Title	Short description in English
Evaluating Large Language Model Robustness Against Jailbreak Attacks in Lithuanian and Developing an Automated Detection Model	This work addresses security issues in large language models (LLMs), focusing on a relatively underexplored area: resilience to "jailbreak" attacks in Lithuanian. While protection mechanisms for English text are being extensively developed, the vulnerability of models in the Lithuanian context remains a significant yet under-evaluated issue. The study will analyze the resilience of popular LLM systems to malicious queries using a specially curated dataset of Lithuanian-language attacks. Based on the findings, an automated attack-detection model tailored to the specifics of Lithuanian text will be designed, trained, and tested, thereby contributing to safer, more reliable use of artificial intelligence systems.
Development and Application of Hierarchical Self-Organizing Maps Architecture for Hierarchical Text Document Clustering	This research investigates and develops a Hierarchical Self-Organizing Maps architecture for hierarchical semantic analysis of large-scale text data. The top level performs coarse document grouping, while subsequent level maps detail data falling into specific cells or clusters. By combining modern text embedding methods with recursive Kohonen network structures, the study creates a tool for visual exploration and navigation of complex document corpora.
Analysis and Prediction of Eurovision Song Contest Voting Patterns Using Machine Learning Methods	This thesis analyzes historical Eurovision Song Contest voting data to identify geopolitical, cultural, or structural patterns. Using machine learning algorithms, the study develops a predictive model aimed at forecasting contest outcomes and points distribution.
Development and Evaluation of an EuroLeague Basketball Match Outcome Prediction Model Based on Team Statistics	This work examines historical performance and match statistics of EuroLeague teams. A machine learning or statistical model is developed and tested to accurately predict the winner of upcoming basketball games.
Prediction of EuroLeague Players' Performance Metrics Using Playing Time and Statistical Indicators	This thesis analyzes the relationship between playing time, standard box-score statistics, and a player's overall efficiency index in the EuroLeague. Regression and predictive modeling techniques are applied to estimate future individual performance.
Influence of Image Count and Object Count on Object Detection Training	The aim is to investigate how two different aspects of training-data quantity affect object detection: the number of unique training images and the number of annotated object instances. Controlled subsets will be created in two complementary settings: approximately fixed image count with varying object count, and approximately fixed object count distributed across different numbers of images. The student will construct and audit the subsets, train object detection models, and analyse mAP50–95, precision, recall, AP by object size and, where appropriate, training-efficiency measures. The study will examine whether broader image diversity or a larger amount of object-level supervision is more important for detector performance.
Influence of Background-Only Images on Object Detection Performance	The aim is to study how object detection training is affected by images that contain none of the target objects. The student will construct an additional background/negative image collection and verify that target classes are absent. Models will be trained with different proportions of such images, for example 0%, 5%, 10%, 20% or more. Two experimental strategies may be compared: replacing part of a fixed training budget with background images, and adding background images while keeping the positive training set unchanged. Evaluation will include mAP50–95, precision, recall, false-positive behaviour and prediction confidence. The study will determine whether, and to what extent, negative examples improve robustness and reduce false detections.
Training Subset Selection Strategies for Data-Efficient Object Detection	The aim is to determine how strongly object detection performance depends not only on the number of training images but also on which images are selected. Several independently generated random subsets, preferably at least five, will provide a baseline distribution and quantify natural subset-selection variability. These will be compared with strategies such as class-balanced selection, diversity- or clustering-based selection and, if the project scope permits, uncertainty- or difficulty-based selection. All

	methods will be evaluated at matched image budgets and under the same training protocol. The analysis will include mAP50–95, class coverage, object distributions, variability across subsets and whether specialised selection methods provide improvements beyond the distribution obtained by random sampling.
Influence of Controlled Object Occlusion on Object Detection Performance	The aim is to quantitatively study how controlled object occlusion affects object detection. Using a controllable synthetic environment or a prepared synthetic-data generation pipeline, paired versions of the same scenes will be generated in which the target object is fully visible or occluded to different degrees by other objects. With full and visible instance masks, the occlusion ratio can be calculated precisely from the visible and original object areas. The student will study how detection performance changes at increasing occlusion levels, for example 0%, 25%, 50% and 75%, and compare models trained on clean, occluded and mixed training data. Evaluation will include mAP, recall, confidence and detection failures as a function of occlusion.
Data Augmentation Methods for Object Detection	The aim is to systematically investigate how different data-augmentation methods affect object detection, particularly under limited training-data conditions. The student will review and compare geometric and photometric transformations, occlusion-based augmentation, Mosaic, MixUp and object-level Copy-Paste. Particular attention will be given to bbox-aware Copy-Paste, where objects are transferred between images while their annotations are updated correctly. Experiments will use matched unique-image budgets and comparable training protocols. Evaluation will include mAP50–95, AP by object size and class, precision, recall, training cost and variability across repeated experiments.
Comparison of Synthetic and Real Image Datasets for Object Detection	This two-year research project will compare object detection models trained on real, photorealistic synthetic and stylized synthetic image datasets. Candidate target objects are stationery items such as markers, erasers, rulers, glue sticks, staplers and highlighters. The student will collect a real dataset and create synthetic datasets in Blender or a similar 3D environment while keeping object classes and class distributions as comparable as possible. The indicative scope is approximately 5,000–6,000 training images per dataset type, around 1,000 validation images and 1,000–1,500 real test images. Final evaluation will be performed on real test images using mAP, precision and recall in order to quantify the synthetic-to-real domain gap and assess the practical usefulness of synthetic data.
