromuscular coordination. Muscle spindles, specialised sensory structures embedded within skeletal muscles, play a vital role in proprioceptionâthe sense of body position and movementâby providing continuous feedback to the nervous system about muscle stretch and tension. This feedback is essential for maintaining posture, coordinating complex movements, and regulating muscle tone. Despite their critical role, the detailed cellular organisation and protein distribution within muscle spindles remain incompletely understood, particularly in comparison to the more abundant extrafusal fibres that generate muscle force. This project aims to deepen our understanding of the spatial distribution of proteins within intrafusal fibres of muscle spindles, focusing on their proximity to the cell nuclei, and to compare this distribution with that of the surrounding extrafusal fibres. Such analysis can provide valuable insights into the specialised structure–function relationships in muscle spindles, potentially revealing molecular mechanisms underlying muscle	No.

Predicting Length of Stay at Hospital using Natural Language Processing	Mamen Romano, Alexander Groh and Binod Bhattarai	m.romano@abdn.ac.uk	Developing models that predict the length of stay of patients in hospital is crucial for optimising resource allocation, reducing healthcare costs, and improving patient flow. Accurate predictions enable hospitals to better manage bed availability, staffing, and treatment plans, ultimately enhancing the efficiency and quality of patient care. In this project we will apply a Natural Language Processing (NLP) approach to predict the length of stay in hospital of patients. We will use the publicly available MIMIC-III dataset, which comprises deidentified health-related data associated with over forty thousand patients admitted to the Beth Israel Deaconess Medical Center (Boston, USA) between 2001 and 2012. The project aims at generating a summary text report from the existing tabular data for each patient. These summaries will then be converted into numerical representations (embeddings) using a language model such as BioBERT. Finally, these embeddings will be input into a neural network model trained to predict each patientâs expected length of hospital stay. The main objectives of the project are: â¢ To evaluate the effectiveness of Large Language Models (LLMs) in generating	No.

Smart Home Sensing, Time-Series Analytics, and a Digital Twin for Indoor Climate & Energy	Marco Thiel	m.thiel@abdn.ac.uk	The student will design and deploy a small-scale smart-home sensing platform (Arduino/ESP32-based) and build an end-to-end data pipeline to ingest, store, visualise, and analyse multi-sensor time series. The core aim is to turn raw sensor readings into actionable inferences (state estimation, prediction, anomaly detection) andâoptionallyâinto a digital twin of a room/house subsystem (e.g., indoor thermal dynamics), calibrated from the collected data. Core tasks Sensing layer: build 1âN sensor nodes (e.g., temperature/humidity, light; optional: COâ proxy / air quality, motion/occupancy proxy) with timestamped telemetry. Data engineering: stream data to a central machine (laptop/mini-PC) via a message broker/protocol, store in a time-series database, and produce dashboards for live monitoring. Data science layer: develop and evaluate items like: short-horizon forecasting (temperature/humidity), anomaly detection (sensor drift, âwindow	None

Optical Microphone (Laser Mic): Recover Spoken Language/Music from a Photodiode Signal via DSP and System Identification	Marco Thiel	m.thiel@abdn.ac.uk	The student will build a controlled âlaser microphoneâ style experiment and develop signal processing methods to recover spoken language or music from an optically measured signal. The experimental setup is intentionally benign and reproducible: build a small box with a plexiglas window, place a phone playing music (or speech) inside, and record the optical receiver output while the measurement system (laser + photodiode/receiver) is located at a distance. The main scientific objective is to reconstruct intelligible audio on a computer by modelling and inverting the measurement chain, then applying modern denoising/enhancement techniques. Core tasks Experimental design: build the plexiglas–window enclosure and create repeatable measurement protocols (distance, angle, surfaces, ambient light conditions). Data acquisition: collect paired recordings (reference audio from the phone track and the measured optical/receiver signal) to enable objective evaluation.	None

