SemEval 2025

Διπλωματική Εργασία 1

Επιβλέπων	Γιώργος Στάμου, Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computational Linguistics
Τύπος Εργασίας	Research

Περιγραφή

The present description involves several shared-tasks, each of which can serve as an individual diploma thesis. Shared tasks in SemEval 2025 are independent from each other and focus in several areas of NLP and multimodal learning. A full list of shared tasks for SemEval 2025 can be found here: https://semeval.github.io/SemEval2025/tasks
Each student participating in any SemEval 2025 task is required to successfully participate in the competition in all subtasks -if applicable-, and also write a research paper describing their approach. The timeline of the competition should be strictly followed, therefore a student wishing to be assigned a SemEval thesis commits to the official deadlines; the deadline for submitting results is end of January 2025, while the paper submission follows within February. The outcome of this endeavor will be a participation and a publication in one of the most prestigious workshops in NLP co-located with one of the major NLP conferences, setting a perfect starting point for further research within AILS laboratory, as well as for the student's future research steps (e.g. boosting their PhD application in competitive universities). Each student can only select one task to participate in, while each task will be assigned to only one student.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

Faithfulness of explanations for Natural Language Inference

Διπλωματική Εργασία 2

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

Natural Language Inference (NLI) is a pivotal NLP task, aiming to define whether a hypothesis entails, contradicts or remains neutral with respect to a given premise. NLI models trained on the seminal e-SNLI dataset are widely susceptible to adversarial attacks, with significant performance drops when such attacks occur. In our recent work, we experimentally prove that only by generating intermediate explanations before predicting our final NLI label can we achieve a significant increase in model robustness. [1]
An open question arising from our previous work is the evaluation of explanations, as their linguistic and semantic quality, as well as their faithfulness to input premise and hypothesis is highly correlated to the robustness of the NLI model. To this end, we seek to first evaluate the faithfulness of generated explanations and as a consequent step advance their faithfulness in practice [2, 3, 4], based on the e-SNLI dataset. Moreover, we test our previous experimentation [1] on LLMs rather than smaller encoder models to seek whether model scale can naturally lead to more faithful explanations, according to our reported metrics. The experimentation can be expanded to more NLI classifiers and other datasets. The faithfulness-robustness relationship will be tested under a variety of adversarial attacks.

[1] ”Enhancing adversarial robustness in Natural Language Inference using explanations”, Alexandros Koulakos, Maria Lymperaiou, Giorgos Filandrianos, Giorgos Stamou https://arxiv.org/abs/2409.07423
[2] “NILE : Natural Language Inference with Faithful Natural Language Explanations”, Sawan Kumar, Partha Talukdar, ACL 2020. https://aclanthology.org/2020.acl-main.771/
[3] “Logical Satisfiability of Counterfactuals for Faithful Explanations in NLI”, Suzanna Sia, Anton Belyy, Amjad Almahairi, Madian Khabsa, Luke Zettlemoyer, Lambert Mathias, AAAI 2023. https://arxiv.org/abs/2205.12469
[4] “Towards Faithful Model Explanation in NLP: A Survey” Computational Linguistics, 2024. https://direct.mit.edu/coli/article/50/2/657/119158/Towards-Faithful-Model-Explanation-in-NLP-A-Survey

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr) και τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr).

A versatile counterfactual editor based on graph optimizations

Διπλωματική Εργασία 3

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

In our previous work [1], we constructed a graph-based counterfactual editor which leverages graph assignment algorithms and GNN-driven speedup to advance optimality, controllability and inference time in comparison to state-of-the-art editors in literature. Our optimization lies in discovering the minimum set of contrastive edits to alter a classifier's decision, even though we can operate in classification-free settings, which was left for future work.
In this thesis, we plan to test our counterfactual editor in various NLP tasks to test its adaptability and its model-agnostic power. For example, we can craft appropriate attacks to test the robustness of NLI models, as in Diploma #2, targeting advanced attack success rate. Moreover, when targeting model interpretability, we can discover linguistic patterns and biases imbued in pre-trained language models across a variety of NLP tasks.
We stress optimality concerns by also experimenting with the beam search function that defines the final contrastive edits to be performed, requesting that the minimum number of source words will be altered. Instead of picking the more contrastive pairs, we can question whether other choices, e.g. altering words with larger attention scores are more influential in comparison to semantic dissimilarity of edits, in terms of classifier flip-rate or other evaluation metrics.

[1] "Optimal and efficient text counterfactuals using Graph Neural Networks", Dimitris Lymperopoulos, Maria Lymperaiou, Giorgos Filandrianos, Giorgos Stamou https://arxiv.org/abs/2408.01969

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr) και τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr).

Causal inference and counterfactuals in the LLM era: a survey

Διπλωματική Εργασία 4

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Causality
Τύπος Εργασίας	Survey

Περιγραφή

Causal discovery has been a core area in AI research, exploring the cause-effect patterns present in data, providing understanding and explainability beyond simple correlations. With the advent of large language models (LLMs), the landscape of causal reasoning has begun to shift. LLMs, known for their ability to generate human-like text and perform a wide range of tasks, have introduced new possibilities for understanding and inferring causality from data. However, despite their success in natural language processing (NLP) tasks, the question remains: can LLMs effectively reason about causality, and what role can they play in advancing causal inference?
This survey paper aims to explore the intersection of causal inference and LLMs. We examine how traditional causal inference techniques, such as Pearl's Causal Model and Rubin's Potential Outcomes Framework, have been applied in the context of LLMs. Additionally, we survey emerging methods that leverage the power of LLMs to assist with causal discovery, reasoning, and prediction across various domains. As LLMs are trained on vast amounts of text data, they offer unique opportunities to extract causal insights from unstructured information, but they also present challenges regarding their capacity to distinguish correlation from causation. We highlight the strengths and limitations of these models in causal reasoning tasks, the new prompting strategies designed to enhance causal understanding, and the potential of LLMs to reshape traditional approaches to causal discovery.
The current survey is planned to be a publishable work on its own, followed by a technical part later in the year to fulfill the diploma thesis requirements imposed by AILS laboratory.

[1] Causal Reasoning and Large Language Models: Opening a New Frontier for Causality https://arxiv.org/pdf/2305.00050

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, causal inference, counterfactuals & XAI. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr) και την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr)

Character guidance for story visualization using diffusion models

Διπλωματική Εργασία 5

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Diffusion Models, AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Story visualization is a challenging task in computer vision where a coherent sequence of images has to be generated based on a corresponding textual sequence. In our prior work, we introduce the technique of character guidance [1], aiming to improve the generation of specific characters present in storyboards of related Story Visualization datasets (Pororo-SV, Flintstones-SV and others). As our prior experimentation showcases the potential of character guidance using generative transformers to produce the story frames, we attempt to expand this claim leveraging even more powerful generators, usually based on diffusion models. Moreover, we exploit character guidance on the closely related Story Continuation task, where the conditioning further involves an initial frame, together with the given story descriptions. Our experimental suite is concluded by employing a variety of Large Language Models (LLMs) to enrich the textual descriptions, so that visual frames can be generated with more details, accurately representing ground truth concepts and characters. Both tasks are evaluated using typical metrics suited for these tasks on all benchmark datasets, as well as comparative human evaluation surveys.

[1] “Masked Generative Story Transformer with Character Guidance and Caption Augmentation”, Christos Papadimitriou, Giorgos Filandrianos, Maria Lymperaiou, Giorgos Stamou https://arxiv.org/abs/2403.08502

Απαιτούμενες/επιθυμητές γνώσεις: Python, όραση υπολογιστών, diffusion models. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr) και τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr).

Automatic harmonization of given musical melody

Διπλωματική Εργασία 6

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Knowledge Representation - AI for Music
Τύπος Εργασίας	Research

Περιγραφή

A musical piece is usually analyzed into three main components: melody, harmony, and rhythm. Each artist chooses how to compose their work, either starting with the melody and then harmonizing it, or by structuring the harmony upon which the melody will be based. The purpose of this diploma thesis is to study and develop tools for the automatic harmonization of a given melody. For this purpose, large datasets of musical works in symbolic form will be collected to extract information, and a Knowledge Base will be developed for the categorization of fundamental harmonization rules.
Next, a system will be developed that, given a melody, will be able to propose and compare different ways of harmonizing it, as well as reproduce the auditory result.
Furthermore, since the study will cover harmonization techniques rooted in both tonal harmony and contemporary approaches, such as the use of chord progressions, a **deep understanding and familiarity with complex musical concepts are necessary**. This project has a strong research and creative aspect, so it would be desirable for the student to have a personal perspective on music composition and be able to propose their own harmonization methods.

Απαιτούμενες/επιθυμητές γνώσεις: Περιγραφικές Λογικές, Γράφοι Γνώσης, Τεχνικές Αναπαράστασης Γνώσης, Οντολογίες, εξοικείωση με κάποια Γλώσσα Προγραμματισμού (ενδεικτικά Python), κατανόηση σε βάθος μουσικών όρων.Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr)

Development of a musical Knowledge Graph for pattern recognition

Διπλωματική Εργασία 7

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Knowledge Graphs - AI for Music
Τύπος Εργασίας	Research

Περιγραφή

The field of Music Information Retrieval (MIR) is closely linked to Computer Science. There is a vast amount of data available in both audio form and symbolic formats (e.g., MIDI, MusicXML, text). The goal of this thesis is to develop a musical Knowledge Graph based on a large volume of symbolic music data, aimed at recognizing patterns based on the harmonic characteristics of musical pieces.
Since MIR combines machine learning techniques with knowledge systems, it is important for the student undertaking this project to have knowledge in both areas. Additionally, as the focus will be on identifying musical patterns based on harmonic features, a **deep understanding and familiarity with complex musical terms are necessary**. The project is highly research-oriented but also creative, making it desirable for the student to have their own perspective on music.

Απαιτούμενες/επιθυμητές γνώσεις: Περιγραφικές Λογικές, Γράφοι Γνώσης, Τεχνικές Αναπαράστασης Γνώσης, εξοικείωση με κάποια Γλώσσα Προγραμματισμού (ενδεικτικά Python), κατανόηση σε βάθος μουσικών όρων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr), Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr)

Robust Visual Inference

Διπλωματική Εργασία 8

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Multimodality
Τύπος Εργασίας	Research

Περιγραφή

Beyond purely linguistic inference (NLI task), extending to multimodal settings that incorporate vision and language poses an abundance of challenges. For example, robustness of visual entailment on the SNLI-VE dataset (as an extension of SNLI with COCO images) is an underexplored domain, being on par with the limited work present in NLI robustness. In our recent work, we experimentally prove that only by generating intermediate explanations before predicting our final NLI label can we achieve a significant increase in model robustness. [1]
In this thesis, we aim to explore whether our proven claim remains relevant in the multimodal setting. We extend our experimentation by also questioning the faithfulness of explanations [2] and whether faithfulness is correlated with advanced model robustness.

