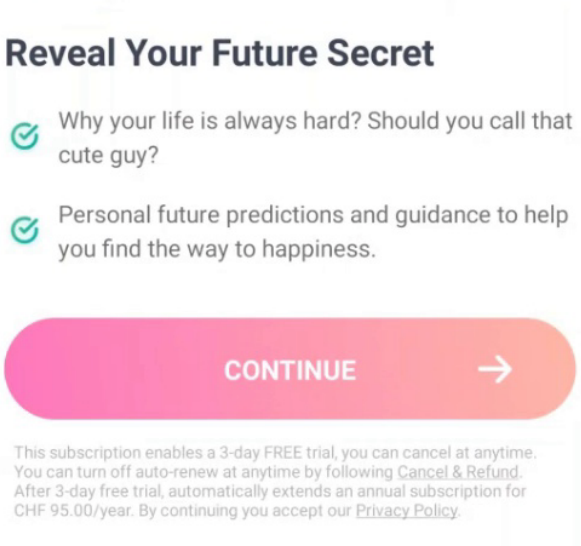
(Bsc/Msc) Enhancing Dark Pattern Detection through Textual Analysis and Contextual Understanding in User Interfaces



Objective: The thesis aims to enhance the detection of dark patterns in user interfaces by incorporating advanced textual analysis and contextual understanding, addressing gaps in current detection methods that primarily focus on visual elements.

How: The research will develop a hybrid model that integrates natural language processing (NLP) techniques with existing visual recognition methods. This model will analyze the language used in UI elements, such as misleading text and ambiguous prompts, while also considering the broader context of user interactions over time.

Outputs:

- (for bachelor thesis): focus can be on building a new dataset/system for a new domain, e.g. social media, games by extending existing ones
- (for master thesis): a comprehensive detection system capable of identifying dark patterns that rely on textual manipulation and contextual cues
- guidelines for designers and regulators to mitigate dark patterns

References:

- Mansur, SM Hasan, et al. "Aidui: Toward automated recognition of dark patterns in user interfaces." *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023 (github repo)
- Gray, Colin M., et al. "An Ontology of Dark Patterns: Foundations, Definitions, and a Structure for Transdisciplinary Action." (2024).
- Sameen, Maria, and Awais Rashid. "Dark Patterns in Video Games: An Exploratory Study." (2024).
- Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23). Association for Computing Machinery, New York, NY, USA, 2362–2374.

(BSc/MSc) Developing a Case-Informed Chatbot for EU261 Air Passenger Rights

Objective:
Develop a chatbot that assists air passengers with their rights under the EU261 regulation, focusing on recent case law to address legal complexities based on the individual cases.

How:

1. **User Interaction**: The chatbot will engage with passengers by asking a series of questions to gather relevant details about their flight issues (e.g., delays, cancellations).
2. **Case Law Integration**: The chatbot will reference recent legal cases and rulings to ensure that its responses account for current interpretations of EU261, particularly in areas with ambiguities (e.g., what counts as "arrival time").
3. **Information Retrieval**: Leveraging a specialized database of EU261 regulations and case law, the chatbot will provide accurate and contextually relevant information to passengers. Different possibilities will be explored in terms of generating answers or limiting the answers to the facts of each case.
4. **Slot filing and Query Decomposition**: The system will operate on the basis of intent identification (e.g. what are my rights? Am I entitled to compensation?) and slot filling (e.g. flight info, dates, etc.) The system will break down complex inquiries into simpler sub-questions to ensure a thorough analysis of the situation

Outputs:

- (for bachelor thesis): based on the cases parsed, focus will be on building a testing dataset using genAI techniques about realistic cases (e.g. generate such conversations) and test them with a basic chatbot design
- (for master thesis): explore chatbot functionality with the end product being an eligibility report with clear legal references and perhaps actionable advice.
- A prototype for the chatbot

References

- EU261 cases (to parse): https://flightdelaypay.com/leading-cases/
- Taranukhin et al. “Empowering Air Travelers: A Chatbot for Canadian Air Passenger Rights” (2024, Pre-print)

(BSc/MSc) Detecting Greenwashing in Corporate Communications and Social Media

Objective:
The aim of this research is to develop an advanced Natural Language Processing (NLP) framework to detect and quantify greenwashing in corporate communications by comparing corporate sustainability reports with external data sources, including news media, NGO reports, and social media content.
One focus area could be to assess greenwashing practices in the airline industry, particularly focusing on European airlines' voluntary carbon offset (VCO) programs. The study aims to evaluate the authenticity and alignment of environmental claims made by airlines, comparing them against regulatory benchmarks like the EU Greenwashing Directive.