Table 2: Project Descriptions

Project Name	Description
The Science behind Football	In recent years, football analysis has seen rapid growth and has greatly impacted the evaluation of football performance [1]â[5]. Several clubs are using the insights from the analysis of data of played games to improve on their performance in future matches. Football is a team sport, and the game has 2 teams, with players trying to score a goal. To analyse a match, it is important to understand the dynamics of passes, the positioning and the movement of not only the individual players but also the whole team. The interest is to find the emergency of patterns in the collective dynamics of players that lead to goals. In this master project, we will be analysing tracking data provided by the company Statsbomb [6], spanning several decades of tracking data for matches happening in several competitions around the world. The interest will be to construct a network representation of a team, based on the way footballers pass the ball. The players and their average position are the nodes of the network, and the amount of passes one player does to another in the duration of a game representing how well connected a player is with another will be the links of the network. We hope to find network structures that enhances defensive performance, or those that favours wins. We will also do exploratory, and time-series based analysis of the matches, for example, understanding the properties of the time-series constructed from observing a player or a group of players moving in the pitch during the match. The analysis will firstly be focussed on an individual game (which contains already a very large amount of data), but could also be developed in more than 1 game for example when the interest is to compare the playing styles of the different schools of football happening in different continents and competitions. 1. D. Bargiotas, An Analysis of FC Barcelona Through Passing Networks and Event Data Analysis, Master Thesis, University of Aberdeen (2021). 2. J. M. BuldÃº et al., Defining a Historic Football Team: Using Network Science to Analyze Guardiolaâs F.C. Barcelona. 9, 13602 (2019). 3. S. Caicedo-Parada S, et al. Passing Networks and Tactical Action in Football: A Systematic Review. Int J Environ Res Public Heal 17, 6649 (2020). 4. What difference can data make to a football team? - Exasol

Machine-learning guided discovery of novel naturally occurring antimicrobials	Bacterial AMR is now the 3rd-leading cause of death globally, behind ischemic heart disease and stroke. The WHO has classified life-threatening community- and hospital-acquired infections due to Gram-negative bacteria, highlighting carbapenem-resistant Enterobacteriaceae (CRE) (including K. pneumoniae and E. coli) and Acinetobacter baumannii (CRAB) as the most critical AMR global health threats. Recent progress in new antibiotic development has been dominated by incremental improvements in well-established drug classes, such as Î²-lactam antibiotics, already impacted by resistance. There are few drugs in development that target a new mode of action, which is essential for combatting the ongoing emergence of bacteria resistant to existing classes of antibiotics. Antimicrobial natural products, compounds discovered in nature, have been playing critical roles of human medicines and are critical for the development of new therapeutics against multidrug-resistant pathogens. Recent discoveries of antibiotics, including repurposed compounds like hygromycin A,1 and novel agents such as lolamicin 2 and abaucin,3 suggest that many novel antibiotics remain undiscovered. The aim of this Master project is to construct training dataset for predictive modelling of antimicrobial compounds against Gram-negative pathogens. Research objectives: 1. To construct training dataset from existing natural product inventory. 2. To master the use of Chemprop software4 and customize its training dataset for predictive modelling of new albopeptide derivatives. This will enable a rapid, data-informed approach to identify candidates with enhanced antimicrobial activity. References: 1. Leimer, N. et al. (2021) Cell. 184, 5405-5418. 2. MuÃ±oz K.A et al. (2024) Nature. 630, 429-441. 3. Liu G. et al. (2023) Nat. Chem. Biol. 2023, 19, 1342-1350. 4. Heid E. et al. (2024) J. Chem. Inf. Model. 64, 9-17.

Advanced Techniques for Time Series Analysis	With the increasing ability to store and process large amounts of data, it is now possible to track a wide range of processes over time. As a result, there is an abundance of data in the form of time seriesâfrom epidemiological records (e.g. COVID-19 case counts or mortality rates), to physical measurements (e.g. seismic activity), to behavioural data (e.g. the motion of insects). While time series can be visualised relatively easily, identifying patterns, trends, or anomalies in large and complex datasets remains a significant challenge. Advanced techniquesâincluding machine learning methodsâcan automate aspects of this analysis, providing faster and deeper insights. Additionally, alternative representations of time series, such as networks, can reveal structures and properties that may remain hidden in traditional approaches. This project will explore modern, data-driven methods for analysing time series. Depending on your interests, you may investigate techniques such as classification, clustering, or forecasting for time series, or explore novel representations like time series networks and their associated metrics. Potential applications include: analysing the motion and interaction of insects, characterising the complexity of physical systems (e.g. material fracture or water flow in porous media), or examining epidemiological data to better understand disease dynamics or mortality patterns. The specific focus of the project will be chosen in consultation at the start, allowing students some flexibility based on their interests. Further reading: Jan Marcel Kezmann âAll 8 Types of Time Series Classification Methodsâ https://medium.com/mlearning-ai/all-8-types-of-time-series-classification-methods-2c8e4b323ea2⌗2fe7. Aghabozorgi, S., Seyed Shirkhorshidi, A. & Ying Wah, T. Time-series clustering â A decade review. Inf. Syst. 53, 16â38 (2015).