Method for Classifying Cybersecurity Requirements into NIST CSF 2.0 Functions Using Large Language Models	This thesis proposes an LLM-based method for classifying cybersecurity requirements into the six functions of the NIST Cybersecurity Framework (CSF) 2.0: Govern, Identify, Protect, Detect, Respond and Recover. In many organizations, this classification is done manually. The CSF 2.0 structure will be used as a labelled dataset.
Method for Detecting Vague Statements in Privacy Policies and Proposing Clear Alternatives According to the GDPR	This thesis proposes a method for detecting vague statements in privacy policies and generating clearer alternatives. Article 12 of the GDPR requires that information about personal data processing is provided in clear and plain language, but many privacy policies contain vague statements. The student will collect and label sentences from public privacy policies as vague or precise. The method will detect vague sentences and use an LLM to suggest a clearer version.
Retrieval-Augmented Generation Approach for Answering Questions About the NIS2 Directive	This work proposes a Retrieval-Augmented Generation (RAG) approach adapted to the structure of legal texts for answering questions about the NIS2 Directive. NIS2 sets cybersecurity obligations for many organizations in the EU, but its text is long and difficult to navigate. A chatbot prototype will be developed using the official NIS2 text, and each answer will include references to the relevant articles.
Method for Automatic Mapping Between NIST CSF 2.0 and NIST SP 800-53 Using Text Embeddings and LLMs	This thesis proposes a two-step method for mapping controls between cybersecurity frameworks: text embeddings select candidate matches, and an LLM confirms or rejects each candidate. Organizations often follow several frameworks at the same time, and creating mappings between them manually takes considerable time. The method will be applied to NIST CSF 2.0 and NIST SP 800-53.
LLM-Based Approach for Building a Knowledge Graph of GDPR Obligations	This thesis proposes an LLM-based approach for building a knowledge graph of GDPR obligations. GDPR obligations are spread across many articles, which makes it difficult to identify which obligations apply to a data controller or a data processor. The

	approach will extract obligations, actors and data types from selected GDPR articles according to a predefined graph schema and store them in a graph database
Knowledge Graph Model of Vulnerabilities and Attack Techniques for Time-Based Asset Risk Analysis	This thesis proposes a knowledge graph model that combines public data on vulnerabilities (CVE), weaknesses (CWE), attack patterns (CAPEC) and attack techniques (MITRE ATT&CK) to analyse how the vulnerability status of software assets changes over time. The model will store publication and update dates and connect this information to a sample inventory of software assets using CPE identifiers. The model will be implemented in a graph database.
LLM-Based Approach for Generating and Validating Security Incident Notifications According to EU Regulatory Requirements	This thesis proposes an LLM-based approach for generating and validating security incident notifications that meet EU regulatory requirements. Organizations must report certain incidents to authorities within short deadlines, and each notification must contain specific information defined by the regulation. The student will select the notification requirements of an EU regulation and prepare incident case notes based on public incident reports, recording which required information is available in each case. In the proposed approach, the LLM first extracts the required information from the case notes, then generates the notification, and finally validates it against a checklist of mandatory elements before it is reviewed by a person
Approach for Translating CIS Benchmark Security Requirements into Automated Python Compliance Checks Using LLMs	This thesis proposes an approach for translating security requirements into automated compliance checks written in Python using LLMs. Many security requirements are still checked manually, and writing automated checks requires time and technical knowledge. The student will select a representative set of requirements from a CIS Benchmark for a Linux operating system that can be checked by reading configuration files, and will prepare compliant and non-compliant sample configuration files. In the proposed approach, the LLM receives the requirement text and the error messages produced when each generated check is executed, and uses them to correct the code.
Method for Classifying Data Practices in Website Privacy Policies According to the GDPR Using Large Language Models	This thesis proposes an LLM-based method for classifying the data practices described in website privacy policies. Privacy policies are long and difficult for users to understand. The student will select a public dataset of privacy policies annotated with data practice categories, and the annotated examples will be used to design the method.
Context-Based Approach for LLM Decision Support in Conflicts Between Incident Response and Service Availability	This work proposes an approach for supporting decisions in security incidents where the response action conflicts with service availability requirements. The approach provides an LLM with structured context about asset importance, dependencies and business rules before it recommends an action. The student will create a set of incident scenarios based on public incident response guidelines. Each scenario will include explicit business rules from which a reference decision can be derived.
Defense Approach Against Poisoning Attacks on a RAG-Based Security Compliance Assistant	This thesis proposes a defense approach against poisoning attacks on a Retrieval-Augmented Generation (RAG) assistant for security compliance. If an attacker inserts false documents into the knowledge base, the assistant may give incorrect security advice. The student will build a RAG assistant based on public security guidelines and reproduce known poisoning attacks described in the literature.
LLM-Based Approach for Detecting Gaps Between Organizational Security Policies and Cybersecurity Framework Requirements	This thesis proposes an LLM-based approach for detecting gaps between an organization's security policies and the requirements of a public cybersecurity framework. Checking whether internal policies cover all framework requirements is usually done manually and takes considerable time. The student will select public security policy templates and create modified versions in which selected requirements are removed, so that the gaps are known in advance. In the proposed approach, the relevant policy sections are retrieved for each framework requirement, and an LLM decides whether the requirement is covered and indicates the supporting text.
Methods for Reducing Hallucinations in Image Captioning	The aim of the thesis is to investigate and compare methods for detecting and reducing hallucinations in image captioning models. The experimental study will use publicly available datasets, including MS COCO and nocaps.

