EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Wirtschafts- und Sozialwissenschaftliche Fakultät

LEHRSTUHL FÜR MARKETING
Prof. Dr. Dominik Papies

Telefon +49 7071 29-76977
Telefax +49 7071 29-5078
dominik.papies@uni-tuebingen.de
www.uni-tuebingen.de/wiwi/marketing

Universität Tübingen · LS für Marketing · Nauklerstr. 47 · 72074 Tübingen

Seminar "New developments at the intersection of Machine Learning, Causal Inference, and Marketing"

I. Type of seminar

In this seminar, students will work on selected topics that involve modern tools for data analysis, e.g., from the domain of Machine Learning or Causal Inference, or at the intersection of these two.

The topics can be chosen either from the list of suggested topics, or students propose their own topics. In the latter case, the suitability of the topic will be discussed with the supervisors.

In this seminar, students will also acquire relevant tools to be prepared for writing a research-based master thesis. This will be supported by an obligatory workshop on academic research as well as an obligatory workshop on presentation skills, which includes a short presentation of each student's current state of the thesis ("research plan presentation"). On top of that, we expect and encourage active participation and interaction between students.

It is expected that students have **at least very solid skills in statistical software (preferably R or Python)**, equivalent to, e.g., a successful completion of DS400 Data Science Project Management and/or DS404 Data Science with Python. In addition, we expect that students are willing to **familiarize themselves** with new methods and approaches as well as new tools in R or Python. The respective supervisor will support students in this.

II. Topics and introductory reading material

Topic 1	**Large Scale Analysis of Nudging Practices in Online Cookie Consent Management**
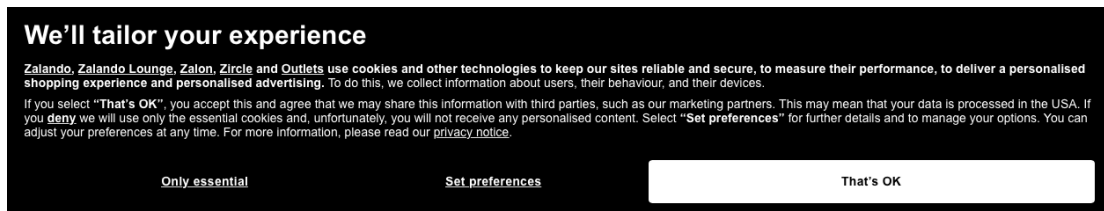
	Cookie consent management systems for several websites could be criticized for not letting the user make a free choice about which cookies to accept. This nudging practice of websites is often manifested in a highlighted "OK" button compared to opting out in smaller underemphasized font or hidden under a second layer of settings. The goal of this thesis is to automatically infer such nudging practices from website screenshots. Students will investigate the prospects of exploiting deep learning methods to extract information relevant to cookie consent management from website screenshot images. Furthermore, students will employ this automated approach to conduct a large-scale analysis of hundreds of popular websites and infer insights about patterns in nudging behavior of online cookie management systems. *Do e-commerce websites employ more aggressive nudging policy compared to news websites? Are websites of businesses headquartered in the EU less likely to nudge users into accepting the use of all cookies?*
Literature	T Gogar, O Hubacek, J Sedivy (2016). Deep Neural Networks for Web Page Information Extraction A Kumar, K Morabia, W Wang, K Chang, A Schwing (2022). CoVA: Context-aware Visual Attention for Webpage Information Extraction.
Data	Dataset to be developed in the project.

Topic 2	**Bad controls in Marketing**
	A common piece of advice in applied econometric analysis is to control for potential confounders, i.e., variables that may affect the focal regressor as well as the outcome. This approach may lead researchers to include a wide range of covariates. At the same time, controlling for so-called "bad control variables" may bias the results, i.e., it may be better not to control for some variables. This problem, however, is not well understood in applied econometric research and in marketing. It is therefore the goal of this project to summarize the current state of the literature on the topic of bad control variables, examine its relevance for marketing, and assess the severity of the problem through simulations.
Literature	Cinelli, C., Forney, A., & Pearl, J. (2022). A Crash Course in Good and Bad Controls. *Sociological Methods & Research*, https://doi.org/10.1177/00491241221099552 Klarmann, M., & Feurer, S. (2018). Control Variables in Marketing Research. *Marketing ZFP*, *40*(2), 26–40. https://doi.org/10.15358/0344-1369-2018-2-26
Data	Own simulations & applications with openly available data from related literature

Topic 3	**Recent Developments in Difference-in-Differences Analysis (DiD)**
	Answering causal questions is at the core of many scientific disciplines. One of the most popular methods for estimating causal effects in (quasi-)experimental research is the Difference-in-Differences (DiD) approach. As the importance of DiD continues to grow in practical applications, so does the breadth of literature on the subject. This leads to a multitude of papers exploring various facets of DiD, making it more and more difficult to keep up with rapidly evolving best practices in the field. In this thesis, students should study how researchers can leverage the various aspects of the DiD approach in different research contexts, with a focus on currently discussed challenges such as differences in treatment timing (staggered DiD), small treatment groups (synthetic DiD), and violations of the parallel pre-treatment trend assumption. They review and discuss the methodology, before applying it to simulated as well as real data from business and economics.
Literature	Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics* Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. https://doi.org/10.2307/j.ctv1c29t27 Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. CRC Press. https://doi.org/10.1201/9781003226055
Data	Own simulations & applications with openly available data from related literature

Topic 4	**Field Study: Exploring the Role of Large Language Models in Programming Education**
	The burgeoning rise of Large Language Models (LLMs) like ChatGPT has sparked interest in their potential to revolutionize learning, particularly in complex fields like programming. This field study investigates the impact of LLMs on programming acquisition by comparing two groups of students: one with unlimited access to ChatGPT and another with traditional learning resources (textbooks, online tutorials). We aim to assess whether ChatGPT-assisted learning leads to: **Improved conceptual understanding:** Deeper grasp of programming concepts, ability to apply knowledge to new problems. **Greater motivation and engagement:** More enjoyable learning experience, increased persistence and confidence. Through carefully designed tasks, assessments, and surveys, we will analyze the learning outcomes and experiences of both groups. This study offers crucial insights into the potential of LLMs as transformative tools for programming education, informing future practices and curriculum development
Literature	Dell'Acqua, Fabrizio and McFowland, Edward and Mollick, Ethan R. and Lifshitz-Assaf, Hila and Kellogg, Katherine and Rajendran, Saran and Krayer, Lisa and Candelon, François and Lakhani, Karim R., Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality (September 15, 2023). Harvard Business School Technology & Operations Mgt.

Topic 5	**Controllable Image Manipulation**

Recent advances in generative image models have enabled the synthesis of high-quality images. Controllable Attribute Manipulation refers to the ability to modify specific aspects of an image in a fine-grained manner, without changing other aspects of the image. For example, it may involve changing the color of a specific object in an image without changing the color of the entire image. While attribute manipulation methods in image editing have advanced significantly, they still come with certain limitations.

1. Limited Attribute Control: Most attribute manipulation methods focus on a specific set of attributes, such as changing the hairstyle or age of a person in an image. They may not support a wide range of attributes simultaneously, and controlling multiple attributes at once can be challenging.

2. Overfitting: Some attribute manipulation models are highly specialized and may overfit to the specific dataset they were trained on. This can lead to unrealistic or inconsistent results when applied to images outside the training distribution.

3. Lack of Fine-grained Control: Achieving fine-grained control over attributes can be challenging. Users might want to make subtle changes, but the model's adjustments can be too pronounced or not precisely aligned with the user's intentions.

4. