PhD projects in the Department of Informatics, AY 25-26 — Natural Language Processing

The PhD projects listed below will be considered for 2025/26 studentships available in the Department of Informatics to start on 1 October 2025 or later during the 2025/26 academic year.

Please note that this list is not exhaustive and potential applicants can alternatively identify and contact appropriate supervisors outlining their background and research interests or proposing their own project ideas.

Each project is designated for a single student, meaning it can only be assigned to one successful applicant. Some projects come with allocated studentships, while others are eligible for "unallocated" studentships. Applicants who apply for projects with allocated studentships and are selected will be offered a full studentship. In the project list, these are marked as "studentship allocated." Applicants chosen for other projects will compete for the unallocated studentships.

We welcome applications from students who have secured, or are applying for, or plan to apply for other funding (within other studentships internal to the university or external schemes) and from self-funded students. See also this list of funding opportunities available at King's for post-graduate research in Computer Science.

PhD projects

- AI and NLP for Multilingual Code-Switching in Education (studentship allocated)
- Leveraging Generative AI for Creativity Education (studentship allocated)
- Character-Centric Systems for Multimodal Story Generation (studentship allocated)
- Towards Robust Reasoning of Large Language Models (studentship allocated)
- Exploring the Diversity of Data Storytelling: A Comparative Analysis Across Cultures, Disciplines, and Media
- Understanding the Complexity of Negotiations
- Exploring Interactive Multi-Dimensional Approaches of Delivery of Communication in Patient Scenarios in Oral Health Education
- Multi-agent Cooperation with RL and LLMs
- Investigating LLM-based Generative AI Applications in Cybersecurity
- Argument mining
- Multilingual argument mining
- Agents powered by foundation models

AI and NLP for Multilingual Code-Switching in Education

Supervisor: Zheng Yuan

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Machine learning / Deep learning, Natural Language Processing, Human-centred computing (human-computer interaction)

Project Description

The rapid growth of multilingualism has led to an increased prevalence of code-switching (CSW) -- the practice of alternating between two or more languages within a single conversation or utterance. Despite its common usage in multilingual communication, current Natural Language Processing (NLP) technologies struggle to handle CSW effectively, particularly in educational contexts. This project aims to address this gap by developing advanced NLP technologies and educational AI systems specifically designed to support multilingual CSW environments. The goal is to create a personalised, inclusive, and engaging AI-powered tutoring system that adapts to the unique linguistic needs of learners. This project will focus on one or more of the following key areas: 1) Development of NLP models that can accurately process and analyse CSW data, distinguishing code-switching from grammatical errors; 2) Creation of an Intelligent Tutoring System (ITS) that provides personalised feedback and assessment tailored to the needs of multilingual learners; 3) Leveraging multilingual Large Language Models (LLMs) to enhance the capabilities of AI in educational settings, particularly in low-resource languages; and 4) Evaluating the impact of educational AI systems in real-world settings, assessing improvements in learning outcomes, learner engagement, and satisfaction.

References

- LLM-based Code-Switched Text Generation for Grammatical Error Correction. Tom Potter and Zheng Yuan. EMNLP 2024.
- Prompting open-source and commercial language models for grammatical error correction of English learner text. Christopher Davis, Andrew Caines, Øistein E. Andersen, Shiva Taslimipoor, Helen Yannakoudakis, Zheng Yuan, Christopher Bryant, Marek Rei and Paula Buttery. ACL 2024 Findings.
- Grammatical Error Correction for Code-Switched Sentences by Learners of English. Kelvin Chan, Christopher Bryant, Li Nguyen, Andrew Caines and Zheng Yuan. LREC-COLING 2024.
- Grammatical Error Correction. Christopher Bryant, Zheng Yuan, Muhammad Reza Qorib, Hannan Cao, Hwee Tou Ng and Ted Briscoe. Computational Linguistics; https://doi.org/10.1162/coli_a_00478
- Building Educational Technologies for Code-Switching: Current Practices, Difficulties and Future Directions. Li Nguyen, Zheng Yuan and Graham Seed. Languages; https://doi.org/10.3390/languages7030220

Leveraging Generative AI for Creativity Education

Supervisor: Zheng Yuan

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Computer vision, Machine learning / Deep learning, Human-centred computing (human-computer interaction), Natural Language Processing