Adapting Foundation Models for Medical Diagnosis	The aim of the thesis is to investigate methods for adapting pretrained Foundation Models to disease diagnosis from medical images and to compare their effectiveness. The publicly available NIH ChestX-ray14 dataset will be used for experimental evaluation.
Generalization of Medical Diagnostic Models	The aim of the thesis is to investigate the ability of medical diagnostic models to maintain their performance when applied to X-ray images from different data sources. The publicly available NIH ChestX-ray14 and RSNA Pneumonia Detection Challenge datasets will be used for experimental evaluation.
Uncertainty Estimation in Medical Diagnostic Models	The aim of the thesis is to investigate methods for estimating uncertainty in medical diagnostic model predictions and their ability to identify unreliable diagnostic decisions. The publicly available NIH ChestX-ray14 dataset will be used for experimental evaluation.
Evaluating Pretrained Generative Restoration Filters as Evasion Vectors Against Lightweight Face Anti-Spoofing Systems	This thesis investigates whether pretrained generative face restoration models (e.g., CodeFormer, GFPGAN) act as unintended evasion vectors against lightweight face presentation attack detection (PAD / anti-spoofing) systems. Edge PAD detectors rely heavily on high-frequency textural cues (screen moiré, print artifacts, specular reflections), which restoration models replace with synthesized, live-like facial micro-textures. The student will construct an evaluation pipeline on standard benchmarks (CASIA-SURF, Replay-Attack), measure error rates (APCER, HTER) across restoration fidelity sweeps, and analyze 2D-FFT spectral attenuation to determine whether consumer-grade restoration filters inadvertently compromise commercial biometric authentication pipelines.
Decision-Threshold Reliability in AI-Generated Image Detection under Limited Calibration Data and JPEG Compression	This thesis investigates the reliability and statistical calibration of decision thresholds in AI-generated image detectors under scarce calibration data and lossy JPEG compression. While synthetic image detectors achieve near-perfect AUROC on academic test sets, practical deployments demand strictly calibrated decision boundaries to bound false alarm rates (e.g., FPR ≤ 1%). The student will evaluate post-hoc calibration strategies and Split Conformal Prediction under cross-generator distribution shifts (Midjourney, Stable Diffusion, DALL-E) and varying compression factors, delivering empirical stability curves and actionable guidelines for reliable deployment in digital forensics and content moderation.
Investigating Accuracy and Explainability in Concept-Based Image Classification with Noisy Concept Annotations	This research examines Concept Bottleneck Models (CBMs), which predict human-interpretable concepts prior to final classification and enable test-time human intervention to correct reasoning errors. In practice, concept annotations are frequently noisy and subjective. The student will investigate the concept leakage paradox and determine the noise threshold at which human concept intervention breaks down, using controlled attribute corruption on the CUB-200-2011 dataset. By comparing joint versus independent training architectures and applying linear probing to intermediate concept representations, the study establishes empirical guidelines for reliable concept-based explainability in safety-critical applications.
Auditing Generative Shortcut Learning in Visual Defect Inspection Using Post-Hoc Explainability	In industrial defect inspection, severe anomaly scarcity is commonly addressed by synthesizing defect samples using generative models (Diffusion, GANs). However, synthetic generation introduces subtle artifacts that may induce shortcut learning, where the classifier recognizes generator artifacts rather than physical defect geometry. Using the MVTec AD benchmark and post-hoc explainability methods (Layer-CAM, Integrated Gradients), the student will conduct a spatial attribution audit comparing attribution heatmaps against pixel-level ground-truth defect masks (Pointing Game hit rate, dilated IoU) to verify whether generative data augmentation preserves genuine spatial localization faithfulness.
The Effect of Prediction-Confidence Filtering on Evaluating Image Classifier Explanation Stability	When benchmarking the stability of explainable AI (XAI) methods under input corruptions (noise, compression), researchers frequently filter the evaluation set to retain only instances where model prediction confidence remains high. This introduces collider stratification bias by restricting evaluation to saturated, easy examples. The student will systematically investigate this phenomenon on the Oxford-IIIT Pets dataset using modern classifiers (ResNet, ConvNeXt) and prominent explainers (Grad-CAM,

	Integrated Gradients). The study aims to map explanation stability across continuous confidence retention thresholds and determine whether confidence filtering artificially inflates stability metrics or inverts explainer rankings.
Evaluating the Evasion Robustness of Generative Latent Watermarking Against Passive and Active Image Transformations	Latent watermarking schemes (e.g., Tree-Ring, Stable Signature) embed cryptographic or statistical signatures into diffusion models to ensure synthetic media provenance. However, watermarked images face downstream social media pipelines and adversarial removal attempts. This research evaluates the robustness of latent watermarks against compound image transformations (resizing, blur, compression) and generative purification attacks (AutoencoderKL re-encoding). The student will construct the Bit Error Rate (BER) and True Positive Rate Pareto frontier against perceptual degradation metrics (LPIPS, SSIM) to establish whether latent watermarks provide secure provenance prior to noticeable visual quality destruction.
Evaluating the Stability and Faithfulness of Image Classification Explanations under Image Corruptions	A key challenge in Explainable AI is that an explanation can remain visually stable without faithfully reflecting the classifier's internal reasoning. When input corruptions (noise, blur, contrast shifts) occur, attribution shifts may indicate explainer instability or a legitimate redirection of classifier focus. This thesis formally decouples explainer stability from classifier faithfulness on the ImageNette benchmark across popular explainers (Grad-CAM, Integrated Gradients, Occlusion Sensitivity). By segregating prediction-preserving from prediction-altering corruptions and employing pixel-removal degradation curves (ROAR/ROAD), the student will determine the operational reliability envelope of visual saliency methods.