For related literature check:
[1] ”Enhancing adversarial robustness in Natural Language Inference using explanations”, Alexandros Koulakos, Maria Lymperaiou, Giorgos Filandrianos, Giorgos Stamou https://arxiv.org/abs/2409.07423
[2] “Logical Satisfiability of Counterfactuals for Faithful Explanations in NLI”, Suzanna Sia, Anton Belyy, Amjad Almahairi, Madian Khabsa, Luke Zettlemoyer, Lambert Mathias, AAAI 2023. https://arxiv.org/abs/2205.12469

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr) και τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr).

Do short-rule trees make short-rule forests?

Διπλωματική Εργασία 9

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

In the domain of tabular data, Decision Trees are one of the most popular Machine Learning models, and Tree Ensembles such as Random Forests, AdaBoost or XGBoost consistently outperform all other models in Machine Learning competitions. One of the most important features of Decision Trees is that they are inherently interpretable, especially when they are small in size. Tree Ensembles on the other hand are opaque due to the large number of trees that are produced in most cases. Many techniques have been developed for illuminating how opaque models make decisions. One such technique is Anchor [1,2] which produces rules that explain which features of an input determine the decision made by the model.

The work in [3] introduces a new technique for producing trees that classify most examples with only a few tests, so even when the resulting tree is large the end-user has a clear picture of how most inputs are classified, and the tree can be approximated by a few short rules. The goal of this thesis is to explore how such trees can be used to create tree ensembles that techniques such as Anchor can easily and accurately explain without sacrificing performance.

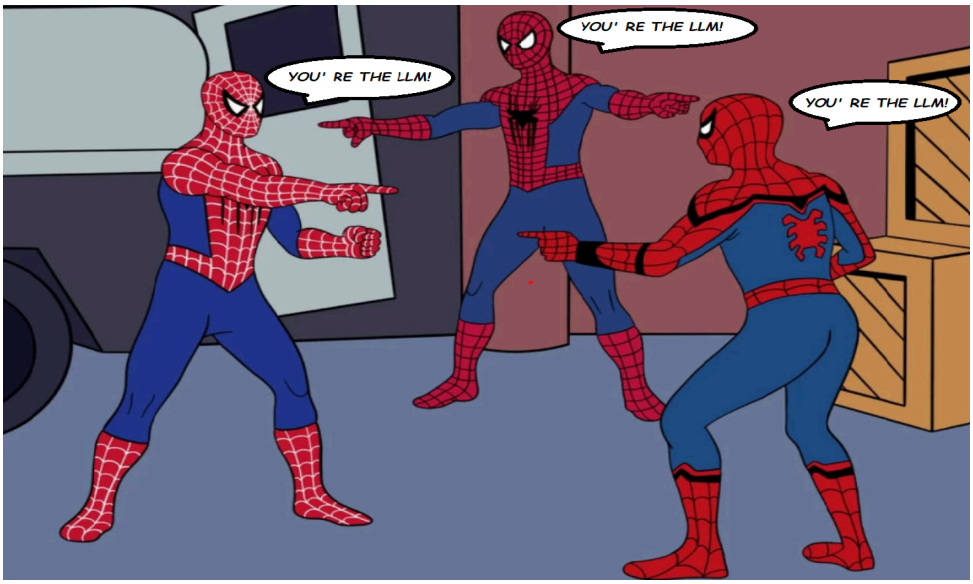
[1] Anchors: high-precision model-agnostic explanations
https://dl.acm.org/doi/abs/10.5555/3504035.3504222
[2] Interpretable Machine Learning §9.4
https://christophm.github.io/interpretable-ml-book/anchors.html
[3] A New Splitting Criterion for Better Interpretable Trees
https://ieeexplore.ieee.org/document/9054987

Απαιτούμενες/επιθυμητές γνώσεις: Python, βασικές γνώσεις για Decision Trees και Tree Ensembles. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), και τον Ιάσονα Λιάρτη (e-mail: jliartis at ails.ece.ntua.gr)

Among Us: Can LLMs detect a LLM?

Διπλωματική Εργασία 10

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research



Περιγραφή

The diploma thesis will explore the capabilities of Large Language Models (LLMs) in both concealing their machine identity and identifying other LLMs in interactive scenarios. The student will develop a framework where three LLMs engage in dialogue with two objectives: first, to avoid being detected as an LLM by the other two models, and second, to identify whether the other discussants are LLMs.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr), τον Κων/νο Θωμά (e-mail: kthomas at ails.ece.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

Understanding memorisation in language models

Διπλωματική Εργασία 11

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

Memorization in large language models (LLMs) has significant ethical implications, especially when these models are adapted with potentially sensitive or proprietary data from specific organizations or companies. While existing research [1] has thoroughly investigated memorization during the pretraining phase—identifying stronger and more persischaracteristent memorization in larger models influenced by data order and learning rate—less is known about memorization during model adaptation. This thesis explores how various adaptation techniques influence memorization, particularly focusing on how changes in model configuration affect its ability to recall and process training data. Incorporating a range of adaptation methods such as LoRA [2], the study investigates the impact of specific hyperparameters (alpha and r) on memorization trends. This analysis aims to provide insights into how tailored adjustments during the adaptation phase can modify memorization tics of LLMs, addressing the need for responsible handling of sensitive data in customized model applications.

[1] Pietro Lesci, Clara Meister, Thomas Hofmann, Andreas Vlachos, and Tiago Pimentel. 2024. Causal Estimation of Memorisation Profiles. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 15616–15635, Bangkok, Thailand. Association for Computational Linguistics.
[2] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." arXiv preprint arXiv:2106.09685 (2021).

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr) και την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr)

Benchmarking Graph Prototypes: A New Evaluation Dataset

Διπλωματική Εργασία 12

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks, Explainable AI
Τύπος Εργασίας	Dataset Creation

Περιγραφή

The discovery of graph prototypes — representative substructures or patterns that capture the essence of a graph class — is an exciting and growing area of research. However, despite numerous studies published in top conferences [1, 2, 3, 4, 5], there is still no well-established dataset for evaluating these prototypes. Researchers often rely on hand-crafted synthetic benchmarks for quantitative analysis, but since different works use different datasets, it leads to inconsistent comparisons and evaluation metrics. For qualitative analysis, protein datasets are sometimes used to assess whether the generated prototypes are representative of their respective classes, but the reliance on a single dataset limits the generalizability of results.
This thesis aims to fill this gap by creating a standardized graph prototype dataset that includes structural graphs along with their node attributes, providing a consistent benchmark for future studies. This can be achieved by drawing inspiration from existing studies and modifying well-known graph benchmarks (https://ogb.stanford.edu/) in order to create new datasets with ground truth prototypes.

[1] https://arxiv.org/pdf/1903.03894
[2]https://arxiv.org/pdf/2011.04573
[3]https://arxiv.org/pdf/2102.05152
[4]https://arxiv.org/pdf/2112.00911
[5]https://arxiv.org/pdf/2210.01974

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr).

Guiding LMM Responses: Causal Interventions for Accurate Question Answering

Διπλωματική Εργασία 13

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Diffusion Models, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

In recent efforts to improve the reliability of Large Multi-modal Models (LMMs) and address their tendency to produce hallucinations, Counterfactual Inception has been proposed [1]. This method introduces counterfactual thinking by generating alternative keywords related to the input data, aiming to enhance the model's contextual understanding without additional training. However, some researchers have noted that some LMMs often lack exceptional reasoning capabilities, which can lead to anchoring effects and therefore inconsistent or inaccurate responses.

This thesis proposes an alternative approach by leveraging causal inference to directly intervene in the model's input to guide its responses. Instead of relying on counterfactual keywords, this method involves modifying the input itself to create a more accurate context for the model. For example, in a scenario where an image shows three cats and the question is "Which cat is yawning?"—if the model incorrectly identifies the cat—interventions would involve altering the image to clearly show the correct yawning cat. By performing these interventions, we aim to correct the model's initial response and ensure it answers the question accurately from the start.

Various experimentation methods will be explored, including using diffusion models to modify images, integrating textual descriptions, and employing external knowledge to enhance the plausibility of the modifications. The performance of this approach will be evaluated against Counterfactual Inception and other hallucination-prevention methods, with the goal of further improving the reliability and contextual accuracy of LMMs.

[1] "What if...?: Thinking Counterfactual Keywords Helps to Mitigate Hallucination in Large Multi-modal Models", Junho Kim Yeonju Kim Yong Man Ro https://arxiv.org/pdf/2403.13513

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr) και τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr).

Incorporating Visual Features for Image Retrieval through Scene Graphs

Διπλωματική Εργασία 14

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Research

Περιγραφή

Image Retrieval is the problem of ranking images, based on their similarity. This field has been dominated by Vision models (CNN-based and ViT-based) for years, but they exhibit severe biases when retrieving similar images, especially regarding low-level image features. For example, when a black-and-white image is given as input, the retrieved images are similar in the black-and-white colors, but they are often irrelevant in the actual content of the image (object, attributes, interactions).
In order to solve this issue, Scene Graphs have been utilized [1, 2], that provide a graph-structured explanation of an image, that contains the objects (nodes) and their attributes, along with the interactions between them (edges). Leveraging scene graphs for Image Retrieval, allows us to find semantically similar images, without taking low-level image features into account. Several papers have introduced methods to utilize purely the scene graphs for Image Retrieval, but only one approach combines the Scene Graphs with Visual Features [3]. This diploma thesis is heavily research-oriented, and aims to combine Graph and Visual Features in a novel framework.

[1] https://arxiv.org/abs/2012.14700
[2] https://arxiv.org/abs/2403.06514
[3] https://dl.acm.org/doi/10.1145/3581783.3612283

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr).

Using audio FX to alter music emotion

Διπλωματική Εργασία 15

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	AI for Music
Τύπος Εργασίας	Research

Περιγραφή

The study of music emotion recognition has gained significant attention in recent years, with advancements in machine learning models enabling deeper insights into emotional responses to music. This thesis will focus on how audio effects (FX) can be used to modify the emotional perception of music. The student will work with existing music datasets, applying state-of-the-art emotion recognition models to classify emotional content. They will then experiment with various audio FX to alter the emotional predictions made by these models, analyzing the impact of different effects on the perceived emotion.

[1] Dash, Adyasha, and Kathleen Agres. "AI-Based Affective Music Generation Systems: A Review of Methods and Challenges." ACM Computing Surveys 56.11 (2024): 1-34.
[2] Chen, Yanxu, Linshu Huang, and Tian Gou. "Applications and Advances of Artificial Intelligence in Music Generation: A Review." arXiv preprint arXiv:2409.03715 (2024).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr) , τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr) και τον Έντι Ντερβάκο (e-mail: eddiedervakos at ails.ece.ntua.gr)

Analyzing Music Through Chord Progressions and Graph Representation Learning

Διπλωματική Εργασία 16

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	AI for Music, Graph Representation Learning
Τύπος Εργασίας	Research

Περιγραφή

The Music Information Retrieval (MIR) community addresses various tasks that rely on the chord progressions of a music track, including chord prediction [1], music structure analysis [2], and harmony analysis [3]. In this project, you will work with the largest dataset containing chord progressions, as far as we know, named *Chordonomicon* [4]. Your tasks will focus on music genre classification and music track similarity. One promising direction is to incorporate latent distance models to analyze the dataset.
Latent distance approaches have been successfully applied to model network data [5, 6], where nodes are represented in a latent space that captures their relationships. A similar approach can be adapted for music, treating each part of a music track as a node. By embedding tracks in this latent space, you can capture similarities and genre characteristics. In this work, latent representation methods will be applied to tackle music track similarity, chord prediction, and genre classification tasks.