Project Description

Creativity is a crucial skill in today's world, driving innovation, problem-solving, and cultural expression. However, teaching and assessing creativity -- especially in fields like creative writing and visual arts -- pose significant challenges due to the subjective nature of creative outputs. The rise of Large Language Models (LLMs) and Generative AI provides new opportunities for enhancing creativity education by generating personalised, adaptive feedback and supporting learners in improving their creative skills across multiple modalities, such as writing and drawing. This project will explore the use of multimodal LLMs and Generative AI to enhance creativity education, offering new ways to assess creativity and helping learners across disciplines such as creative writing and digital arts. By leveraging the capabilities of multimodal models, this research will investigate how AI can support, nurture, and assess creativity in a personalised and scalable manner. Research Questions: 1) How can LLMs and Generative AI effectively assess creativity in different forms, such as written stories, poems, or drawings? 2) What are the most effective ways for AI systems to provide feedback that nurtures creativity, without stifling originality? 3) How can multimodal AI systems enhance cross-disciplinary creative education (e.g. combining writing and drawing) to create richer, more engaging learning experiences? 4) What metrics and frameworks can be developed to evaluate the success of AI-generated feedback and creativity assessment systems?

Character-Centric Systems for Multimodal Story Generation

Supervisor: Lin Gui/Yulan He

Areas: Machine learning / Deep learning, Artificial Intelligence (symbolic AI, logic, etc.), Natural Language Processing

Project Description

The primary goal of this project is to design and develop a character-centric multimodal system capable of generating rich, coherent narratives from multimodal inputs. This system will focus on story generation based on different types of data—such as images or audio—and could be applied in a variety of settings, including museums, medical diagnostics, or educational explanations. Specifically, the system would be able to generate detailed descriptions, historical accounts, or explanations from a given image or set of multimodal data. The research problem consists of several interconnected challenges that need to be addressed to achieve the goal: multimodal input interpretation, text generation based on input data, character-centric storytelling, and cross-domain adaptability. By focusing on a character-driven approach and cross-domain adaptability, the proposed system will not only engage users but also deliver accurate, contextually relevant content based on diverse input types. This system holds great potential for enhancing user experience in numerous real-world applications, driving innovation in AI-based storytelling and explanation systems.

References

1. Silin Gao, Beatriz Borges, Soyoung Oh, Deniz Bayazit, Saya Kanno, Hiromi Wakaki, Yuki Mitsufuji, Antoine Bosselut: PeaCoK: Persona Commonsense Knowledge for Consistent and Engaging Narratives. ACL (1) 2023: 6569-6591.

2. Melanie Sclar, Sachin Kumar, Peter West, Alane Suhr, Yejin Choi, Yulia Tsvetkov: Minding Language Models' (Lack of) Theory of Mind: A Plug-and-Play Multi-Character Belief Tracker. ACL (1) 2023: 13960-13980.

3. Jingkang Yang, Yuhao Dong, Shuai Liu, Bo Li, Ziyue Wang, Chencheng Jiang, Haoran Tan, Jiamu Kang, Yuanhan Zhang, Kaiyang Zhou, Ziwei Liu: Octopus: Embodied VisionLanguage Programmer from Environmental Feedback. CoRR abs/2310.08588 (2023).

4. Yujie Wang, Hu Zhang, Jiye Liang, Ru Li: Dynamic Heterogeneous-Graph Reasoning with Language Models and Knowledge Representation Learning for Commonsense Question Answering. ACL (1) 2023: 14048-14063.

5. Runcong Zhao, Wenjia Zhang, Jiazheng Li, Lixing Zhu, Yanran Li, Yulan He, Lin Gui: NarrativePlay: Interactive Narrative Understanding. CoRR abs/2310.01459 (2023).

6. Shenzhi Wang, Chang Liu, Zilong Zheng, Siyuan Qi, Shuo Chen, Qisen Yang, Andrew Zhao, Chaofei Wang, Shiji Song, Gao Huang: Avalon's Game of Thoughts: Battle Against Deception through Recursive Contemplation. CoRR abs/2310.01320 (2023).

7. Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, Yang Liu: Exploring Large Language Models for Communication Games: An Empirical Study on Werewolf. CoRR abs/2309.04658 (2023).

8. Lixing Zhu, Runcong Zhao, Lin Gui, Yulan He: Are NLP Models Good at Tracing Thoughts: An Overview of Narrative Understanding. CoRR abs/2310.18783 (2023).