Investigating the Effects of Reward Shaping on Sample Efficiency and Generalization in Reinforcement Learning	In reinforcement learning with sparse rewards, reward shaping heuristics are commonly used to improve sample efficiency. Although Potential-Based Reward Shaping (PBRS) guarantees policy invariance in the infinite-sample limit, real-world training is constrained by finite interaction budgets. This thesis investigates how reward shaping impacts sample efficiency and out-of-distribution generalization in deceptive MiniGrid topologies (e.g., cul-de-sacs) using the PPO algorithm. By analyzing state-visitation heatmaps, action-distribution entropy, and navigation success on unseen room configurations, the student will quantify the trade-off between accelerated early learning and topological exploration collapse.
Adaptive Urban Traffic Signal Control Using Deep Reinforcement Learning Under Fluctuating Traffic Demands	This thesis investigates deep reinforcement learning (RL) methods for adaptive traffic light control to minimize vehicle queue lengths and idle delays at urban intersections. Using the open-source SUMO simulation platform and standard RL algorithms (e.g., PPO, DQN), the student will design a multi-lane intersection model and train an intelligent controller responsive to fluctuating traffic densities. Experiments will benchmark RL-based policies against conventional fixed-time and actuated signal controllers under abrupt demand surges and emergency priority interventions.
Evaluating Lightweight Segment Anything Models for Zero-Shot Defect Segmentation in Industrial Inspection	This thesis evaluates the capability of lightweight foundation segmentation models (e.g., MobileSAM, FastSAM) to segment industrial surface defects in a zero-shot regime. Conventional inspection pipelines require extensive per-task labeled data. Using the MVTec AD benchmark, the student will investigate prompt-guided segmentation (point and bounding-box prompts), measure intersection-over-union (IoU) accuracy across structural and texture anomalies, and analyze edge-device latency to determine the feasibility of promptable zero-shot inspection in manufacturing.
Evaluating the Robustness of Lightweight Object Detectors Under Synthetic Weather Degradations	This thesis investigates the robustness of modern lightweight object detectors (e.g., YOLO, RT-DETR) under adverse environmental conditions such as fog, rain, and low illumination. In real-world autonomous systems, optical corruptions often degrade detector reliability. The student will curate controlled corruption benchmarks using public datasets (e.g., COCO, BDD100K), quantify mAP degradation curves across model scales, and evaluate data augmentation strategies to mitigate performance drops in safety-critical edge applications.
Self-Supervised Deep Anomaly Detection for Early Fault Diagnosis in Rotating Manufacturing Equipment	Fault detection in industrial equipment is difficult because failures are relatively rare, while most collected sensor data represent normal operation. Furthermore, changes in machine load or operating speed may be mistakenly classified as faults. In this thesis, student is expected to develop a self-supervised deep-learning framework that learns normal equipment behaviour and detects abnormal conditions without requiring large quantities of labelled fault data. Vibration signals will be processed using time-

	domain and frequency-domain representations such as FFT, STFT, or wavelets. Student will adopt classical anomaly-detection methods, such as Isolation Forest etc. and compared with deep-learning approaches. Using public benchmarked dataset such as CWRU dataset, , the student will investigate models' robustness under fault severities. Suitable public datasets include the Case Western Reserve University Bearing Dataset and the NASA IMS Bearing Dataset.
Semi-Supervised Learning for Online Grooming Detection in Social-Media Conversations under Limited Labelled Data	Online grooming detection remains challenging because grooming behavior develops gradually across conversations, while labelled examples are limited and highly imbalanced. Traditional supervised classifiers may therefore fail to identify subtle or previously unseen behavioral patterns. The goal of this thesis is to develop a context-aware semi-supervised learning framework that can exploit both labelled and unlabelled conversational data to improve grooming detection while reducing dependence on expensive manual annotation. The student will compare conventional machine-learning methods with transformer-based language models and semi-supervised approaches. Conversation-level representations will be investigated to capture contextual and temporal relationships between messages. The experiment will be done using public benchmarked dataset such as PAN 2012 Sexual Predator Identification Corpus andcurated PJ Dataset.
Explainable Deep Learning for Depression Detection Severity Estimation	This thesis aims to develop an explainable deep-learning framework for estimating depression symptom severity from clinical interview videos. Visual and behavioural features such as facial action units, facial landmarks, gaze direction, and head pose will be extracted and represented as temporal sequences from publicly available dataset such as DAIC-WOZ dataset. The student will investigate the baseline performance of learning models such as GRU, LSTM, temporal CNN, or transformer-based models for severity prediction. Furthermore, student should employ explainability techniques such as SHAP, Integrated Gradients, attention visualization, and feature ablation to identify influential behavioral patterns.
Uncertainty-Aware and Explainable Deep Learning for Reliable Skin Lesion Classification under Class Imbalance	Deep-learning models for skin-lesion classification can produce highly confident predictions even when images are ambiguous, noisy, or belong to underrepresented diagnostic classes. This thesis aims to investigate uncertainty-aware and explainable deep-learning methods for improving the reliability of automated skin-lesion classification. The student will develop and compare deep-learning architectures such as EfficientNet, ResNet, and Vision Transformers for multi-class classification of dermoscopic images. Data augmentation and class-balancing techniques will be applied to reduce the effects of dataset imbalance. To estimate predictive uncertainty, approaches such as Monte Carlo Dropout, Deep Ensembles, and Evidential Deep Learning will be investigated and compared. Model performance will be evaluated using classification metrics together with uncertainty and calibration measures. Experiments will be conducted using public benchmark datasets from the ISIC Archive and ISIC Challenges.