[1] Müller, Meinard, and Meinard Müller. "Music structure analysis." *Fundamentals of music processing: Audio, analysis, algorithms, applications* (2015): 167-236.
[2] Garoufis, Christos, Athanasia Zlatintsi, and Petros Maragos. "An LSTM-based dynamic chord progression generation system for interactive music performance." *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2020.
[3] Kantarelis, Spyridon, Edmund Dervakos, and Giorgos Stamou. "The Music Part Ontology." *Description Logics*. 2023.
[4] https://github.com/spyroskantarelis/chordonomicon
[5] Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. "Latent space approaches to social network analysis." *Journal of the american Statistical association* 97.460 (2002): 1090-1098.
[6] Krivitsky, Pavel N., et al. "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models." *Social networks* 31.3 (2009): 204-213.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr) και τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr) και τον Νίκο Χάιδο (email: nchaidos at ails.ece.ntua.gr)

Graph Counterfactuals in Music Genre Recognition

Διπλωματική Εργασία 17

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	AI for Music, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

Music tracks are frequently represented through chord progressions, providing a structured means to analyze and understand musical compositions. One of the most extensive datasets available for this purpose, to our knowledge, is the Chordonomicon [1]. This dataset encodes chord progressions as weighted graphs, offering a rich analytical framework.
Graph counterfactuals, a subfield in explainable artificial intelligence (XAI), focus on identifying the minimal edits needed to change a graph's classification. Several notable works in this area exist [2, 4], with one of the most prominent being CLEAR [3]. In this study, you will explore current methodologies related to graph counterfactuals in graph classification. Specifically, you are going to apply these techniques to determine the minimal graph edits required to change the classification of a music track from one genre to another.

[1] https://github.com/spyroskantarelis/chordonomicon
[2] Prado-Romero, Mario Alfonso, et al. "A survey on graph counterfactual explanations: definitions, methods, evaluation, and research challenges." ACM Computing Surveys 56.7 (2024): 1-37.
[3] Ma, Jing, et al. "Clear: Generative counterfactual explanations on graphs." Advances in neural information processing systems 35 (2022): 25895-25907.
[4] Guo, Zhimeng, et al. "Counterfactual learning on graphs: A survey." arXiv preprint arXiv:2304.01391 (2023).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr), τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr).

EEG Analysis with Explainable Artificial Intelligence (XAI)

Διπλωματική Εργασία 18

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Explainable AI
Τύπος Εργασίας	Survey

Περιγραφή

Electroencephalography (EEG) plays a pivotal role in deciphering dynamic brain processes, offering valuable insights into both normal and pathological conditions, particularly when integrated with Brain-Computer Interface (BCI) technologies. Artificial Intelligence (AI) is commonly applied to analyze and interpret EEG signals, yet its outcomes often fall short of expectations. This is largely due to EEG's unique characteristics, such as significant inter- and intra-individual variability, pronounced non-stationarity, and limited spatial resolution compared to other biosignals. These factors make generalizing EEG features for classification, regression, and anomaly detection tasks a compelling challenge in AI.
Explainable Artificial Intelligence (XAI) has the potential to address this challenge in two key ways: enhancing the interpretability of AI system outputs and improving system performance. XAI techniques can aid in identifying EEG features and transformations with greater stability, thereby enabling more robust generalization and interpretability [1].
Numerous studies have explored the use of Explainable AI (XAI) techniques to analyze EEG data [2-4]. In this project, your goal is to identify and review key research that applies XAI methodologies specifically to EEG analysis.

[1] https://xaiworldconference.com/2024/explainable-ai-for-eeg-research/
[2] Dutt, Micheal, et al. "SleepXAI: An explainable deep learning approach for multi-class sleep stage identification." *Applied Intelligence* 53.13 (2023): 16830-16843.
[3] Sujatha Ravindran, Akshay, and Jose Contreras-Vidal. "An empirical comparison of deep learning explainability approaches for EEG using simulated ground truth." *Scientific Reports* 13.1 (2023): 17709.
[4] Raab, Dominik, Andreas Theissler, and Myra Spiliopoulou. "XAI4EEG: spectral and spatio-temporal explanation of deep learning-based seizure detection in EEG time series." *Neural Computing and Applications* 35.14 (2023): 10051-10068.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) τον Νίκο Σπανό (e-mail: nspanos at ails.ece.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

EEG Analysis in Music Understanding

Διπλωματική Εργασία 19

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	AI for Music
Τύπος Εργασίας	Survey

Περιγραφή

Recent literature has significantly advanced our understanding of the relationship between brain activity and music perception through the use of EEG signals. Xiao et al. [1] introduced MEEG and AT-DGNN, which enhance emotion recognition from EEG data by integrating music and graph learning, demonstrating the growing sophistication of EEG analysis techniques. Cui et al. [2] provided a comprehensive review of music-emotion recognition studies, emphasizing how EEG can capture the intricate emotional responses elicited by music. Daly et al. [3] showed that combining brain activity with acoustic features enables accurate prediction of music-induced emotions, further deepening our knowledge of the neural basis of music perception. Additionally, Turchet et al. [4] explored emotion recognition in musicians using EEG, ECG, and acoustic signals, offering insights into the real-time emotional and cognitive processes involved during music performance. Building on these studies, this thesis will conduct a thorough survey of research using EEG to investigate music cognition, with a focus on emotional processing, rhythm perception, and related cognitive tasks.

[1] Xiao, Minghao, et al. "MEEG and AT-DGNN: Advancing EEG Emotion Recognition with Music and Graph Learning." *arXiv preprint arXiv:2407.05550* (2024).
[2] Cui, Xu, et al. "A review: Music-emotion recognition and analysis based on EEG signals." *Frontiers in neuroinformatics* 16 (2022): 997282.
[3] Daly, Ian, et al. "Music-induced emotions can be predicted from a combination of brain activity and acoustic features." *Brain and cognition* 101 (2015): 1-11.
[4] Turchet, Luca, et al. "Emotion recognition of playing musicians from EEG, ECG, and acoustic signals." *IEEE Transactions on Human-Machine Systems* (2024).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr), τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr) και τον Έντι Ντερβάκο (e-mail: eddiedervakos at ails.ece.ntua.gr).

Emotion Recognition using Biosignals

Διπλωματική Εργασία 20

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Survey

Περιγραφή

Emotion is a complex and subjective concept that is challenging to define precisely. Many existing approaches rely on facial pattern recognition and other techniques that struggle to capture the brain's inner workings. However, EEG signals are widely known for their ability to measure brain activity that correlates with emotional states [1]. Several datasets have been developed to explore the relationship between EEG data and emotions [2]. Most emotion-detection models using EEG data employ deep learning techniques [3], while others utilize graph neural networks (GNNs) [4] or even foundational models [5]. Lots of work generally exists also in the domain of emotion recognition by using physiological data [6]. This work proposes a survey covering the key physiological emotion datasets and the models employed for emotion detection.

[1] Bird, Jordan J., et al. "Mental emotional sentiment classification with an eeg-based brain-machine interface." *Proceedings of the International Conference on Digital Image and Signal Processing (DISP'19)*. 2019.
[2] Koelstra, Sander, et al. "Deap: A database for emotion analysis; using physiological signals." *IEEE transactions on affective computing* 3.1 (2011): 18-31.
[3] Zheng, Wei-Long, and Bao-Liang Lu. "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks." *IEEE Transactions on autonomous mental development* 7.3 (2015): 162-175.
[4] Zhong, Peixiang, Di Wang, and Chunyan Miao. "EEG-based emotion recognition using regularized graph neural networks." *IEEE Transactions on Affective Computing* 13.3 (2020): 1290-1301.
[5] Cui, Wenhui, et al. "Neuro-GPT: Towards A Foundation Model For EEG." *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024.
[6] Egger, Maria, Matthias Ley, and Sten Hanke. "Emotion recognition from physiological signal analysis: A review." *Electronic Notes in Theoretical Computer Science* 343 (2019): 35-55.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr), τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr) και τον Έντι Ντερβάκο (e-mail: eddiedervakos at ails.ece.ntua.gr).

How Model Scale Influences Post-Hoc Explainability in Large Language Models

Διπλωματική Εργασία 21

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

As large language models (LLMs) grow in size to enhance performance, there remains a limited understanding of how this expansion affects their explainability. Some studies have explored the relationship between model size and the quality of explanations provided by LLMs [1]. In this project, you will apply standard explainability techniques, including SHAP values and perturbation-based post-hoc methods [2], to different models varying in size such as LLaMA3-7B, LLaMA3-13B, LLaMA3-70B, Mistral 7B, and Mixtral 8x7B. For post-hoc explainability methods, since most focus on classification, the exploration could involve tasks like sentiment analysis, topic classification, toxic detection, etc. Additionally, different generative tasks may be explored by reformulating them into classification tasks, for example, testing them in Natural Language Understanding benchmarks like Glue [3] or BLiMP [4]. The goal is to investigate how model size influences the interpretability of these explanations, evaluating them with metrics such as fidelity.

[1] Heyen, Henning, et al. "The Effect of Model Size on LLM Post-hoc Explainability via LIME." *arXiv preprint arXiv:2405.05348* (2024).
[2] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Anchors: High-precision model-agnostic explanations." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. No. 1. 2018.
[3] Wang, Alex. "Glue: A multi-task benchmark and analysis platform for natural language understanding." *arXiv preprint arXiv:1804.07461* (2018).
[4] Warstadt, Alex, et al. "BLiMP: The benchmark of linguistic minimal pairs for English." *Transactions of the Association for Computational Linguistics* 8 (2020): 377-392.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) , τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

Contrastive Feature Disentanglement for Domain Generalization

Διπλωματική Εργασία 22

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

In medical image analysis, deep learning models have demonstrated remarkable success in tasks such as segmentation; however, their performance often degrades when applied to new healthcare settings due to variations in imaging protocols, equipment, and patient populations. To address this issue, domain adaptation and domain generalization methods have been introduced, with the goal of improving model robustness across diverse clinical environments. While unsupervised domain adaptation and multisource domain generalization methods have shown promise, they pose significant challenges for practical implementation, particularly due to the high cost of acquiring target-domain data and concerns surrounding the redistribution of data from multiple sources. Recent research has shifted toward single-domain generalization techniques, which aim to enhance model generalizability using only a single source domain. These approaches often incorporate contrastive learning and feature disentanglement to separate domain-specific characteristics, such as style, from domain-invariant structural features. Such methods have shown potential to outperform traditional techniques, offering a more scalable and privacy-preserving solution for robust medical image segmentation across varying clinical domains. In this work, we want to review making any improvements towards the different datasets by changing the architecture or the methods used for single domain generalization.