Towards Robust Reasoning of Large Language Models

Supervisor: Yulan He

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Machine learning / Deep learning, Natural Language Processing

Project Description

Context Reasoning is a core aspect of human intelligence, essential for tasks such as critical thinking, evaluation and making decisions. With the advancements of large language models (LLMs), we have witnessed their impressive performance in various natural language processing tasks that require reasoning. For an intelligent system to be effective, it must thoroughly analyse key information within a given context and provide accurate responses by leveraging its internal knowledge and external resources. This is a complex process as LLMs need to stay current with new information, remain robust in noisy contexts, and be capable of utilising external tools for validation when necessary.

Project: Despite advancements in the reasoning capabilities of LLMs, there remains uncertainty regarding the extent to which LLMs can reason beyond memorisation. Recent empirical studies have highlighted their susceptibility to challenges posed by noisy contexts, new information, and novel tasks. Therefore, our goal is to create a robust reasoning framework that enables LLMs to reason effectively when presented with new and unfamiliar inputs. To achieve this, example tasks include:

- Enhancing reasoning through tool augmentation based on a neuro-symbolic approach. LLMs can improve their reasoning by leveraging neuro-symbolic methods with the help of external interpreters, particularly in more complex tasks.
- Facilitating model adaptation to reason with the most recent knowledge. This involves model editing and fine-tuning LLMs with new information while ensuring they retain their reasoning abilities for previously encountered tasks.
- Encouraging collaboration among multiple LLM agents to support reasoning across diverse domains. When faced with an input from an unfamiliar domain, integrating knowledge from multiple trained LLM agents based on its relevance to the specific input could enhance reasoning performance.

References

- Mirzadeh I, Alizadeh K, Shahrokhi H, Tuzel O, Bengio S, Farajtabar M. 2024. GSM-symbolic: Understanding the limitations of mathematical reasoning in large language models. [pdf]
- Jie H, Kevin Chen-Chuan C. 2023. Towards Reasoning in Large Language Models: A Survey. [pdf]
- Collin B, Haotian Y, Dan K, Jacob S. 2022. Discovering Latent Knowledge In Language Models Without Supervision. [pdf]
- Almog G, Elad V, Colin R, Noam S, Yoav K, Leshem C. 2023. Knowledge is a Region in Weight Space for Fine-tuned Language Models. [pdf]
- Luyu G, Aman M, Shuyan Z, Uri A, Pengfei L, Yiming Y, Jamie C, Graham Ng. 2023. PAL: Program-aided Language Models. [pdf]
- Marco F, Florian W, Luca Z, Alessandro A, Emanuele R, Stefano S, Bernhard S, Francesco L. 2023. Leveraging sparse and shared feature activations for disentangled representation learning. [pdf]
- Jonas P. Sebastian R. Ivan V.. Edoardo M. P∗.2023. Modular Deep Learning. [pdf]

Exploring the Diversity of Data Storytelling: A Comparative Analysis Across Cultures, Disciplines, and Media

Supervisor: Alfie Abdul-Rahman

Areas: Human-centred computing (human-computer interaction), Natural Language Processing

Project Description

Data storytelling is a rapidly evolving field that integrates data analysis, narrative techniques, and visualizations to convey complex insights. Despite its growing significance, the diversity of approaches to data storytelling across cultural contexts, disciplines, and media platforms remains largely unexplored. This PhD project aims to investigate how these variations influence data stories' effectiveness, ethics, and reception. By examining this diversity, this PhD aims to reveal the impact of cultural, disciplinary, and media-related factors on how data is interpreted and communicated. Traditionally, data storytelling has been viewed through the lens of information visualization and communication. However, recent studies have broadened the scope to include narrative techniques, audience reception, and the cultural context of both the storyteller and the audience. Most existing research focuses either on the technical aspects of data visualization or narrative elements, often neglecting cross-cultural and cross-disciplinary analysis. As media platforms diversify, understanding how these platforms mediate data stories and influence engagement becomes increasingly important. This research addresses these gaps by providing a comprehensive and comparative analysis of data storytelling. It explores how the cultural and disciplinary backgrounds of both creators and audiences and the media platforms used to shape the form and impact of data stories.