Developing a Retrieval-Augmented Generation (RAG) System for Extraction and Consolidation of Data from Public Sector PDF Documents	Government and public sector bodies (e.g., Scottish Health Boards, Local Authorities) routinely publish critical operational and financial data within lengthy, unstructured PDF reports. Analysts are frequently required to synthesize key information across multiple documents to answer specific policy or financial questions. This project proposes the development of a fully automated workflow that combines document scraping and state-of-the-art Artificial Intelligence (AI) techniques to build a robust Retrieval-Augmented Generation (RAG) system, drastically reducing manual analysis time and improving data consistency. The primary goal is to create a functional, scalable prototype that can efficiently answer complex, cross-document queries based solely on official PDF documents. More specifically, you will: 1. Develop a Multi-Document RAG Pipeline: Implement a pipeline capable of processing multiple PDFs from identified domains and generating consolidated answers. 2. Ensure Data Grounding: Guarantee that all extracted summaries and answers are directly traceable and grounded to specific pages and documents within the source material. 3. Produce Structured Output: Consolidate extracted data into a clean, queryable format (e.g., Markdown or CSV tables). The primary data source is left flexible, but a likely focus will be on Scottish Public Sector bodies (e.g., health boards annual accounts or local authority complaints reports.

Statistical and Non-linear Analysis of Electroencephalographic signals	Electroencephalography (EEG) is a method to record the spontaneous electrical activity of the brain, which shows a broad range of complex signal behaviours. To understand cortical function, the EEG signal is typically divided into bands by frequency (delta, theta, alpha, beta, and gamma), isolating the rhythmic activity. These bands are related to different physiological functions and produced by different neuronal circuits. However, the bands show different characteristics for different cortical locations and states of mind, such as during the sleep-wake cycle states, which includes the states of wakefulness, rapid-eye movement, and non-rapid-eye movement sleep. Since considerable motivation exists for the development of an adequate statistical model for spontaneous EEG activity, we propose to study the main EEG bands recorded during the sleep-wake state using methods from Statistics and Chaos theory. Those taking on this project will be provided with an EEG dataset, that they will need to filter into frequency bands (pre-processing), obtain the main statistical descriptors for these bands (feature extraction), create surrogate signals for statistical hypothesis testing (statistical modelling), and analyse the phase portraits created by combining some bands (data exploration), which can lead to a state-dependent model for the EEG activity during the sleep-wake cycle and improvements to classification accuracy of machine and deep learning algorithms.

Artificial Intelligence in Biotechnology	Background: Biotechnology includes the engineering of biological elements to develop products that can improve human life and the application of machine learning is revolutionizing the field. Inspired by a recent industrial collaboration, one potential project is to use machine learning to accelerate the development of high-temperature RNA thermoswitches. RNA thermoswitches are RNA sequences that can control expression – for example, by only becoming active once a certain temperature has been reached – often via temperature-dependent structural changes . Naturally occurring thermoswitches are optimised for temperatures around 37Â°C and often show undesired basal ('leaky') gene expression at lower temperatures. Owing to their customisable length and base compositions, the ability to design synthetic RNA thermoswitches responsive to higher temperatures (>60Â°C) with reduced leakiness at lower temperatures has significant industrial applications. This project aims to computationally design and evaluate novel RNA thermoswitches specifically optimised for use in the bacterium Bacillus subtilis, which is widely used in industrial biotechnology. Project Aims: Utilise bioinformatics software (primarily NUPACK and RNAStructure) to design and evaluate RNA thermoswitches. Identify structural and sequence features of known thermoswitches (e.g., FourU, ROSE, Hsp17, Listeriaâs PrfA) that correlate with optimal high-temperature performance. Generate and analyse RNA thermoswitch designs predicted to function effectively at specified industrially relevant high temperatures. Examine how RNA thermoswitch structures interact with essential genetic elements, both upstream (e.g., promoters, untranslated regions) and downstream (e.g., ribosome binding sites, coding sequences) and design riboswitches that factor in the functions of these additional elements. Assess computational predictions to identify designs that minimise leakiness while maintaining efficient gene expression at target temperatures.