[1] Devil is in the Channels https://arxiv.org/abs/2306.05254
[2] Domain Generalization: A Survey https://arxiv.org/abs/2103.02503

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

U-Mamba: State Space Model for Medical Image Analysis and Domain Generalization

Διπλωματική Εργασία 23

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Convolutional Neural Networks (CNNs) and Transformers have long dominated biomedical image segmentation tasks, yet both architectures face limitations in effectively modeling long-range dependencies due to their inherent locality or computational complexity. To overcome these challenges, we introduce U-Mamba, a versatile network designed for biomedical image segmentation. Drawing inspiration from State Space Sequence Models (SSMs), known for their strength in handling long sequences, U-Mamba incorporates a hybrid CNN-SSM architecture. This design leverages the local feature extraction capabilities of convolutional layers while utilizing SSMs to capture long-range dependencies. Additionally, U-Mamba features a self-configuring mechanism, enabling it to adapt automatically to diverse datasets without manual tuning. This model demonstrates superior performance across various biomedical segmentation tasks, surpassing existing CNN-based and Transformer-based networks, and offering a new approach to efficient long-range dependency modeling in biomedical image analysis. In this diploma thesis, we want to combine the U-Mamba model with invasive and non-invasive data augmentation methods (input/feature/contrastive loss) in order to optimize its performance in different medical datasets and tasks.

[1] Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model https://arxiv.org/abs/2401.09417
[2] MaxStyle: Adversarial Style Composition for Robust Medical Image Segmentation https://arxiv.org/abs/2206.01737
[3] U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation https://arxiv.org/abs/2401.04722

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr), την Π. Θεοφίλου (e-mail: partheofilou at ails.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

SegMamba: Model Long-range Sequential Modeling Mamba for 3D Imagery

Διπλωματική Εργασία 24

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

VMamba, as a State Space Model (SSM), has recently emerged as a notable approach for modeling long-range dependencies in sequential data, excelling in the field of natural language processing with its remarkable memory efficiency and computational speed. As a result, it has expanded into computer vision tasks by replacing attention mechanisms in Vision Transformers and has seen great improvements, both in performance and in lowering the necessary resources for training. SegMamba is used for long range sequences, such as in 3D medical imaging. In this thesis, we want to review the architecture and modify it for different 3D imagery tasks, such as pose and skeleton recognition and evaluate its performance in enhancing complex long-range sequence tasks.

[1] Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model https://arxiv.org/abs/2401.09417
[2] SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation https://arxiv.org/abs/2401.13560

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr), την Π.Θεοφίλου (e-mail: partheofilou at ails.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

Infinigen: Conditioning LLMs for 3D Scene Generation through Blender Code

Διπλωματική Εργασία 25

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Dataset Creation, Research

Περιγραφή

3-D Vision has known significant improvement in the computer vision the pas few years, due to advancements in hardware and model architectures. It lacks, however, serious support in datasets due to the fact that labeling is extremely difficult and time-consuming for human annotators. As a result, new methods have arisen for generating data from physics based engines that provide along with the images information for labeling. We want to review the possibility of conditioning an LLM into generating appropriate code for image generation and then review how we can use the tool as an augmentation method for different tasks. The task will require a familiarity with the tool and code generation of different scenes, through which we will proceed to appropriate model training and research.

[1] Infinite Photorealistic Worlds using Procedural Generation https://arxiv.org/abs/2306.09310
[2] Infinigen Indoors: Photorealistic Indoor Scenes using Procedural Generation https://arxiv.org/abs/2406.11824

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr) και την Π.Θεοφίλου (e-mail: partheofilou at ails.ece.ntua.gr).

Video Storyboarding with Large Language and Vision Assistance Models

Διπλωματική Εργασία 26

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Concepts involved in long-form videos such as people, objects, and their interactions, can be viewed as following an implicit prior. They are notably complex and continue to pose challenges to be comprehensively learned. In recent years, generative pre-training (GPT) has exhibited versatile capacities in modeling any kind of text content, even visual locations. Some recent work, namely Long Form Video Prior, has introduced a new dataset called Storyboard 20K in regards to combating the problem of small datasets. For this subject, we would like to explore Video Storyboarding and Generation further, and move on into utilizing the dataset in a diploma thesis for further scientific research, by utilizing LLaVA models.

[1] LLaVA: Large Language and Vision Assistant https://arxiv.org/abs/2304.08485
[2] Learning Long-form Video Prior via Generative Pre-Training https://arxiv.org/abs/2404.15909, https://github.com/showlab/Long-form-Video-Prior

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr), την Π.Θεοφίλου (e-mail: partheofilou at ails.ece.ntua.gr) και τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr).

Multimodal Instruction for Image Generation

Διπλωματική Εργασία 27

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	Diffusion Models, AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Image generation has been gaining track in recent years, with the constant improvement of diffusion models. However, we are reaching a stand-still of creating images of high fidelity, but without following the textual prompt given. Recently, more and more multimodal instruction methods have been appearing. In this thesis, we would like to explore how diffusion models can benefit from multi-modal image synthesis and how we can add further modalities into such instructions, such as text, graphs and knowledge bases to further improve semantic content in any type of generated images or different approaches with different methods, such as zero-shot approaches.(faces, objects, sceneries, artwork etc).

[1] Instruct-Imagen: Image Generation with Multi-modal Instruction https://arxiv.org/abs/2401.01952
[2] Adding Conditional Control to Text-to-Image Diffusion Models https://arxiv.org/abs/2302.05543

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr) και τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

Data Collection Strategies for Pretraining LLMs with Low-Resource Languages

Διπλωματική Εργασία 28

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

This thesis aims to thoroughly explore current methodologies and the unique challenges associated with collecting data for low-resource languages intended for pretraining Language Learning Machines (LLMs). These languages are often underrepresented on digital platforms and lack substantial linguistic resources, creating significant obstacles in data acquisition and processing. The thesis will meticulously review existing literature to identify effective data collection strategies and highlight gaps in current approaches.

Additionally, the student will actively participate in the practical aspects of data collection through involvement in an ongoing LLM project that targets a specific low-resource language. This will include developing data collection protocols and utilizing tools to efficiently gather, clean, and annotate linguistic data. This hands-on experience is intended to provide deep insights into the complexities of data collection in real-world settings, thereby enhancing the student's capability to contribute to the development of robust machine learning models capable of handling diverse linguistic inputs.

[1] Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., ... & Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv preprint arXiv:2104.08758.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

Data Collection Strategies for Instruction Tuning in LLMs with Low-Resource Languages

Διπλωματική Εργασία 29

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

This thesis delves into the exploration of data collection strategies crucial for instruction tuning in Language Learning Machines (LLMs) using low-resource languages. These languages are frequently marginalized on global digital platforms, lacking robust linguistic databases, which poses substantial challenges for effective data procurement and language model training. The thesis will critically evaluate existing research to uncover effective strategies for data collection and identify significant methodological gaps.

The project will actively involve the student in a current LLM project focusing on instruction tuning specifically tailored to a low-resource language. This involvement includes the development and implementation of targeted data collection protocols, and the use of advanced tools to gather, refine, and annotate linguistic data for instruction tuning purposes. Such practical engagement is aimed at providing insightful perspectives on the real-world complexities of data collection, thereby enhancing the student's ability to contribute meaningfully to the development of LLMs that can efficiently process and understand instructions in diverse languages.

[1] Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., ... & Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv preprint arXiv:2104.08758.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

On the synergy of large-scale learners in optimization tasks

Διπλωματική Εργασία 30

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Optimization
Τύπος Εργασίας	Research

Περιγραφή

Optimization tasks often seek to maximize or minimize a specific value under given constraints. Focusing on competitive environments, the optimal value may be reached forming a zero-sum game: only one competitor at a time can be benefitted, while the overall value should be maximized. Such environments can be auctions or buyer-seller scenarios, which have been thoroughly explored in classic algorithmic theory. However, such optimizations may be computationally expensive, while they can benefitted from advancements in deep learning, such as having powerful learners to perform a competitive strategy.
In this work, we aim to explore competitive scenarios, harnessing data from related environments, such as stock market and auctions, placing Large Language Models (LLMs) as the competitors. Each LLM tries to maximize their gain while minimizing the gain of the other, while the overall gain of the competitive game should be maximized. We examine various LLM roles -such as proponent, opponent, or mediator-, while also setting various rules (turn-taking, rebuttals, the use of evidence or logical reasoning etc) and constrains. We proceed with analyzing the strategies that emerge from the debate, using game-theoretic principles like Nash equilibrium, to understand how LLMs optimize their positions and adapt to the strategies of others. Several prompting strategies can be explored to achieve this equilibrium in models of ascending size, so that the cognitive abilities of LLMs in competition will be evaluated on par with their general reasoning perception. Finally, we explore whether larger models tend to specialize in certain types of arguments or strategies, and whether smaller models can generalize better across different topics.

Related work:
[1] Efficacy of Language Model Self-Play in Non-Zero-Sum Games Efficacy of Language Model Self-Play in Non-Zero-Sum Games (arxiv.org)
[2] Large Language Models Playing Mixed Strategy Nash Equilibrium Games https://arxiv.org/pdf/2406.10574
[3] Can Large Language Models Serve as Rational Players in Game Theory? A Systematic Analysis https://arxiv.org/abs/2312.05488
[4] How Well Can LLMs Negotiate? NEGOTIATIONARENA Platform and Analysis https://arxiv.org/pdf/2402.05863

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, γνώσεις αλγορίθμων ή/και θεωρίας παιγνίων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Κων/νο Θωμά (e-mail: kthomas at ails.ece.ntua.gr).

Knowledgeability of LLMs in competitive settings

Διπλωματική Εργασία 31

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Optimization
Τύπος Εργασίας	Research

Περιγραφή

LLMs can be considered as competitive agents that attempt to approach and achieve equilibrium states in related games by leveraging appropriate prompting strategies. The contribution of LLMs in solving challenges within competitive environments was described in the previous diploma thesis.
One open research question is whether LLMs can appropriately learn the optimal strategy when debating or if they have just memorized existing games, thus replicating the winning strategies they have learned during pre-training. To this end, we pursue to unlock their reasoning vs memorization balance by slightly tweaking competitive games and evaluate changes in winning strategies; in fact, we perform some counterfactual interventions on existing games (what would have happened if we minimally alter the game parameter X by replacing it with Y?). Moreover, we enhance our experimentation focusing on conversational exchanges between the LLMs and the user, while also prompting the intermediate reasoning steps of LLMs of varying parameter sizes. After experimenting with quantized and non-quantized models of different scales, we will be able to evaluate whether model performance in game-theoretic settings is mostly attributed to memorization or reasoning capabilities of related models.