Understanding the Complexity of Negotiations

Supervisor: Alfie Abdul-Rahman

Areas: Natural Language Processing, Human-centred computing (human-computer interaction)

Project Description

A negotiated text is the product of a formal decision-making process where a text has been negotiated and drafted over a period of time. Many of the foundational texts of the modern world have not been written by individuals, by negotiated by groups of people in formal settings. For example, treaties between states such as the Universal Declaration of Human Rights or the Treaty of Versailles; or constitutions, such as the one negotiated by the American states in the Constitutional Convention of 1787. During such negotiations, it is important for us to keep track of the delegations and their involvements in order to grasp their influence on the negotiation process either using techniques such as close reading, distance reading, or machine learning. Even relatively short historical documents written collectively in this way have been the product of thousands of specific proposals and decisions. This project will apply a visual analytics approach towards the understanding of the complexity of a negotiation and the influence of the delegations during a negotiation process. Possible research questions: - Developing new static and interactive visualization to assist with data discovery and insight generation in large datasets of events within interacting timelines. - Developing new approaches to show the evolution of complicated, technical documents over the period of months or years. - Developing new approaches for indexing the datasets related to the negotiation of documents, and more intuitive displays of the results. - Developing natural-language-based approaches to relate information captured in 'informal' archives (such as private diaries, letters, social media feeds etc.) to the formal records of a negotiation.

Exploring Interactive Multi-Dimensional Approaches of Delivery of Communication in Patient Scenarios in Oral Health Education

Supervisor: Informatics: Dr Alfie Abdul-Rahman & Dr Lin Gui FoDOCS: Dr Melanie Nasseripour & Dr Ana Angelova

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Natural Language Processing, Human-centred computing (human-computer interaction), Machine learning / Deep learning, Education

Project Description

This is a joint project between the Department of Informatics and the Faculty of Dentistry, Oral & Craniofacial Sciences (FoDOCS). Communication in patient scenarios in oral health education can be cost-intensive in terms of time and resources. In this project, we propose exploring interactive multidimensional approaches such as immersive technology, text-to-text, and voice-to-voice communication delivery in patient scenarios in oral health education. These approaches enhance the learning experience and offer a cost-effective solution, making the delivery of communications in patient scenarios in oral health education more feasible and sustainable. This project aims to design and create adaptable, contextually relevant patient scenarios, offering engaging and realistic interactions for students. The beauty of these approaches is their adaptability. Whether it is a VR interactive tool, text-based, or voice-based conversation, they can all respond to students' inquiries and actions in real-time, mimicking the interaction they would normally have in the clinic. This adaptability ensures the relevance and effectiveness of the project in various educational settings. We aim to examine the Generative Language Models (GLMs) to generate customized case studies and simulation scenarios so that each learner can practice specific skills repeatedly in a controlled environment. This encourages the acquisition and refinement of skills, such as explaining the importance of oral hygiene and discussing dietary habits. The critical focus is patient-clinician communication, behaviour change, professionalism, etc.

Multi-agent Cooperation with RL and LLMs

Supervisor: Yali Du

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Machine learning / Deep learning, Natural Language Processing, Robotics

Project Description

Multi-agent systems (MAS) have become increasingly relevant in fields such as robotics, finance, and autonomous systems. However, achieving effective cooperation among multiple agents remains challenging, especially in dynamic and uncertain environments. RL has been a powerful method for training agents, but traditional approaches often struggle with scalability and communication bottlenecks. Meanwhile, LLMs have demonstrated remarkable capabilities in language understanding and generation, which can be leveraged to facilitate communication and strategy development among agents. This study aims to explore how reinforcement learning (RL) can be combined with large language models (LLMs) to improve multi-agent cooperation in complex environments. The goal is to enhance communication, decision-making, and coordination between agents, enabling them to solve tasks that require a high level of collaboration and safety. This project explores the questions of 1) How can LLMs be integrated into multi-agent systems to enhance cooperation and communication among agents trained using RL? 2) What are the optimal communication protocols that maximize the synergy between LLMs and RL in multi-agent scenarios? 3) How can this combination be scaled to large numbers of agents while maintaining efficiency and performance? Dr Du's early attempts explored how to leverage LLMs for communication, and incorporated human instructions to ensure safe and cooperative control, with examples including the game of Werewolf, football, and safe robot control. This research will contribute to the field of multi-agent systems by developing new techniques for improved cooperation using cutting-edge LLMs. The findings could be applicable in various industries, including autonomous vehicles, robotics, and distributed AI systems, where multi-agent cooperation is critical for success.