Waves in the Martian atmosphere	There are many different types of atmospheric wave, which are periodic disturbances in atmospheric fields such as temperature or density. These range from sound waves, with wavelengths measured in metres or less, to planetary-scale Rossby waves, with wavelengths measured in thousands of kilometres. Intermediate are gravity waves; these have gravity as the restoring force, similar to a wave on the surface of the ocean. Waves are an important part of the climate system as they transport momentum and energy from place to place without bulk motion of the air. Waves occur in other planetsâ atmospheres as well as on Earth. On Mars, the motion of the atmosphere is a major unknown in our understanding of that planetâs climate, because it is difficult to measure remotely. Identifying waves can tell us something about how the atmosphere transports momentum and energy, linking remote parts of the atmosphere through the general circulation. We can identify waves by studying images of clouds taken in visible or ultraviolet wavelengths, filtered using a high-pass filter to remove large-scale structures. The Emirates Mars Mission (EMM) has been in orbit around Mars since 9 February 2021. Its Emirates eXploration Imager (EXI) instrument is a visible / ultraviolet camera that is unique among Mars orbiters as it takes regular observation sequences that are suitable for tracking clouds and other small-scale phenomena like waves. This project will use EXI images to look for waves in Marsâ atmosphere. Processed EXI images are publicly available from the EMM Science Data Center. As of 3 December 2024, there are 229 suitable images in 34 sequences between November 2021 and December 2023. Students will identify waves and characterise their wavelengths, frequencies, directions, etc. in terms of location, time of day, season, and so on. It may be possible to measure the dispersion relation for some waves, which is related to the properties of the surrounding atmosphere. This project would suit students interested in image analysis and spectral analysis. A background in physics or mathematics would be quite helpful, but is not essential.

Evaluating Systemic Inequality in Access to debt advice in Scotland	About Your Financial Wellbeing (YFW) Your Financial Wellbeing is a social enterprise working to change how financial wellbeing support is offered to people who are struggling with money. We understand that financial stress affects mental, physical and social wellbeing, and can place huge pressure on individuals, families and communities. This is why we design services that are more accessible, reduce barriers, and help people get support sooner to prevent finances spiralling out of control. We provide money advice (omni channel), financial wellbeing workshops, and develop ethical, AI tools that support both clients as well as the advisors delivering frontline services. Project Overview This project will use data-science methods to explore who is and, crucially, who is not accessing money and debt-advice services across Scotland when they need it. By analysing public datasets (e.g., the Financial Conduct Authorityâs Financial Lives Survey, the Scottish Index of Multiple Deprivation, local-authority crisis-fund data, and Office for National Statistics wellbeing datasets), students will model demographic and geographic patterns of exclusion - comparing factors such as gender, income, housing status and rurality. Findings will help identify underserved groups and inform our approach to designing more accessible and inclusive financial wellbeing services. Methods Students will: â¢ Clean and prepare multiple open/public datasets â¢ Merge datasets and carry out exploratory data analysis (EDA) to identify patterns of exclusion â¢ Use regression modelling (e.g., to examine predictors of non-access) â¢ Produce geospatial visualisations (e.g., local authority or data-zone maps) showing where exclusion or under-provision is concentrated Key Datasets & Links Money and Pensions Service (MaPS) â Financial Wellbeing Survey