Related work:
[1] Efficacy of Language Model Self-Play in Non-Zero-Sum Games Efficacy of Language Model Self-Play in Non-Zero-Sum Games (arxiv.org)
[2] Large Language Models Playing Mixed Strategy Nash Equilibrium Games https://arxiv.org/pdf/2406.10574
[3] Can Large Language Models Serve as Rational Players in Game Theory? A Systematic Analysis https://arxiv.org/abs/2312.05488
[4] Head-to-Tail: How Knowledgeable are Large Language Models (LLMs)? A.K.A. Will LLMs Replace Knowledge Graphs? https://aclanthology.org/2024.naacl-long.18.pdf

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, , γνώσεις αλγορίθμων ή/και θεωρίας παιγνίων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ecentua.gr), τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Κων/νο Θωμά (e-mail: kthomas at ails.ece.ntua.gr).

LLMs as Predictors in Dynamic Games: Real-Time Strategy Adaptation

Διπλωματική Εργασία 32

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, Optimization
Τύπος Εργασίας	Research

Περιγραφή

Dynamic games are scenarios where players must make decisions sequentially, with each player's strategy potentially evolving based on the observed actions of others. Unlike static games, where decisions are made simultaneously and without knowledge of others' choices, dynamic games require ongoing adaptation and prediction. In such settings, Large Language Models (LLMs) could be employed to predict the strategies of opponents and adjust their own strategies in real time, offering a novel application of AI in strategic decision-making.

In this work, we investigate the ability of LLMs to rapidly adapt their strategies in response to the actions of other players in dynamic game environments. We choose or design dynamic games that involve sequential decision-making (examples might include classical games like chess, negotiation scenarios, or even more complex real-time strategy games). More specifically, we implement LLMs that can observe the moves or decisions made by opponents in real time and adjust their strategies accordingly. The LLM should be able to predict likely future moves based on current observations. Moreover, we measure how quickly the LLM can adjust its strategy following an opponent's move and how accurately it predicts the opponent's future actions. The focus would be on real-time processing and the LLM's ability to maintain or gain strategic advantage as the game progresses.

As a second research objective, we explore how different prompting strategies can guide LLMs through multi-stage decision-making processes, optimizing for long-term gains rather than short-term victories. For example, we can provide the LLM with detailed historical context and ask it to weigh future outcomes based on this history, use decision tree prompts where the LLM must choose a path at each stage, with each choice leading to different consequences, and finally implement prompts that encourage the LLM to evaluate the risk and reward of each possible move, guiding it to balance immediate gains against long-term objectives.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, , γνώσεις αλγορίθμων ή/και θεωρίας παιγνίων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr) τον Γιώργο Φιλανδριανό (e-mail: geofila at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Κων/νο Θωμά (e-mail: kthomas at ails.ece.ntua.gr) .

Explainability in Large Vision-Language Models (LVLMs) survey

Διπλωματική Εργασία 33

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models, AI Computer for Vision
Τύπος Εργασίας	Survey

Περιγραφή

Large Vision-Language Models (LVLMs) have dominated several multimodal tasks, following the surge of purely linguistic Large Language Models (LLMs). Despite their impressive capabilities, there are still underexplored aspects such as model interpretability, transparency and fairness. Explainability issues in LLMs have been thoroughly studied in recent survey papers [1], even though related endeavors in LVLMs have not yet been published.
In this work, we aim to investigate explainability literature in LVLMs, and discuss potential opportunities and challenges related to existing art. To this end, we gather all existing papers in the area, propose an appropriate organization of them and analyze them in a comparative and contrastive manner. Ultimately, we deliver a survey paper that delves into the area of LVLM explainability with the goal of framing novel future research pathways.
Upon completing the survey paper, we proceed to the technical part of the thesis after acknowledging interesting research gaps.

[1] Explainability for Large Language Models: A Survey https://arxiv.org/abs/2309.01029

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, όραση υπολογιστών. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (e-mail: marialymp at ails.ece.ntua.gr).

Diffusion Models for Creative Image Editing

Διπλωματική Εργασία 34

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	Diffusion Models, AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Diffusion models, known for their capacity to model complex and high-dimensional distributions, have demonstrated exceptional performance in various tasks such as image synthesis, inpainting, and style transfer. However, the challenge of applying these models to fine-grained image editing tasks that require intricate, localized modifications, remains a significant hurdle.
To address this challenge, this thesis aims to systematically study existing diffusion-based methods for image editing and explore ways to extend their capabilities. It will provide an in-depth analysis of current diffusion model approaches, evaluating their performance in handling detailed and nuanced editing tasks. The research will focus on identifying the strengths and limitations of these models in contexts such as precise object manipulation, detailed texture adjustments, and complex style transformations.

[1] Contrastive Denoising Score for Text-guided Latent Diffusion Image Editing
https://arxiv.org/abs/2311.18608
[2] TiNO-Edit: Timestep and Noise Optimization for Robust Diffusion-Based Image Editing
https://arxiv.org/abs/2404.11120
[3] DiffEditor: Boosting Accuracy and Flexibility on Diffusion-based Image Editing
https://arxiv.org/abs/2402.02583

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

From Text to Motion: Leveraging Diffusion Models for Video Generation

Διπλωματική Εργασία 35

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision, Diffusion Models
Τύπος Εργασίας	Research

Περιγραφή

The burgeoning field of text-to-video generation aims to create dynamic video content directly from textual descriptions, revolutionizing how we generate and interact with multimedia content. Diffusion models, known for their impressive results in image synthesis and manipulation, offer significant potential for advancing this technology by bridging the gap between textual input and coherent video output.
In this context, the current thesis will begin with a comprehensive review of the latest methods employed in text-to-video generation using diffusion models. This will include an exploration of current state-of-the-art techniques, assessing their capabilities and limitations in producing high-quality, temporally consistent video sequences from text prompts. The review will provide a foundation for understanding how diffusion models can be effectively applied and improved in this emerging field.
Building on this foundation, the thesis will then focus on advancing these methods to enhance the performance of diffusion models in video generation. This will involve developing and refining techniques to improve the temporal coherence, visual fidelity, and contextual relevance of the generated videos.

[1] TI2V-Zero: Zero-Shot Image Conditioning for Text-to-Video Diffusion Models
https://arxiv.org/abs/2404.16306
[2] Grid Diffusion Models for Text-to-Video Generation
https://arxiv.org/abs/2404.00234
[3] Generative Rendering: Controllable 4D-Guided Video Generation with 2D Diffusion Models
https://arxiv.org/abs/2312.01409

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr) και το Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

Leveraging Diffusion Models for speech and emotional driven 3D Body Animation

Διπλωματική Εργασία 36

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	Diffusion Models, AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

This thesis aims to explore the use of diffusion models to generate 3D body animations directly from speech, with a special focus on capturing and representing emotional content. The research lies at the intersection of machine learning, digital signal processing, and computer graphics, seeking to bridge the gap between verbal expression and embodied physical movement. The thesis will begin with a comprehensive study of diffusion models, which excel at generating complex, high-dimensional data by progressively refining noisy inputs. Next, it will examine existing techniques for speech-driven 3D body animation, focusing on how features like tone, pitch, and rhythm are used to drive movement. After identifying limitations in capturing emotional depth, the focus will shift to enhancing these models by integrating emotion recognition from speech. This will enable the generation of animations that not only synchronize with speech but also reflect emotional content, resulting in more expressive and dynamic body movements.

[1] Emotional Speech-driven 3D Body Animation via Disentangled Latent Diffusion
https://arxiv.org/abs/2312.04466

[2] DiffSHEG: A Diffusion-Based Approach for Real-Time Speech-driven Holistic 3D Expression and Gesture Generation
https://arxiv.org/abs/2401.04747

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

Diffusion Models for 3D Data Generation: A Survey

Διπλωματική Εργασία 37

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	Diffusion Models, AI for Computer Vision
Τύπος Εργασίας	Survey

Περιγραφή

In this thesis we will conduct a comprehensive survey of diffusion models in the context of 3D data generation, with a particular emphasis on text-to-3D and image-to-3D model generation. Diffusion models, which have gained significant attention in recent years for their ability to generate complex data structures, offer promising solutions for synthesizing realistic 3D models from different modalities. The thesis will begin by exploring the foundational concepts of diffusion processes, followed by an in-depth review of state-of-the-art approaches in 3D data generation. A specific focus will be placed on recent advancements that leverage diffusion models to generate 3D models from textual descriptions and 2D images, highlighting the challenges, applications, and performance metrics. The survey aims to provide insights into the current trends, methodologies, and future directions for integrating diffusion models in the evolving field of 3D content creation. Through this, the thesis will contribute to a better understanding of the potential and limitations of diffusion-based techniques in 3D generation and their role in bridging the gap between various data modalities.

[1] Direct2.5: Diverse Text-to-3D Generation via Multi-view 2.5D Diffusion
https://arxiv.org/abs/2311.15980
[2] GaussianDreamer: Fast Generation from Text to 3D Gaussians by Bridging 2D and 3D Diffusion Models
https://arxiv.org/abs/2310.08529
[3] Bayesian Diffusion Models for 3D Shape Reconstruction
https://arxiv.org/abs/2403.06973
[4] Wonder3D: Single Image to 3D using Cross-Domain Diffusion
https://arxiv.org/abs/2310.15008

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr) τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr) και τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

Creation of Music Listening Dataset

Διπλωματική Εργασία 38

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	AI for Music
Τύπος Εργασίας	Dataset Creation

Περιγραφή

Previous studies have focused on creating datasets for music emotion recognition, often incorporating biosignals due to their strong correlation with emotional responses. In this thesis, the student will develop a music listening dataset that integrates both biosignals and emotional annotations. Participants will listen to various musical pieces while wearing a headband to record EEG data, along with sensors for ECG (electrocardiogram) and GSR (galvanic skin response). During the sessions, they will use annotation tools to log their emotional reactions in real-time and answer additional questions to capture their subjective experience. The aim is to create a comprehensive dataset that combines physiological data with emotional and cognitive annotations, advancing research in music cognition.

[1] Zhang, Kejun, et al. "The PMEmo dataset for music emotion recognition." *Proceedings of the 2018 acm on international conference on multimedia retrieval*. 2018.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr), τον Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr) και τον Έντι Ντερβάκο (e-mail: eddiedervakos at ails.ece.ntua.gr).

Collision Avoidance

Διπλωματική Εργασία 39

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Sense and Avoid (SAA) techniques are critical in autonomous systems, especially in unmanned aerial vehicles (UAVs) and other autonomous robots. These techniques allow the system to detect obstacles or potential hazards in its environment and take action to avoid collisions. The challenge becomes even greater when dealing with non-cooperative aerial vehicles, which provide little to no external information about their movements. To build more reliable SAA frameworks, recent advancements have focused on using optical sensors to develop end-to-end systems for detecting and avoiding non-cooperative UAVs. This thesis will explore the key components of collision avoidance frameworks—detection, tracking, and distance estimation—by addressing their current limitations and enhancing their robustness and generalization in challenging conditions.

[1] Domain Generalization: A Survey
[2] NEFELI: A deep-learning detection and tracking pipeline for enhancing autonomy in Advanced Air Mobility
[3] Common Corruptions for Evaluating and Enhancing Robustness in Air-to-Air Visual Object Detection

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr) και τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

Surface Normal Guided Models for Robust Depth Estimation

Διπλωματική Εργασία 40

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Monocular Depth Estimation (MDE) is a fundamental problem in computer vision that involves predicting the depth (distance from the camera) of each pixel in an image using only one image. Recent methods like MiDaS and Depth Anything showcased promising results in zero-shot generalization across multiple unseen datasets.

Other methods focus on introducing additional information to achieve robust models. Methods like NDDepth integrate surface normal information offering a robust framework for enhancing monocular depth estimation. Although the model showcased promising results, limited experiments were conducted on their generalization capabilities. This thesis aims to study monocular depth estimation techniques and enhance their robustness in out-of-distribution (ODD) data.

[1] NDDepth Normal-Distance Assisted Monocular Depth Estimation
[2] Rethinking Inductive Biases for Surface Normal Estimation

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

Enhancing Visual Concept Comprehension During Inference

Διπλωματική Εργασία 41

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Domain adaptation remains a significant problem for visual models. Usually, adaptation to new tasks requires either meta-learning or fine-tuning, both of which are resource-expensive. Moreover, the paradigm of fine-tuning at inference time, widely used in large language models, has yet to be primarily explored for visual models. Notably, recent works, such as CAML, demonstrated surprisingly good performance by reformulating meta-learning as sequence modeling, similar to in-context learning in LLMs. Such techniques have achieved remarkable progress in learning visual concepts at inference. Based on this, this thesis aims to comprehensively review in-context learning techniques on visual concepts and further advance those techniques generalization ability towards extracting knowledge in visual context.

[1] CONTEXT-AWARE META-LEARNING
[2] A Closer Look at Few-shot Classification Again
[3] What Makes Good Examples for Visual In-Context Learning?

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

3D scene synthesis by leveraging single or sparse views

Διπλωματική Εργασία 42

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Survey

Περιγραφή

The field of 3D synthesis has experienced significant advancements, primarily driven by the growing demand for high-quality 3D content across various industries. Traditionally, the creation of 3D models required extensive input data and the use of specialized tools, rendering the process both time-consuming and technically complex. However, recent developments have introduced more efficient methodologies that enable the generation of detailed 3D models from minimal inputs, such as a single image or sparse views. Notably, diffusion models have emerged as a promising solution to the challenges inherent in 3D synthesis. These models offer a novel approach to generating coherent 3D structures from incomplete or limited data by simulating multiple perspectives and integrating them into a comprehensive 3D reconstruction. Such advancements are transforming 3D content creation, allowing for faster workflows without sacrificing quality. This thesis will undertake a comprehensive review of the latest progress in 3D scene generation from single or sparse views, focusing on the role of diffusion models in enhancing efficiency and accessibility in this domain.

[1] CAT3D: Create Anything in 3D with Multi-View Diffusion Models
[2] MVDREAM: MULTI-VIEW DIFFUSION FOR 3D GENERATION

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr) και τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr).

Occluded normal surface completion for 3d reconstruction

Διπλωματική Εργασία 43

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

The diploma thesis aims to enhance 3D scene reconstruction techniques by integrating multi-view normal map data with methods for completing occluded surfaces. Traditional 3D reconstruction methods often face challenges in accurately capturing fine surface details and dealing with areas that are hidden from view, which impacts the overall realism and completeness of the resulting 3D models. This research seeks to overcome these issues by developing a hybrid approach that combines two advanced techniques: leveraging multi-view normal maps to capture intricate surface details and employing depth-based methods to reconstruct occluded surfaces. Through this integration, the thesis will investigate how to produce more accurate, detailed, and complete 3D reconstructions of real-world scenes.

[1] Behind the Veil: Enhanced Indoor 3D Scene Reconstruction with Occluded Surfaces Completion
[2] SuperNormal: Neural Surface Reconstruction via Multi-View Normal Integration

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr) και τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr).

The in-n-outs of LLMs: Data Provenance and Output Watermarking

Διπλωματική Εργασία 44

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

New York Times vs. OpenAI, Getty Images vs. Stability AI—these legal battles reflect a growing concern in the world of AI: how can we prove that specific datasets were used to train powerful generative models like Large Language Models (LLMs) and image generator? As proprietary data gets swept into the training of these models, it raises critical questions about data attribution and the rights of content creators.
At the same time, another crucial task arises: watermarking AI-generated outputs to ensure their authenticity and traceability. Watermarking the outputs of LLMs can serve as a safeguard, enabling us to trace the origin of AI-generated content and verify its authenticity. This is crucial not only for protecting intellectual property but also for establishing accountability in a legal and ethical landscape where AI-generated material is becoming indistinguishable from human-created content.

And then a crucial question arises: could these two problems be connected? What can we learn from data watermarking to support data attribution or provenance? Can methodologies from watermarking outputs be adapted to facilitate tracing the datasets that trained these models? In this thesis, we will explore these two areas. The study will begin with an extensive survey of data attribution and data watermarking techniques. Then, we will investigate the possibility to combine methods from the two areas to improve the performance of both tasks, giving a special focus on data attribution.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), και τον Ορφέα Μενή Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

Development of Expressive Graph Neural Networks based on Heuristic Algorithms

Διπλωματική Εργασία 45

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Research

Περιγραφή

Graph Neural Networks (GNNs) have recently emerged as a powerful tool for performing machine learning on graphs. A considerable amount of recent work has focused on characterizing the expressive power of GNNs [1,2,3]. Expressive power of GNNs is typically measured in terms of their ability to distinguish non-isomorphic graphs. For example, standard GNNs were shown to be at most as powerful as the Weisfeiler-Leman (WL) algorithm (a well-known heuristic for deciding whether two graphs are isomorphic) in terms of distinguishing non-isomorphic graphs [1,2]. To develop models that are more powerful than WL, a common approach is to take an algorithm (which is known to be more powerful than WL), and build a neural network that follows the steps of that algorithm. For instance, several recent models are inspired from high order variants of WL [3,4] or from other algorithms [5]. The objective of this thesis is to capitalize on already-established heuristics for the graph isomorphism problem, and derive neural network models from them. This will result into models that can be as powerful as their associated heuristics in terms of distinguishing non-isomorphic graphs. The models that will be developed will also be empirically evaluated on the BREC dataset, a dataset that contains hard to distinguish pairs of non-isomorphic graphs such as instances of the CFI graphs.

[1] Xu, K., Hu, W., Leskovec, J., & Jegelka, S. "How powerful are graph neural networks?". In ICLR'19.
[2] Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., & Grohe, M. "Weisfeiler and leman go neural: Higher-order graph neural networks". In AAAI'19.
[3] Feng, J., Chen, Y., Li, F., Sarkar, A., & Zhang, M. "How powerful are k-hop message passing graph neural networks". In NeurIPS'22.
[4] Morris, C., Rattan, G., & Mutzel, P. "Weisfeiler and leman go sparse: Towards scalable higher-order graph embeddings". In NeurIPS'2020.
[5] Thiede, E., Zhou, W., & Kondor, R. Autobahn: Automorphism-based graph neural nets. In NeurIPS'21.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Extending DiGress for Generating Higher-Quality Graphs

Διπλωματική Εργασία 46

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Graph Neural Networks
Τύπος Εργασίας	Research

Περιγραφή

One of the most challenging tasks of machine learning on graphs is that of graph generation [1]. Graph generation has attracted a lot of attention recently and its main objective is to create novel and realistic graphs. For instance, in chemo-informatics, graph generative models are employed to generate novel, realistic molecular graphs which also exhibit desired properties (e.g., high drug-likeness) [2]. Among the various generative models that have been proposed recently, DiGress has attracted a lot of attention [3]. DiGress is a discrete denoising diffusion model which can generate graphs with categorical node and edge attributes. With regards to the graph structure, the model considers only two edge types: (1) an edge exists between two nodes, and (2) no edge exists between the two nodes. This leads to transition matrices of dimension 2 x 2 which might prevent the model from capturing the full complexity of the input data. The objective of this thesis is to modify DiGress by increasing the edge types. The types can for example capture the shortest path distances between nodes. This would clearly lead to more than two edge types. The model that will be developed will be evaluated on both molecular and non-molecular benchmarks.

[1] Zhu, Y., Du, Y., Wang, Y., Xu, Y., Zhang, J., Liu, Q., & Wu, S. "A survey on deep graph generation: Methods and applications". In LoG'22.
[2] Zang, C., & Wang, F. "Moflow: an invertible flow model for generating molecular graphs". In KDD'20.
[3] Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., & Frossard, P. "Digress: Discrete denoising diffusion for graph generation". In ICLR'23.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Investigating Tasks/Settings where Graph Transformers can be Superior to MPNNs

Διπλωματική Εργασία 47

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Research

Περιγραφή

Over the past few years, Graph Neural Networks (GNNs) have achieved great success in solving machine learning problems on graph data. Most GNNs can actually be seen as Message Passing Neural Networks (MPNNs) where vector messages are exchanged between nodes and updated using neural networks [1]. In the past years, Graph Transformers have emerged as a promising alternative to MPNNs [2,3]. In Graph Transformers, nodes cannot only communicate with their direct neighbors, but also with the rest of the nodes of the graph. Besides their attractive properties, it was recently shown that MPNNs, when augmented with a virtual node, can approximate the self-attention mechanism of Graph Transformers under certain conditions [4]. It was also empirically shown that Graph Transformers do not necessarily outperform MPNNs in tasks where long-range dependencies between nodes need to be captured [5]. Given the increased complexity of Graph Transformers compared to that of MPNNs, there might exist some scenarios where Graph Transformers can succeed, but MPNNs fail. The objective of this thesis is to identify such scenarios.

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. "Neural message passing for quantum chemistry". In ICML'17.
[2] Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., et al. "Do transformers really perform badly for graph representation?". In NeurIPS'21.
[3] Rampášek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., & Beaini, D. "Recipe for a general, powerful, scalable graph transformer". In NeurIPS'22.
[4] Cai, C., Hy, T. S., Yu, R., & Wang, Y. "On the connection between mpnn and graph transformer". In ICML'23.
[5] Tönshoff, J., Ritzert, M., Rosenbluth, E., & Grohe, M. (2025). "Where did the gap go? reassessing the long-range graph benchmark". TMLR.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Graph Structure Learning for Protein Representation Learning

Διπλωματική Εργασία 48

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Research

Περιγραφή

In this project, we aim to advance protein representation learning by applying graph structure learning techniques. Unlike traditional GNN approaches which work with static graph structures, our model will dynamically learn the graph structure during training. This involves constructing initial graphs where nodes represent amino acids and edges represent known physical interactions. The model will then adaptively modify these edges and potentially introduce new connections based on patterns learned from data. This adaptive connectivity is enabled through techniques such as attention mechanisms, which allow the model to focus on and reinforce connections that are most informative for predicting protein functions and interactions. By allowing the model to learn and optimize the graph structure, we anticipate more accurate representations of protein dynamics and interactions, leading to better predictions of their biological functions.