References

[1] Safe Multi-agent Reinforcement Learning with Natural Language Constraints. Ziyan Wang, Meng Fang, Tristan Tomilin, Fei Fang, Yali Du. Arxiv 2024.
[2] Understanding, Rehearsing, and Introspecting: Learn a Policy from Textual Tutorial Books in Football Games. Xiong-Hui Chen, Ziyan Wang, Yali Du, Meng Fang, Shengyi Jiang, Yang Yu, Jun Wang. NeurIPS 2024 Oral.
[3] Learning to Discuss Strategically: A Case Study on One Night Ultimate Werewolf. Xuanfa Jin, Ziyan Wang, Yali Du, Meng Fang, Haifeng Zhang, Jun Wang. NeurIPS 2024.

Investigating LLM-based Generative AI Applications in Cybersecurity

Supervisor: Ievgeniia Kuzminykh/Hannan Xiao

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Cybersecurity, Natural Language Processing, Human-centred computing (human-computer interaction)

Project Description

The rapid development and deployment of GPT-based agents in cybersecurity mark a significant leap forward in approaching digital security challenges from a practitioner standpoint. Under this topic you can explore the ways generative AI is impacting the cybersecurity industry, from both sides, such as gen AI for security and security of gen AI. From one side, malicious attackers are seizing the potential of generative AI to launch cyber attacks that are harder to detect and defend against. OWASP top 10 for Generative AI [1] lists out the top 10 vulnerabilities impacting the applications usilising LLM. Prompt Injection, Insecure Output Handling and Data Poisoning take the top 3 spots and are also the root causes for the other type of vulnerabilities (Overreliance, Insecure Plugins etc) as shown in research paper [2]. From another side, Gen AI is also helping make security teams more accurate, efficient, and productive in defending their organisations. Examples of utilising of generative AI for security operations could be [3] : - Supplementing understaffed security teams - Detecting threats in real time - Improving incident response. The potential topics in this project area could include but not limited to: 1. Optimisation of prompts for security related topics. Through clever prompt engineering (called jailbreaking [4]), LLMs can be made to reveal internal mechanisms, share private data, produce offensive speech, or perform unintended workloads. LLMs thus pose a security risks [1, 5, 6]. 2. Prompt injection detection mechanisms. 3. Ensuring online safety using LLM. AI seems like the perfect response to the growing challenges of content moderation on social media platforms: the immense scale of the data, the relentlessness of the violations, and the need for human judgments without wanting humans to have to make them. The paper [7] elaborates on the topic of prompt/response classifiers. The prompts and answers could be classified into groups such as safe and harmful. Typical examples of a harm would be Child Safety, Exfiltrating PII/SPII, Sexually Explicit Content, Malicious/Dangerous content. 4. Content moderation using LLM. Similar to previous but can be extended to the detection of harassment and throlling [8]. 5. Understanding Generative AI for Cloud Security. Generative AI can make new data from existing patterns. For cloud security, this means it has the potential to: 6a. Simulate Threat Scenarios: Generative AI can create realistic threat scenarios, allowing security teams to test and validate their Cloud infrastructure's resilience. By simulating potential attack vectors, organizations can proactively identify vulnerabilities and take steps to ensure they are protected against them before they are exploited. 6b. Optimize Security Configurations: AWS offers a number of services, each with its own set of security configurations. With Generative AI, we can analyze existing configurations, simulate various combinations, and ask Generative AI to provide recommendations based on our specific needs. 6c. Enhance Monitoring and Alerts: By training on historical security logs and events, Generative AI can predict potential security breaches or anomalies. The key word here is "potential." Knowing what "could" happen allows security teams time to prepare and allows for more rapid action to be taken. 6. Understanding Generative AI for firewall optimisation Generative AI could simulate web traffic patterns based on your historical log data and compare that to your existing WAF or firewall rules, ensuring that malicious requests are blocked while legitimate traffic flows seamlessly. 7. 10. Understanding of Gen AI for Qualitative audit of security policies Each organisation is governed by a security policy, which technically or conceptually specifies a number of guidelines for ensuring IT security. You will investigate whether GenAI can be employed to translate a security policy for wider staff [9, 10]..