Retrieval-Augmented Large Language Models for Evidence-Grounded Clinical Question Answering and Hallucination Reduction	Large Language Models can answer complex biomedical questions but may generate incorrect or unsupported information, commonly referred to as hallucination. This limitation is particularly important in healthcare, where generated information should be factual and traceable to reliable evidence. This thesis aims to develop and evaluate a Retrieval-Augmented Generation framework to improve factuality, evidence grounding, and reliability of biomedical question answering. The student is expected to compare a standalone open-source LLM with several RAG configurations. Biomedical documents from public datasets such as PubMedQA, MedQA, and MedMCQA will be used. This thesis will investigate semantic retrieval, hybrid semantic-keyword retrieval, reranking, different chunk sizes, and different numbers of retrieved documents.
Deep Learning for Lung Nodule Localization and Malignancy-Risk Assessment in Thoracic CT	Automated lung-nodule analysis is affected by variations in CT image quality, including differences in resolution, noise, reconstruction parameters, and slice thickness. This thesis aims to develop a deep-learning framework for lung-nodule localization or segmentation and malignancy-risk assessment. Image-quality based on controlled degradation will be used to test robustness of deep learning models. This thesis will evaluate and compare conventional and quality-aware models on publicly benchmarked dataset such as LIDC-IDRI to determine deep model reliability.

Comparative Evaluation of Lightweight Deep Learning Models for Retinal Disease Detection, Classification, and Lesion Analysis	Deep-learning methods have demonstrated strong performance in automated retinal disease analysis; however, many state-of-the-art models are computationally intensive and may be unsuitable for deployment on resource-constrained devices or in clinical environments with limited computing infrastructure. This thesis aims to evaluate and compare lightweight deep-learning architectures for automated retinal disease classification, with particular emphasis on balancing predictive performance and computational efficiency. The student will investigate lightweight architectures such as MobileNetV2/V3, EfficientNet, ShuffleNet, SqueezeNet, GhostNet, and MobileViT for classifying retinal fundus images. Data preprocessing, augmentation, transfer learning, and class-balancing techniques will be applied to improve model robustness. Model-compression approaches such as pruning, quantization, and knowledge distillation will be investigated to further reduce computational requirements. Public benchmark retinal datasets such as IDRiD will be used for experimental evaluation.
Segmentation and Classification of Colorectal Lesions using Attention-Based Deep Neural Networks	Automated analysis of colorectal histopathology images is challenging because lesion boundaries can be irregular and visually similar tissue patterns may belong to different pathological categories. This thesis aims to develop and evaluate attention-based deep neural networks for simultaneous colorectal lesion segmentation and classification. The student will investigate architectures such as U-Net, Attention U-Net, ResNet, EfficientNet, and transformer-based or attention-enhanced CNN models. The study will compare baseline models with attention-based approaches on public benchmarked dataset such as EBHI-Seg and NCT-CRC-HE-100K and evaluate whether attention improves both segmentation accuracy and pathological classification performance.
The Impact of AI Manipulation on Human Decision-Making	The study could experimentally assess how different forms of AI manipulation (e.g., recommendation delivery, personalization, information framing, or choice restriction) affect human decision-making. An experimental AI system could be developed to compare decision quality, autonomy, and trust across manipulated and control scenarios.
AI Methods for Detecting Manipulative Design Practices	The study would develop and experimentally evaluate an AI model capable of automatically identifying manipulative user interface elements or “dark patterns.” The accuracy, explainability, and practical suitability of different AI methods for monitoring manipulative design practices would be compared.
The Application of Explainable AI to Strengthen User Decision-Making Autonomy	The study would assess whether explanations provided by Explainable AI (XAI) help users better understand AI recommendations and make more autonomous decisions. A user experiment could compare different types of explanations and measure decision confidence, understanding, and reliance on AI recommendations.
A Risk Assessment Model for Prohibited AI Practices under Article 5 of the AI Act	The study would focus on developing a formal risk assessment model that uses defined criteria to determine whether a particular AI system falls within the categories of prohibited practices under Article 5 of the EU AI Act. The model could be tested on real-world or hypothetical AI system cases and evaluated in terms of classification accuracy, consistency, and decision explainability.
The Application of Explainable AI for Detecting Manipulative and Deceptive AI Techniques	The study would develop an XAI-based method for identifying when an AI system employs manipulative or deceptive techniques and for explaining which system characteristics led to such a determination. The research would assess not only detection performance but also whether the explanations are understandable and useful to humans responsible for AI system oversight.
Protecting Vulnerable Groups from the Impact of AI Systems	The study would analyse how AI systems affect vulnerable user groups differently and what governance measures could mitigate potential negative impacts. A risk assessment framework could be developed to identify higher-risk situations and evaluated using different user or AI decision-making scenarios.
Formalizing Significant Harm Criteria for AI System Assessment	The study would seek to translate the abstract concept of “significant harm” into concrete and measurable criteria that can be applied when assessing AI system risks. The resulting set of criteria could be validated through expert evaluation and applied to the analysis of real-world or synthetic AI use cases.