Skills:
- Familiarity with Python and deep learning frameworks (PyTorch, TensorFlow).
- Experience with graph neural networks (GNNs).
- Familiarity with attention mechanisms.

Relevant References:
[1] Qian, Chendi, et al. "Probabilistically Rewired Message-Passing Neural Networks." arXiv preprint arXiv:2310.02156 (2023).
[2] Kazi, Anees, et al. "Differentiable graph module (dgm) for graph convolutional networks." IEEE Transactions on Pattern Analysis and Machine Intelligence 45.2 (2022): 1606-1617.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Linearization for graphs and combination with LLMS.

Διπλωματική Εργασία 49

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks, Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

This project aims to develop methods to linearize graph structures so they can be effectively used with large language models (LLMs) in sequence-based transformer models. Graph structures, such as those used to represent networks, molecules, or social connections, are inherently non-linear, posing a challenge when integrating them into traditional transformer architectures designed for sequential data.
The goal of this project is to explore and design techniques that convert graphs into sequences while preserving essential structural information. By representing the nodes and edges of a graph in a linear format, we can leverage the power of LLMs to process and extract meaningful patterns from these structures. Various strategies, such as depth-first search (DFS), breadth-first search (BFS), and novel serialization approaches, will be explored for graph-to-sequence transformation.
Once the graphs are linearized, they can be fed into pre-trained transformers or LLMs for tasks such as node classification, graph generation, or property prediction. This combination of graph theory and LLMs could unlock new potential for analyzing complex data structures across diverse domains such as biology, social networks, and chemical analysis.

Skills:
- Foundation in graph theory and deep learning (especially transformer models).
- Proficiency in Python and libraries like PyTorch or TensorFlow.
- Experience with graph neural networks (GNNs) and LLMs.

Relevant References:
[1] Abdine, Hadi, et al. "Prot2text: Multimodal protein's function generation with gnns and transformers." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 10. 2024.
[2] Min, Erxue, et al. "Transformer for graphs: An overview from architecture perspective." arXiv preprint arXiv:2202.08455 (2022).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Molecule-Protein-Cell to Text

Διπλωματική Εργασία 50

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

This project focuses on developing AI models that can generate meaningful textual descriptions for various biological entities, including molecules, proteins, and cells. These descriptions may include functional roles, interactions, or any relevant biological characteristics. The goal is to bridge the gap between raw biological data and human-readable, interpretable text, making complex biological information more accessible for research and communication.
The approach involves training large-scale models that can understand and predict text based on the inherent properties and interactions of biological entities. Using advanced deep learning techniques, such as transformer architectures, the model will learn to map biological data (such as molecular structures, protein sequences, or cellular characteristics) to natural language descriptions. This could be used for various applications, including drug discovery, personalized medicine, and educational resources.

Skills:
- Understanding of deep learning, particularly transformers and language models.
- Familiarity with biological data formats (e.g., SMILES, FASTA) and databases (e.g., PubMed, UniProt).
- Familiarity in Python and deep learning libraries (e.g., PyTorch, TensorFlow).

Relevant References:
[1] Abdine, Hadi, et al. "Prot2text: Multimodal protein's function generation with gnns and transformers." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 10. 2024.
[2] Zhao, Suyuan, et al. "LangCell: Language-Cell Pre-training for Cell Identity Understanding." arXiv preprint arXiv:2405.06708 (2024).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Large Models for Language and Graphs

Διπλωματική Εργασία 51

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks, Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

LLMs for Standard and Dialectal Arabic: Arabic language has a rich history and profound cultural significance, with its intricate script, vast lexicon, and complex grammar, making it a unique linguistic entity. The objective will be developing or enhancing the dialectal capabilities of current Arabic LLMs mainly trained on standard Arabic, e.g., JAIS model built by MBZUAI [1], involving tasks like data collec- tion, training and benchmark dataset creation, distributed training, and evaluation. Arabic related tasks include dialect identification, tashkeel/diacritics restoration, poetry analysis and generation, dialect lan- guage to official/dialect language translation and analysis.

[1] Sengupta, Neha, et al. ”Jais and jais-chat: Arabic-centric foundation and instruction-tuned open generative large language models.” arXiv preprint arXiv:2308.16149 (2023).
Synthetic Data Creation: Recently, synthetic data has emerged as a significant contributor to LLM development, often surpassing the performance of human-curated data. Techniques such as Self-Instruct and RLAIF have become prevalent for generating data used in both instruction-tuning and preference-tuning. The aim is to develop innovative methods for generating high-quality synthetic data by leveraging existing LLMs, as demonstrated in works like [2, 3].
[2] Cheng, Daixuan, Shaohan Huang, and Furu Wei. ”Adapting large language models via reading co
m-
prehension.” The Twelfth International Conference on Learning Representations. 2023.
[3] Li, Ming, et al. ”Selective reflection-tuning: Student-selected data recycling for llm instruction-tuning.” arXiv preprint arXiv:2402.10110 (2024).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

LLM Evaluation from Novel Perspectives

Διπλωματική Εργασία 52

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks, Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

This research direction aims at the evaluation of LLMs from novel perspectives, such as reasoning, factuality, knowledge, bias, and linguistic diversity [1], diverging from the usual emphasis on multi-task-solving benchmark performance. In addition, we are interested in the evaluation of model collapse [2], which refers to performance degradation caused by recursively training models on synthetic data produced by previous models.

[1] Guo, Yanzhu, et al. ”The curious decline of linguistic diversity: Training language models on synthetic text.” arXiv preprint arXiv:2311.09807 (2023).
[2] Shumailov, Ilia, et al. ”AI models collapse when trained on recursively generated data.” Nature 631.8022 (2024): 755-759.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Data Efficient Training with Curriculum Learning

Διπλωματική Εργασία 53

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

Training LLMs requires substantial energy and electricity resources, leading to both ecological and economic challenges. This highlights the importance of developing data-efficient training approaches. The objective is to create curriculum learning metrics that establish an optimal ordering of LLM training data, allowing for the same performance to be achieved with fewer training steps. Such a metric could be diversity, information density, uncertainty, and skills [6].

[6] Chen, Mayee, et al. ”Skill-it! a data-driven skills framework for understanding and training language models.” Advances in Neural Information Processing Systems 36 (2024).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Multimodal Models & LLMs for Graphs

Διπλωματική Εργασία 54

Επιβλέπων	Μιχάλης Βαζιργιάννης, Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks, Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

Recently, the use of multi-modality in LLMs has become popular, yielding impressive results in generating images, videos, and sounds. However, graph—a ubiquitous representation commonly used to model social networks, knowledge graphs, bio/chemical data—remains a comparably emerging and underdeveloped modality in this research area, leaving room for new and effective approaches. The objective will be to investigate methods for augmenting LLMs with capacities of graph understanding and generation, e.g., Prot2Text [7] and Talk like a graph [8].

[7] Abdine, Hadi, et al. ”Prot2text: Multimodal protein’s function generation with gnns and transformers.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 10. 2024.
[8] Fatemi, Bahare, Jonathan Halcrow, and Bryan Perozzi. ”Talk like a graph: Encoding graphs for large language models.” arXiv preprint arXiv:2310.04560 (2023).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Μ. Βαζιργιάννη (e-mail: mvazirg@lix.polytechnique.fr) και τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr).

Graph Neural Networks Robustness on Node Classification

Εργασία Μαθήματος 1

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Survey

Περιγραφή

Graph Neural Networks (GNNs) have emerged as a powerful paradigm for learning representations of graph-structured data, finding applications in diverse domains such as social networks, recommendation systems, and bioinformatics. Their ability to capture complex relationships among entities in graphs has led to significant advancements in tasks like node classification, link prediction, and graph classification. However, despite their effectiveness, GNNs are not immune to challenges, particularly when confronted with noisy or adversarial data. Understanding the robustness of GNNs, especially in critical tasks like node classification, is essential for ensuring their reliability and applicability in real-world scenarios.

Node classification, a fundamental task in graph analysis, involves predicting the class labels of nodes based on their attributes and the topology of the graph. The performance of GNNs on node classification tasks has been extensively studied across various datasets and scenarios. While GNNs often exhibit impressive performance in ideal conditions, their vulnerability to perturbations in graph structure or node features raises concerns about their robustness in practical settings. Adversarial attacks targeting GNNs have demonstrated the potential vulnerabilities of these models, highlighting the need for robustness-enhancing techniques to mitigate such threats. This literature review aims to offer a critical analysis of existing studies, elucidating the challenges, methodologies, and findings surrounding the robustness of Graph Neural Networks in node classification tasks. The main focus will be the different types of adversarial attacks/perturbations, their impact on existing GNNs, and proposed techniques designed to make GNNs robust.

Example papers:
[1]https://www.semanticscholar.org/paper/FocusedCleaner%3A-Sanitizing-Poisoned-Graphs-for-Node-Zhu-Tong/e01f18b468b8e4c3634b1ef3ca5faca019c3ee2c
[2]https://www.semanticscholar.org/paper/Robust-Node-Classification-on-Graphs%3A-Jointly-from-Zhuang-Hasan/0e8edb00c2e822bddccba8fc34c662ce19d6133d
[3]https://www.semanticscholar.org/paper/RoboGNN%3A-Robustifying-Node-Classification-under-Guan-Ma/548dfca32d996da0934a3b3e924d818c1fcbcd18

Απαιτούμενες/επιθυμητές γνώσεις: Python, θεωρητικό υπόβαθρο και μεθοδολογίες για γράφους, εξοικείωση με pytorch και PyG. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr)

Reproducing and Testing GNN Causal and Counterfactual Explanations

Εργασία Μαθήματος 2

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks, Explainable AI
Τύπος Εργασίας	Reproduction

Περιγραφή

This exercise focuses on reproducing and extending a state-of-the-art method for generating explainable Graph Neural Network (GNN) predictions [1]. The paper you will work with introduces the CF2 framework, which combines both Counterfactual and Factual reasoning from causal inference to generate high-quality explanations for GNN predictions. Unlike traditional GNN explainability methods that only consider one causal perspective, CF2 optimizes both, making it a more comprehensive approach to interpreting complex structural data.
After reproducing the original paper's results, the primary goal of this exercise is to extend the evaluation by applying CF2 to a new dataset: Graph-SST2 [2], a sentiment analysis dataset represented as graphs, often used in explainability research. The experiments will focus mainly on generating qualitative insights, assessing how well the CF2 framework provides meaningful and interpretable explanations in this new context. By exploring its performance on a dataset outside of the original study's scope, this exercise will offer a deeper understanding of the method's generalizability across different domains. The student is expected to gain a deeper understanding of graph techniques, counterfactual reasoning and a brief introduction to causal inference.