References

[1] https://owasp.org/www-project-top-10-for-large-language-model-applications/
[2] Knowledge Bases and Language Models: Complementing Forces. https://link.springer.com/chapter/10.1007/978-3-031-45072-3_1
[3] https://secureframe.com/blog/generative-ai-cybersecurity
[4] Liu, Y., et al.: Jailbreaking ChatGPT via prompt engineering: an empirical study. arXiv preprint arXiv:2305.13860 (2023)
[5] https://simonwillison.net/2023/May/2/prompt-injection-explained/
[6] https://www.geeksforgeeks.org/what-is-jailbreak-chat/
[7] Llama Guard: LLM-based Input-Output Safeguard for Human-AI Conversations https://arxiv.org/html/2312.06674v1
[8] Content moderation, AI, and the question of scale https://journals.sagepub.com/doi/full/10.1177/2053951720943234
[9] Introducing Rules Genie: Generative AI for Automating Policy Creationhttps://www.opsmx.com/blog/introducing-rules-genie-generative-ai-for-automating-policy-creation/
[10] The LLM Police: Between Firewalls and Policies for Generative AI https://www.fairly.ai/blog/the-llm-police-between-firewalls-and-policies-for-generative-ai

Argument mining

Supervisor: Oana Cocarascu

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Machine learning / Deep learning, Natural Language Processing

Project Description

In everyday life, decisions are often based on arguments, counter-arguments, and facts. While arguments are claims backed by reasons that are supported by evidence, facts can be proven with clear and objective data. Automatically identifying and presenting facts and arguments can not only facilitate and challenge debates, but also aid humans and automated systems in reaching decisions, hence the societal impact of this task is tremendous.
Computational argumentation is a research area in natural language processing which encompasses several tasks such as argument mining, argument reasoning, and argument generation amongst others. Much progress has been made in recent years on argument mining whereby the task is to determine whether a text represents an argument, followed by identifying the arguments for or against an issue. Argument mining has been applied to several areas: persuasive essays, scientific articles, Wikipedia articles, news articles, online debates, product reviews, social media, legal documents, and political debates.
The project aims to develop computational methods that find, extract, and evaluate arguments in text as well as deal with incomplete arguments, i.e. arguments that can be understood using background knowledge.

References

[1] https://aclanthology.org/2022.tacl-1.80.pdf
[2] https://aclanthology.org/2024.acl-long.126.pdf

Multilingual argument mining

Supervisor: Oana Cocarascu

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Machine learning / Deep learning, Natural Language Processing

Project Description

In everyday life, decisions are often based on arguments, counter-arguments, and facts. While arguments are claims backed by reasons that are supported by evidence, facts can be proven with clear and objective data. Automatically identifying and presenting facts and arguments can not only facilitate and challenge debates, but also aid humans and automated systems in reaching decisions, hence the societal impact of this task is tremendous.
Computational argumentation is a research area in natural language processing which encompasses several tasks such as argument mining, argument reasoning, and argument generation amongst others. Much progress has been made in recent years on argument mining whereby the task is to determine whether a text represents an argument, followed by identifying the arguments for or against an issue. Argument mining has been applied to several areas: persuasive essays, scientific articles, Wikipedia articles, news articles, online debates, product reviews, social media, legal documents, and political debates.
Despite the growing interest in computational argumentation, the majority of datasets are in English. The project will focus on argument mining in low-resource languages and will develop novel corpora and algorithms for multilingual argument mining.

References

[1] https://aclanthology.org/2022.tacl-1.80.pdf
[2] https://aclanthology.org/2024.acl-long.126.pdf
[3] https://aclanthology.org/2024.acl-long.628.pdf
[4] https://aclanthology.org/2020.findings-emnlp.29.pdf

Agents powered by foundation models

Supervisor: Helen Yannakoudakis

Areas: Machine learning / Deep learning, Artificial Intelligence (symbolic AI, logic, etc.), Natural Language Processing

Project Description

With the expansive capabilities of foundation models, the concept of building agents powered by these models (like large language models) has recently emerged. Several demonstration projects, such as AutoGPT, GPT-Engineer, and BabyAGI, illustrate this potential. Foundation models offer possibilities beyond creating images, well-crafted text, stories, essays, and code—they can serve as powerful general problem solvers. This project aims to develop agents driven by foundation models that can observe, take action, and respond to feedback in a continuous loop with external environments, including interactions with humans, tools, and the physical world. The focus will be on two key areas: specialization and multi-modality.