Machine Learning Analysis of Well-Being and Behaviour Data	Company Background Hey Joey is a digital companion app designed to support parents of neurodivergent children. The app helps parents record daily mood and sensory data, identify patterns, and receive personalised insights to make everyday life easier. Currently in beta testing, Hey Joey aims to launch a scalable version of the platform in early 2026. We are now exploring the use of conversational AI to enhance user experience and engagement. Roles and responsibilities â¢ Clean and prepare anonymised synthetic datasets for analysis. â¢ Apply supervised and/or unsupervised learning techniques to detect correlations and trends (e.g. daily patterns, triggers, or positive periods). â¢ Develop a prototype model or dashboard to visualise findings. â¢ Evaluate ethical and privacy implications of predictive features in family wellbeing contexts. â¢ Present a final report outlining findings, accuracy, and recommendations for next steps.

Synthetic Fraud Pattern Modelling & Anomaly Detection for Digital Music Uploads	Brief Abstract XYNQ Ltd is developing REDFLAG, an AI-driven verification engine designed to detect fraudulent or manipulated music uploads before they reach distributors or streaming platforms. One of the core challenges in early development is the lack of publicly available fraud-labelled datasets, which limits experimentation with anomaly detection techniques. This 12-week MSc project will focus on generating synthetic fraud datasets, designing feature representations for music metadata and behavioural patterns, and prototyping anomaly detection methods (e.g., statistical baselines, clustering, or ML approaches). The outputs will directly support the feasibility stage of REDFLAG and contribute valuable groundwork for our collaborative Interface research project. The work is highly aligned with applied data science and gives the student exposure to real-world anomaly detection and synthetic data engineering challenges. Expected Outcomes * Synthetic fraud dataset that simulates: * Metadata anomalies (e.g., mismatched names, duplicated ISRCs, suspicious title patterns). * Impersonation attempts (e.g., similar artist names, spoofed profiles). * Behavioural upload irregularities (e.g., burst uploads, repeated catalogue submissions). * Bot-like or mass-generated content behaviours. * Exploratory Data Analysis (EDA) report describing: * Dataset structure and assumptions. * Class distributions and anomaly injection strategies. * Limitations and realism of the synthetic data. * Prototype anomaly detection approaches, such as:

Spatial Characterisation of Protein Distribution in Muscle Spindles	Project Overview Understanding the structure and function of muscle fibres is critical for advancing our knowledge of motor control and neuromuscular coordination. Muscle spindles, specialised sensory structures embedded within skeletal muscles, play a vital role in proprioceptionâthe sense of body position and movementâby providing continuous feedback to the nervous system about muscle stretch and tension. This feedback is essential for maintaining posture, coordinating complex movements, and regulating muscle tone. Despite their critical role, the detailed cellular organisation and protein distribution within muscle spindles remain incompletely understood, particularly in comparison to the more abundant extrafusal fibres that generate muscle force. This project aims to deepen our understanding of the spatial distribution of proteins within intrafusal fibres of muscle spindles, focusing on their proximity to the cell nuclei, and to compare this distribution with that of the surrounding extrafusal fibres. Such analysis can provide valuable insights into the specialised structure–function relationships in muscle spindles, potentially revealing molecular mechanisms underlying muscle sensitivity and control. Research Focus Advances in imaging technologies, particularly fluorescence microscopy, have transformed the way we study cellular structures. Fluorescence techniques allow precise visualisation of protein distribution within cells, making it possible to quantify differences in protein expression across muscle fibre types. In this project, we will work with sets of three complementary images of muscle cross–sections, capturing different cellular features: â¢ Brightfield (light microscopy) image: provides a general structural view of the muscle tissue, highlighting the overall arrangement of muscle fibres. â¢ Nuclear stain (blue fluorescence) image: shows the distribution and density of cell nuclei within the muscle fibres, which helps distinguish intrafusal fibres within spindles from surrounding extrafusal fibres. â¢ Protein of interest (red fluorescence) image: visualises the concentration and spatial distribution of a specific protein, providing insights into its potential functional role within the muscle spindle. Project Objectives â¢ Protein localisation analysis: characterise the