Application of AI-Based Adaptive Learning Solutions for Individuals with ADHD	This work will explore the potential of artificial intelligence to adaptively present learning materials available in Moodle to students with attention deficit hyperactivity disorder (ADHD). The student will develop a solution that analyses uploaded course content, assignment descriptions, and learning objectives, adapting the presentation of materials to individual needs by dividing them into shorter learning steps, highlighting key information, and providing clear instructions for completing tasks. The adaptations will be refined based on students' preferences and feedback.
Completion of Unfinished Musical Works Using Artificial Intelligence	This work explores the use of artificial intelligence to reconstruct missing passages or complete unfinished musical works, drawing on stylistic patterns in other works by the same composer from a similar period. The student will analyse selected compositions, adapt and compare AI models, and generate musical passages consistent with the available material. The results will be evaluated for musical coherence, stylistic consistency, and compatibility with the surviving score, treating the reconstructions as plausible interpretations rather than definitive restorations of the composer's intentions.
Hyperspectral Imaging and Artificial Intelligence for Textile Material Classification	This work explores hyperspectral imaging and artificial intelligence for textile material classification, aiming to improve material identification for sorting and recycling. The student will analyse existing hyperspectral datasets, supplemented with additional measurements if needed, and train and compare machine learning models to distinguish textile materials. The study will evaluate classification performance and the potential of the approach for automated textile sorting.
Computer Vision Methods for Textile Fibre Identification in Recycling Processes	This work will investigate the identification of textile fibres through the analysis of microscopic images. The student will compare classification methods based on fibre shape and texture features with deep learning models and improve the solutions by adapting existing models and developing hybrid models. Recognition accuracy, processing time, and model performance on worn, damaged, or contaminated fibres will be evaluated to assess their suitability for textile sorting and composition verification in recycling processes.
Automated Discovery and Exploitation of Prompt Injection Vulnerabilities in Self-Hosted LLM Intermediary APIs	Self-hosted LLM intermediaries like Ollama expose unauthenticated API surfaces vulnerable to indirect prompt injections in RAG workflows. In this Master's thesis, an automated red-teaming methodology will be developed to discover and exploit prompt injection vectors targeting self-hosted LLM intermediary architectures. A Python fuzzing framework will systematically inject malicious payloads into data streams, and a dynamic Python middleware filter will be constructed to sanitize inference requests via contextual boundary enforcement. Mitigation effectiveness will be verified using standardized LLM security benchmarks. Choosing this topic requires proficiency in Python, local LLM orchestration, API security testing, and adversarial machine learning.
Influence of Quantization and Context Truncation on Local LLM Accuracy in Cybersecurity Regulatory Mapping	Model quantization and context window truncation in local LLMs risk degrading reasoning accuracy during complex cybersecurity regulatory mapping. In this Master's thesis, a quantitative evaluation methodology will be developed to determine the impact of quantization levels and context truncation on local LLM accuracy during control cross-referencing. Using Python and the Ollama engine, an automated benchmarking pipeline will measure semantic drift, hallucination rates, and computational overhead to establish an optimized configuration threshold. Mappings will be verified against a human-validated ground-truth dataset. Choosing this topic requires proficiency in Python, LLM parameter manipulation, local AI engine deployment, and security governance frameworks.
Investigation of Hallucination Detection and Mitigation Methods in Vision-Language Models	To analyze and investigate methods for detecting and reducing hallucinations in image and language models. This work will examine different types of hallucinations related to the generation of objects, attributes, and relationships between objects, as well as their impact on the reliability of the models. This will require an analysis of the scientific literature, existing methods, and evaluation systems, as well as the formulation of a research methodology and evaluation criteria. The study will assess the impact of various hallucination mitigation methods—such as contrastive decoding and preference-based model blending—on model performance. Experiments will be conducted using publicly available evaluation datasets and tasks designed to assess the factual accuracy and reliability of image and language models. The developed experimental solution will be tested and evaluated based

	on hallucination frequency, response accuracy, and the overall quality and effectiveness of the models. The thesis concludes with findings and recommendations regarding the practical application of the most effective methods for reducing hallucinations.
Optimizing the Efficiency of Large Language Models in an Environment with Limited Computational Resources	To analyze and investigate methods for optimizing the efficiency of large language models designed for execution in environments with limited computational resources. This work will examine the impact of various optimization methods—such as quantization, model architectural solutions, and the application of Mixture-of-Experts and dense models—on model quality, performance, and resource utilization. This will require an analysis of the scientific literature and existing technological solutions, as well as the formulation of evaluation criteria and an experimental methodology. During the study, various large language models will be selected and tested, evaluating their accuracy, reasoning capabilities, hallucination rate, response generation speed, memory consumption, and energy efficiency. The developed experimental solution will be tested using standardized evaluation datasets and real-world usage scenarios. The thesis concludes with findings and recommendations regarding optimal strategies for deploying and using large language models in environments with limited computational resources.
Explainability Algorithms Reliability Assessment in Neural Network Based Medical Image Analysis	Research aim is to develop and evaluate a systematic method for assessing the reliability of explainability algorithms applied to neural network models in the medical image analysis The main objectives are the following: to analyze the behavior of selected explainability algorithms across different neural network architectures, to measure the stability and consistency, to design reliability metrics suitable for evaluating explainability methods in medical imaging tasks. to validate the proposed assessment method, to identify limitations and potential failure modes.