How:

- Start by looking at existing datasets and expand if needed
- Develop algorithms to detect inconsistencies between a company's internal reports and external perceptions, particularly focusing on areas where social media reveals a different narrative. For tis, literature will be used (peer-relative greenwashing score, greenwashing severity index, etc.) and standard methodologies (e.g. ICAO carbon emissions calculators).

Outputs:

- Comparison of the detected practices with regulatory standards to determine the extent to which airlines' environmental claims align with legal and ethical guidelines.
- Analysis of differences across airlines and identification of patterns in greenwashing strategies

References

- Sharma, U., Rudinac, S., Demmers, J., van Dolen, W., Worring, M. (2024). GreenScreen: A Multimodal Dataset for Detecting Corporate Greenwashing in the Wild. In: Rudinac, S., *et al.* MultiMedia Modeling. MMM 2024. Lecture Notes in Computer Science, vol 14565. (data repo)
- Rosario, Macario. "Assessing the efficacy of EU greenwashing directive: A study of European airlines' voluntary carbon offset programs." *Journal of the Air Transport Research Society* 3 (2024): 100028.
- Yu, Ellen Pei-yi, Bac Van Luu, and Catherine Huirong Chen. "Greenwashing in environmental, social and governance disclosures." *Research in International Business and Finance* 52 (2020): 101192.

(BSc/MSc) Analyzing Queer Linguistic Lexicons in Social Media: Cross-Language Usage and LLM insights

Objective:
This research aims to explore the use of queer linguistic lexicons across different languages (e.g. Lubunca in Turkish, Kaliarnta in Greek etc.) on social media and evaluate the ability of large language models (LLMs) to understand and accurately respond to these lexicons. The study will utilize established benchmarks, such as VinoQueer, to systematically assess model performance.

How:

- Data collection: Compiling multi-lingual datasets that feature queer-specific lexicons
- Use benchmarks like VinoQueer to analyze and categorize queer lexicons by language, context, and cultural significance.
- Evaluate LLMs against queer linguistic benchmarks, such as VinoQueer, to measure their ability to recognize, interpret, and accurately respond to queer-specific lexicons across languages

Outputs:

- A detailed analysis of queer linguistic lexicons' usage across languages on social media
- A performance evaluation of prominent LLMs using the developed benchmarks, highlighting bias and errors when it comes to queer lexicons.
- An assessment of cross-linguistic differences in LLM performance, with recommendations for addressing any identified biases.
- Practical guidelines and strategies for improving LLM training processes to better incorporate and understand queer lexicons, contributing to more inclusive AI technologies.

References:

- (article) How does Queer slang differ around the world?
- (workshops) Queer in AI workshops (1,2), C3NLP
- Anne Lauscher, Archie Crowley, and Dirk Hovy. 2022. Welcome to the Modern World of Pronouns: Identity-Inclusive Natural Language Processing beyond Gender. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1221–1232, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Robust Pronoun Fidelity with English LLMs- Are they Reasoning, Repeating, or Just Biased? (preprint)
- Dhingra, Harnoor, et al. "Queer people are people first: Deconstructing sexual identity stereotypes in large language models." *arXiv preprint arXiv:2307.00101* (2023).
- Virginia Felkner, Ho-Chun Herbert Chang, Eugene Jang, and Jonathan May. 2023. WinoQueer: A Community-in-the-Loop Benchmark for Anti-LGBTQ+ Bias in Large Language Models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 9126–9140, Toronto, Canada. Association for Computational Linguistics

(BSc/MSc) Social media strategies of airlines: Content and ads

Objective:
Apply NLP techniques to analyze and enhance the effectiveness of social media strategies used by airlines, focusing on the content and tone of ads and posts.

How:

- Construction of the dataset: An existing dataset of facebook posts exists for different airlines across the globe. we will enhance this with the Meta Ad libraryLinks to an external site. to have a bigger longitudinal dataset.
- Using NLP models identifying the tone/language used in media as well as the type of content (informational, entertainment, promotional, advertising, etc.).