[1] https://dl.acm.org/doi/pdf/10.1145/3485447.3511948
[2] https://github.com/divelab/DIG/tree/main/dig/xgraph/datasets

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr), τον Βασίλης Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr)

GNN Confidence Analysis on Graph Classification

Εργασία Μαθήματος 3

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Graph Neural Networks
Τύπος Εργασίας	Reproduction

Περιγραφή

Graph Neural Networks (GNNs) have emerged as a powerful paradigm for learning representations of graph-structured data, finding applications in diverse domains such as social networks, recommendation systems, and bioinformatics. Their ability to capture complex relationships among entities in graphs has led to significant advancements in tasks like node classification, link prediction, and graph classification. However, despite their effectiveness, GNNs are not immune to challenges, particularly when confronted with noisy or adversarial data. Understanding the robustness of GNNs, especially in critical tasks like node classification, is essential for ensuring their reliability and applicability in real-world scenarios.

Node classification, a fundamental task in graph analysis, involves predicting the class labels of nodes based on their attributes and the topology of the graph. The performance of GNNs on node classification tasks has been extensively studied across various datasets and scenarios. While GNNs often exhibit impressive performance in ideal conditions, their vulnerability to perturbations in graph structure or node features raises concerns about their robustness in practical settings. Adversarial attacks targeting GNNs have demonstrated the potential vulnerabilities of these models, highlighting the need for robustness-enhancing techniques to mitigate such threats. This literature review aims to offer a critical analysis of existing studies, elucidating the challenges, methodologies, and findings surrounding the robustness of Graph Neural Networks in node classification tasks. The main focus will be the different types of adversarial attacks/perturbations, their impact on existing GNNs, and proposed techniques designed to make GNNs robust.

Example papers:
[1]https://www.semanticscholar.org/paper/FocusedCleaner%3A-Sanitizing-Poisoned-Graphs-for-Node-Zhu-Tong/e01f18b468b8e4c3634b1ef3ca5faca019c3ee2c
[2]https://www.semanticscholar.org/paper/Robust-Node-Classification-on-Graphs%3A-Jointly-from-Zhuang-Hasan/0e8edb00c2e822bddccba8fc34c662ce19d6133d
[3]https://www.semanticscholar.org/paper/RoboGNN%3A-Robustifying-Node-Classification-under-Guan-Ma/548dfca32d996da0934a3b3e924d818c1fcbcd18

Απαιτούμενες/επιθυμητές γνώσεις: Python, θεωρητικό υπόβαθρο και μεθοδολογίες για γράφους, εξοικείωση με pytorch και PyG. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Νίκο Χάιδο (e-mail: nchaidos at ails.ece.ntua.gr), την Αγγελική Δημητρίου (e-mail: angelikidim at ails.ece.ntua.gr) και τον Βασίλη Λυμπεράτο (e-mail: vaslyb at ails.ece.ntua.gr)

Deep Learning and Image Amodality

Εργασία Μαθήματος 4

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Survey

Περιγραφή

Amodal completion is the task of predicting the whole shape and appearance of objects that are not fully visible, and this ability is crucial for many downstream applications in vision, graphics, and robotics. What makes amodal completion challenging compared to other synthesis tasks is that it requires grouping for both the visible and hidden parts of an object. To complete an object, we must be able to first recognize the object from partial observations, then synthesize only the missing regions for the object. Existing computer vision systems can compete with humans in understanding the visible parts of objects, but still fall far short of humans when it comes to depicting the invisible parts of partially occluded objects. Image amodal completion aims to equip computers with human-like amodal completion functions to understand an intact object despite it being partially occluded. In this work, we want to formulate a survey with recent advances in image amodality and review contributions as a continuation for a diploma thesis.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

Optical Flow Estimation

Εργασία Μαθήματος 5

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Survey

Περιγραφή

Motion analysis is a key and complex problem in computer vision, with numerous applications in areas like autonomous driving, action recognition, scene understanding, and robotics. Typically, the motion between consecutive frames is categorized into two types: optical flow and scene flow. Optical flow captures the pixel-level motion between adjacent frames, while scene flow represents the 3D movement within the dynamic scene between frames. Traditional methods for estimating optical and scene flow often rely on variational techniques, solved through energy minimization. Recently, deep learning has become a powerful tool for learning feature representations directly from data. Especially with the rise of Neural Radiance Fields, optical flow estimation has known significant progress. In this work, we want to formulate a survey with recent advances in optical flow estimation and review contributions as a continuation for a diploma thesis.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Ν. Σπανό (e-mail: nspanos at ails.ece.ntua.gr).

Mathematical Foundations of Diffusion Models: A Probabilistic and Stochastic Approach

Εργασία Μαθήματος 6

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	Diffusion Models, AI for Computer Vision
Τύπος Εργασίας	Survey

Περιγραφή

This survey will explore the mathematical background and principles of diffusion models, tracing their development from early theoretical foundations to modern applications. Beginning with the origins of diffusion theory in physics and probability, the research will examine the historical evolution of diffusion models, focusing on foundational concepts such as Brownian motion and the Wiener process.
The study will then delve deeper into the mathematical formulations that describe the forward diffusion process and will explore various methods used to formulate and solve the inverse diffusion process. The survey will culminate in an analytical presentation of the most significant mathematical formulations that govern the principles of diffusion models, providing a comprehensive understanding of both their theoretical structure and practical implications.

[1] Denoising Diffusion Probabilistic Models
https://arxiv.org/abs/2006.11239
[2] Score-Based Generative Modeling through Stochastic Differential Equations
https://arxiv.org/abs/2011.13456
[3] Generative Modeling by Estimating Gradients of the Data Distribution
https://arxiv.org/abs/1907.05600
[4] Denoising Diffusion Implicit Models
https://arxiv.org/abs/2010.02502

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr).

Optimal Decision Tree Training with Mixed Integer Programming Formulations

Εργασία Μαθήματος 7

Επιβλέπων	Γιώργος Στάμου
Status	Διαθέσιμη
Περιοχή	Explainable AI, Optimization
Τύπος Εργασίας	Reproduction

Περιγραφή

The goal of this project is to reproduce results from papers which use Mixed Integer Programming to train Optimal Decision Trees and which do not publicly provide code for their implementation.
The most widely adopted algorithms for decision tree training use greedy top-down induction (e.g. CART and C4.5), but the results of such methods are usually suboptimal [1]. The problem of finding the best possible decision tree for a training objective is known to be NP-complete, but advancements in computational power and algorithms has made this problem relatively tractable. Algorithms for finding optimal decision trees mostly involve Mixed Integer Programming [2], Branch and Bound [3] or MaxSAT formulations [4].
The main advantage of MIP formulations is the ease with which they can be modified to include different objectives or regularization terms and that they can be often formulated in a few lines of code, with an existing MIP solver doing the heavy lifting. Unfortunately, many works which use MIP formulations to compute optimal trees do not provide their implementations publicly, which makes further research unnecessarily difficult. For this project, you will be called to create a public implementation of one such publication. Several are available and more can be found upon request. [5-7]

[1] Optimal or Greedy Decision Trees? Revisiting their Objectives, Tuning, and Performance
https://arxiv.org/abs/2409.12788
[2] Optimal classification trees
https://link.springer.com/article/10.1007/S10994-017-5633-9
[3] Optimal Sparse Decision Trees
https://arxiv.org/abs/1904.12847
[4] Learning Optimal Decision Trees with SAT
https://www.ijcai.org/proceedings/2018/189
[5] Optimal decision trees for categorical data via integer programming
https://arxiv.org/abs/1612.03225
[6] Robust Optimal Classification Trees under Noisy Labels
https://arxiv.org/abs/2012.08560
[7] New optimization models for optimal classification trees
https://www.sciencedirect.com/science/article/abs/pii/S0305054823003799

Απαιτούμενες/επιθυμητές γνώσεις: Python, βασικές γνώσεις για Decision Trees, Mathematical Programming και Optimization. Για περισσότερες πληροφορίες επικοινωνήστε με τον Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και τον Ιάσονα Λιάρτη (e-mail: jliartis at ails.ece.ntua.gr).

Advancing Few-Shot Classification with Corrupted Image Modeling: A Unified Approach Across Architectures

Εργασία Μαθήματος 8

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Reproduction

Περιγραφή

Few-shot classification poses the challenge of requiring models to generalize to new classes with minimal labeled examples, a task that traditional supervised learning methods struggle to address effectively. Masked Image Modeling (MIM) has demonstrated success in self-supervised pre-training, particularly for Vision Transformers (ViTs). However, the reliance on [MASK] tokens creates a misalignment between the pre-training and fine-tuning phases, which negatively impacts performance, especially in few-shot learning scenarios. Furthermore, utilizing this technique poorly adapts to CNN-based architectures due to the information leakage inherent in CNNs' sliding-window architecture
To overcome these challenges, Corrupted Image Modeling (CIM) provides a more flexible and effective alternative. Rather than using [MASK] tokens, CIM employs a trainable auxiliary generator to corrupt input images by replacing patches with plausible alternatives, resulting in more natural corruptions.
The aim of this diploma thesis is to explore improvements in generalization and adaptability across different architectures, particularly in fields such as medical imaging, where available data is limited.

[1] Rethinking Generalization in Few-Shot Classification
[2] Corrupted image modeling for self-supervised visual pre-training

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr).

Addressing Challenges in Monocular 3D Object Detection: Enhancing Depth Estimation and 3D Feature Representation.

Εργασία Μαθήματος 9

Επιβλέπων	Αθανάσιος Βουλόδημος
Status	Διαθέσιμη
Περιοχή	AI for Computer Vision
Τύπος Εργασίας	Survey

Περιγραφή

Monocular 3D object detection is a computer vision task aiming to detect and estimate objects' 3D position, size, and orientation using only a single camera image. It has garnered significant attention due to its wide range of applications in fields such as autonomous driving, robotics, and augmented reality, where cost-effectiveness and simplicity are crucial. Recent advances have introduced techniques that leverage deep learning and geometrical cues to predict 3D information from 2D images, making it feasible to perform 3D detection without expensive LiDAR sensors. These methods, such as those incorporating Neural Radiance Fields (NeRF) and multi-depth estimations, have notably improved detection accuracy and performance. However, despite these advances, monocular 3D object detection still faces fundamental limitations, particularly in accurately estimating depth and handling occluded or distant objects. Many current approaches depend heavily on precise depth estimations, which remain challenging to obtain reliably from a single image.
Additionally, these methods often struggle to generate dense and accurate 3D feature representations, especially in complex environments or when data comes from diverse domains like indoor and outdoor scenes. This research aims to conduct an extensive review of the state-of-the-art techniques in monocular 3D object detection, identify their limitations, and propose new methods to address the challenges related to depth estimation, 3D feature representation, and generalization across various environments.

[1] MonoCD: Monocular 3D Object Detection with Complementary Depths
[2] UniMODE: Unified Monocular 3D Object Detection
[3] MonoNeRD: NeRF-like Representations for Monocular 3D Object Detection
[4] Lidar Point Cloud Guided Monocular 3D Object Detection

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, και εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Α. Βουλόδημο (τηλ. 210-7722280, e-mail: thanosv at mail.ntua.gr), την Π. Τζούβελη (e-mail: tpar at image.ece.ntua.gr), τον Β. Καραμπίνη (e-mail: vkarampinis at ails.ece.ntua.gr) και τον Η. Μήτσουρα (e-mail: iliasmits at ails.ece.ntua.gr).