Reliability and explainability assessment method development for AI models predicting multidimensional health indicators	This research focuses on evaluating the reliability and explainability of artificial intelligence models to predict multidimensional health indicators. The physiological measurements, the laboratory results, the lifestyle related metrics, the risk scores are used. The research examines how explainable AI methods represent model decision processes. A statistical reliability framework is applied to assess robustness, stability, and interpretability of model outputs. The study supports the development of transparent AI systems for clinical decision support and digital health applications
Human-Informed Federated Chest X-Ray Classification with Calibrated Uncertainty and Explanation Alignment.	Develop a federated chest X-ray classifier using existing labelled datasets and expert annotations. Compare centralized, local, FedAvg, FedProx, and one personalized method. Add uncertainty calibration and explainability, then evaluate AUROC, calibration, explanation alignment, client-level performance, and communication cost. Potential data: MIMIC-CXR-JPG + BRAX/VinDr-CXR + REFLACX.
Personalized Federated Lung Segmentation under Cross-Site Domain Shift with Uncertainty-Aware Prediction.	Investigate federated lung segmentation across heterogeneous hospital-like datasets using U-Net. Compare centralized, local, FedAvg, FedProx, and FedBN or another personalization strategy. Study domain shift, generate uncertainty maps, and evaluate Dice, IoU, Hausdorff distance, client-specific performance, uncertainty quality, and communication cost using existing segmentation masks. Potential data: CheXmask and CXLSeg.
Parameter-Efficient Federated Vision-Language Learning for Chest X-Ray Image–Report Retrieval.	Use a pretrained medical vision-language model to learn from chest X-rays and associated reports across simulated institutions. Apply LoRA or adapters for efficient federated fine-tuning. Compare centralized, local, and federated approaches using image-to-text and text-to-image retrieval metrics, client-level performance, and communication cost. Existing datasets remove the need for new clinical annotation. Potential data: MIMIC-CXR + LATTE-CXR or REFLACX.
Federated Multimodal Clinical Prediction under Missing-Modality Conditions.	Develop a federated multimodal prediction framework using two medical data types, such as chest X-rays and physiological data. Compare unimodal, centralized multimodal, and federated models, then introduce missing-modality conditions using modality dropout or robust fusion. Evaluate AUROC, AUPRC, calibration, client-specific performance, and resilience when one modality is unavailable. Potential data: Symile-MIMIC.

Communication-Efficient Federated Medical Imaging for Resource-Constrained Healthcare Environments.	Study how to reduce communication and computational demands in federated medical imaging without major loss of model performance. Use one fixed imaging task and compare strategies such as fewer communication rounds, quantization, sparsification, or parameter-efficient fine-tuning. Evaluate predictive utility, bytes transmitted, runtime, convergence, memory use, and client-level performance. Potential data: MIMIC-CXR-JPG or CXLSeg.
Governance and Continuous Monitoring of Federated Medical AI under Cross-Site Distribution Shift.	Design a monitoring and governance framework for federated medical AI across multiple simulated sites. Track performance, calibration, uncertainty, drift, subgroup behaviour, communication cost, and model versions. Introduce controlled distribution shifts, define engineering alert thresholds, and compare aggregate versus site-aware monitoring. Produce model documentation covering risks, deployment decisions, and lifecycle management. Potential data: MIMIC-CXR-JPG + VinDr-CXR/BRAX.
Multimodal Privacy Violation Detection in Social Media Posts with Policy-Constrained Decision Rules	This project will develop a multimodal AI system for detecting privacy risks in social media posts by analysing both images and accompanying text. The student will compare image-only, text-only, and multimodal models for detecting privacy categories such as identity exposure, location disclosure, contact information, and sensitive personal information. A simple policy-rule layer will be added to ensure that model decisions comply with predefined privacy requirements. The system will be evaluated using classification and policy-compliance metrics on existing labelled datasets. Potential datasets: FineViP and PRISM, with optional supplementary use of VISPR, VizWiz-Priv, and a self-disclosure text corpus.
Comparison of Explainable AI Methods in Blood Test-Based Disease Prediction	The study would apply existing machine learning models to predict the likelihood of a selected condition (e.g., liver or chronic kidney disease) from blood test parameters, using public datasets (e.g., from the UCI repository). Different explainability methods (e.g., SHAP, LIME, permutation feature importance) would be compared to assess whether they identify the same key parameters, whether the explanations are stable, and whether they are consistent with medical knowledge.
Evaluating the Reliability of Visual Explanations in X-ray Image Classification Models	The study would use pre-trained chest X-ray classification models (e.g., from the TorchXRayVision library) and apply different visual explanation methods (e.g., Grad-CAM, Grad-CAM++, Integrated Gradients). Using public datasets with annotated pathology regions (e.g., RSNA Pneumonia), the research would assess how well the highlighted regions correspond to patological areas.
Applying Explainable AI to Interpret Decisions of MRI Image Classification Models	The study would adapt a pre-trained image classification model to recognise brain tumours in MRI scans using a public dataset (e.g., BraTS). The model's decisions would be interpreted using visual explainability methods, and the resulting explanations would be compared with the tumour segmentation masks provided in the dataset. The research would assess whether the model bases its decisions on the tumour region or on other, clinically irrelevant parts of the image.
Using Explainable AI to Detect Shortcut Learning in Medical Imaging Models	The study would analyse whether pre-trained X-ray classification models base their decisions on incidental image elements (e.g., text markers, medical devices, or image borders) rather than on medical features. Using public datasets (e.g., NIH ChestX-ray14 or CheXpert) and explainability methods, such cases would be identified and their impact on model reliability would be assessed.
Assessment the impact of GNNs on the quality of missing data	Eurostat regional statistics often contain missing values, and neighbouring regions tend to be similar. Graph neural networks (GNNs) can exploit this similarity, but their performance depends heavily on how the regional graph is built. The student will construct several graphs: shared borders, geographic distance, indicator similarity, administrative membership and a learned structure. These will be compared with GCN, GAT and GraphSAGE models, using XGBoost without a graph as the baseline.