Output:

Insights on what type of content is posted by airlines in social media content and ads, more specifically:

- frequency, timing, context of phrases or other signals that might correlate with high engagemnet
- linguistic patterns and content strategies and how they evolve over time

References:

- Prados-Peña MB, Crespo-Almendros E, Porcu L. COVID-19 and social media communication strategies: A comparative study of the effectiveness of Facebook posts during the lockdown and the “new normal” in the airline industry. Journal of Air Transport Management. 2022 Aug 1;103:102255.
- Heiets, Iryna, et al. "Social media activities of airlines: What makes them successful?." *Journal of the Air Transport Research Society* 2 (2024): 100017.
- http://viveksck.github.io/langchangetrack/

(MSc) Comparative Automated Analysis and Summarization of Terms and Conditions Across Digital Platforms

Objective:
The Digital Services Terms and Conditions Database is a resource that tracks and compiles the terms and conditions, privacy policies, and other legal agreements of digital service providers (e.g. platforms like Amazon, Meta, Google).

Our goal istTo develop an NLP-based system that automates the extraction, categorization, and comparative analysis of key elements from the terms and conditions of digital platforms. The system aims to enhance transparency, assess compliance with the P2B Regulation, and identify commonalities and differences across platforms.
How:

- Data collection: The EU Platform Contracts Database is available for downloading. Since we are the first ones to work on this, a good exploratory analysis is necessary. If also needed, the annotation of data needs to be considered
- Named Entity Recognition (NER) and Text summarization/simplification techniques to parse and condense the terms and conditions of each platform.
- Classification of critical clauses related to user rights, obligations, and compliance issues (for this, annotated data might be needed)

Outputs:

- Summarization/Simplification System: A system capable of generating concise summaries of terms and conditions from digital platforms, emphasizing key points and potential legal implications.
- Comparative Analysis Reports: Detailed reports that compare terms of service across different platforms, highlighting commonalities, differences, and potential risks
- (optionally) a dashboard or interface that displays summarized content, flags potential compliance concerns, and allows for easy comparison of terms across platforms.

References

- Galassi, Andrea, et al. "Unfair clause detection in terms of service across multiple languages." *Artificial Intelligence and Law* (2024): 1-49.
- Sweta Agrawal and Marine Carpuat. 2024. Do Text Simplification Systems Preserve Meaning? A Human Evaluation via Reading Comprehension. *Transactions of the Association for Computational Linguistics*, 12:432–448.
- The Force Awakens: Artificial Intelligence for Consumer Law
- (links) http://claudette.eui.eu/index.html, https://tosdr.org/

(MSc) From Scrolls to Scripts: Analyzing the Corpus Juris Civilis using NLP

Objective:
The Corpus Juris Civilis (translated: Body of Civil Law) is a foundational collection of legal texts compiled and codified under the direction of the Byzantine Emperor Justinian I in the 6th century CE. It is one of the most important works in the history of law, serving as a basis for civil law systems in many parts of the world, particularly in Europe.
The goals is to develop and apply NLP techniques to analyze, interpret, and categorize the texts of the Corpus Juris Civilis, addressing both legal and computational questions related to the understanding, translation, and application of ancient Roman law in modern contexts.

How:

- Data collection: There are different version of Corpus Juris Civilis (in the original version) which we will use along translated versions.
- NLP techniques to use (depending on the actual task)
    - Text Classification and Categorization: Apply text classification techniques to categorize different sections of the Corpus Juris Civilis.
    - Named Entity Recognition (NER): Reuse/develop NER models to identify legal entities, concepts, and relationships within the texts.
    - Translation with semantic analysis: Use translation models combined with semantic analysis to intrepret roman legal concepts

Outputs:

- Textual Analysis and Structure
  - How is the structure of legal *argumentation* in the Corpus Juris Civilis organized? (topic modeling /clustering)
  -What are the most common linguistic structures used in legal language within the Corpus Juris Civilis? (syntactic and semantic patterns that are prevalent in the text) and how they are related/connected?
- Authorship Attribution: Can we identify different authors or sources within the Corpus Juris Civilis based on linguistic patterns?
- Comparative Legal Analysis: How does the legal language in the Corpus Juris Civilis compare to other historical legal texts?
  Comparison of legal concepts in the Corpus Juris Civilis with modern legal systems

References

Dingledy, Frederick W. "The corpus juris civilis: a guide to its history and use." *Legal Reference Services Quarterly* 35.4 (2016): 231-255.

Corpus:
https://droitromain.univ-grenoble-alpes.fr/corpjurciv.htm
https://constitution.famguardian.org/2-Authors/sps/sps.htm
https://archive.org/details/institutesofjust0000unse_j1t4/page/6/mode/2up
https://www.perseus.tufts.edu/hopper/collection?collection=Perseus:collection:Greco-Roman
ML/NLP for ancient languages (resources and ideas for tasks)
https://www.ancientnlp.com/alp2023/
https://www.ml4al.com/