Predicting Length of Stay at Hospital using Natural Language Processing	Developing models that predict the length of stay of patients in hospital is crucial for optimising resource allocation, reducing healthcare costs, and improving patient flow. Accurate predictions enable hospitals to better manage bed availability, staffing, and treatment plans, ultimately enhancing the efficiency and quality of patient care. In this project we will apply a Natural Language Processing (NLP) approach to predict the length of stay in hospital of patients. We will use the publicly available MIMIC-III dataset, which comprises deidentified health-related data associated with over forty thousand patients admitted to the Beth Israel Deaconess Medical Center (Boston, USA) between 2001 and 2012. The project aims at generating a summary text report from the existing tabular data for each patient. These summaries will then be converted into numerical representations (embeddings) using a language model such as BioBERT. Finally, these embeddings will be input into a neural network model trained to predict each patientâs expected length of hospital stay. The main objectives of the project are: â¢ To evaluate the effectiveness of Large Language Models (LLMs) in generating text from tabular data, compared to deterministic template-based methods. â¢ To analyse the influence of various demographic and clinical variables on length-of-stay predictions. â¢ To compare the performance of this integrated NLP approach with a more traditional random forest model. References: Johnson, A., Pollard, T., & Mark, R. (2016). MIMIC-III Clinical Database (version 1.4). PhysioNet. https://doi.org/10.13026/C2XW26. Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics, Volume 36, Issue 4, February 2020, Pages 1234â1240, https://doi.org/10.1093/bioinformatics/btz682

Smart Home Sensing, Time-Series Analytics, and a Digital Twin for Indoor Climate & Energy	The student will design and deploy a small-scale smart-home sensing platform (Arduino/ESP32-based) and build an end-to-end data pipeline to ingest, store, visualise, and analyse multi-sensor time series. The core aim is to turn raw sensor readings into actionable inferences (state estimation, prediction, anomaly detection) andâoptionallyâinto a digital twin of a room/house subsystem (e.g., indoor thermal dynamics), calibrated from the collected data. Core tasks Sensing layer: build 1âN sensor nodes (e.g., temperature/humidity, light; optional: COâ proxy / air quality, motion/occupancy proxy) with timestamped telemetry. Data engineering: stream data to a central machine (laptop/mini-PC) via a message broker/protocol, store in a time-series database, and produce dashboards for live monitoring. Data science layer: develop and evaluate items like: short-horizon forecasting (temperature/humidity), anomaly detection (sensor drift, âwindow openâ events, abnormal heating behaviour), occupancy inference (privacy-preserving proxies), missing-data robustness and sensor fusion. Digital twin extension (recommended): construct a simple physics-based model (e.g., RC thermal network), calibrate parameters from data, and use it for prediction and/or control-oriented what-if simulation. Evaluation / success criteria Quantitative metrics (forecast error, detection delay/false alarm rates, calibration error, energy/comfort trade-offs if control is attempted). Reproducible pipeline (documented deployment, scripts/containers, logged experiments). Clear failure-mode analysis (when/why the models break).

Optical Microphone (Laser Mic): Recover Spoken Language/Music from a Photodiode Signal via DSP and System Identification	The student will build a controlled âlaser microphoneâ style experiment and develop signal processing methods to recover spoken language or music from an optically measured signal. The experimental setup is intentionally benign and reproducible: build a small box with a plexiglas window, place a phone playing music (or speech) inside, and record the optical receiver output while the measurement system (laser + photodiode/receiver) is located at a distance. The main scientific objective is to reconstruct intelligible audio on a computer by modelling and inverting the measurement chain, then applying modern denoising/enhancement techniques. Core tasks Experimental design: build the plexiglas-window enclosure and create repeatable measurement protocols (distance, angle, surfaces, ambient light conditions). Data acquisition: collect paired recordings (reference audio from the phone track and the measured optical/receiver signal) to enable objective evaluation. Signal processing (minimum expected): preprocessing (DC drift removal, filtering, normalisation), demodulation / reconstruction (envelope/Hilbert approaches as appropriate), denoising and enhancement (e.g., Wiener/spectral subtraction; optional learned denoiser on CPU), system identification (estimate an effective transfer function from optical signal â audio). Evaluation: compare reconstruction pipelines using objective measures (SNR improvement, spectral distortion metrics; intelligibility-focused metrics where feasible) and structured listening tests. Ethical boundary The work must be done only on self-generated