Data Imputation Attributes in the Eurostat Regional Knowledge Graph Using GNNs	A student during research treats missing-value imputation as a knowledge graph. Regions, countries and indicators are represented as entities, and their connections as typed relations such as "borders", "belongs to country" and "has indicator". The student will build such a knowledge graph from Eurostat data, for example, in RDF format. Missing numeric values will be predicted with relational GNN models (R-GCN, GRAPE, MrAP), and the results will be compared with XGBoost and TabPFN.

Software System Requirements Validation Model Using a Large Language Model with GraphRAG	Students should investigate requirements, traceability, and compliance using LLM+GraphRAG. For the investigation, GraphRAG parameterisation and quality assessment of data compliance and semantic compliance to the study should be applied.
Reliability Research of the MITRE CyberSecurity Threat Knowledge Graph	A student should use the validated MITRE CyberSecurity Threat Knowledge Graph. During research applužy different kinds of perturbation and assess the reliability of KG and KG+LLM (context)
Robustness Research of the MITRE CyberSecurity Threat Knowledge Graph	A student should use the validated MITRE CyberSecurity Threat Knowledge Graph. During research apply different kinds of poisoning and assess the robustnees of KG and KG+LLM (context)
Explainable Deep Learning for Vocal Fold Segmentation and Vibration Analysis	The aim of this research is to develop an automated and interpretable framework for the analysis of vocal fold vibrations from high-speed videoendoscopy. Deep learning methods will be used to segment the glottis and vocal folds in video frames, while temporal and signal-processing methods will extract quantitative vibration characteristics such as glottal area, amplitude, periodicity, opening and closing patterns, and left-right asymmetry. Explainable AI methods will then be investigated to determine which extracted characteristics contribute most strongly to distinguishing normal and disordered vocal fold vibration.
Classification of Vocal Fold Pathologies from Laryngoscopic Images	The research is aimed at classifying vocal fold pathologies based on laryngoscopic images. It is envisaged that the model will classify the images into eight categories, encompassing normal vocal folds and seven different pathological conditions.
Voice Pathology Detection from Audio Recordings with Limited Training Data	This research focuses on mashine learning methods for detecting voice pathologies from audio recordings with limited labeled data. Different audio representations and learning strategies will be investigated to improve classification performance in low-data settings.
Early Prediction of Students' Academic Performance Based on Learning Activity Dynamics	This research investigates machine learning methods for the early prediction of students' academic performance based on learning activity dynamics.
Improving the Reliability of Skin Lesion Recognition under Image Distortions	This study investigates methods for improving the robustness of neural network models for skin lesion recognition under various visual distortions, such as image blur, noise, illumination changes, compression artifacts, and reduced resolution.
A study of methods for detecting hallucinations in large language models	Practical tasks involve creating a test set of questions, obtaining responses from several LLMs, and developing a classifier or metric to identify potential hallucinations.
Uncertainty calibration in LLM-based agents	Study whether agents that use tools (search, code execution) know when they are likely to fail, and develop methods for them to abstain or ask for help. Evaluation benchmarks are still immature here, so well-designed benchmarks are required.
Efficient fine-tuning for low-resource languages	Compare parameter-efficient methods (LoRA variants, adapters, vocabulary extension) for adapting LLMs to a low-resource language such as Lithuanian.
Synthetic data for training and its failure modes.	Study when LLM-generated or simulator-generated training data helps versus when it causes model collapse or bias amplification in a specific task.
Detecting and mitigating benchmark contamination in open-weight models	Suggest methods to estimate whether a model has seen test data during training, using techniques like perplexity analysis or rephrased test sets.
Controlled comparison of Socratic LLM tutors vs. answer-giving LLM tutors	Design a guardrailed tutor that refuses to give direct solutions and instead scaffolds with hints, then run an A/B study measuring learning gains and student frustration.
Design and Implementation of a Subject-Agnostic Real-Time Socratic LLM Tutor with Lecturer-Controlled Scaffolding for In-Class Learning	This thesis designs and implements a real-time Socratic LLM tutor that guides students with hints and questions instead of giving direct answers, and adapts to any subject using course-specific materials. The lecturer monitors student questions through a live dashboard and adjusts the tutor's hint level during class.

Robustness of Voice-Clone Detection Methods for the Lithuanian Language.	Tests whether deepfake audio detectors trained mostly on English speech work on Lithuanian, and adapts them using a Lithuanian dataset.
Early Prediction of Student Performance from Moodle Activity Logs: Model Transferability Across Study Programmes and Groups	Predicts students' final module outcomes from early behavioural features extracted from Moodle activity logs, comparing classical machine learning models in terms of accuracy, earliness and interpretability. The main focus is transferability, i.e. whether a model trained on some study programmes and groups remains accurate on unseen ones.
Identifying Student Learning Behaviour Profiles from Moodle Activity Logs and Their Relationship to Outcomes Across Study Programmes	To extract students' behavioural features and apply clustering methods to identify learning behaviour profiles. To evaluate the quality and stability of these profiles and to examine whether they recur across different study programmes and are associated with final outcomes.
Process Mining of Learning Trajectories in Moodle: Comparing Actual Student Paths with the Intended Course Design	To reconstruct students' learning process models and compare them with the intended course structure. To analyse how the trajectories of higher- and lower-performing students differ, and to formulate recommendations for improving the course.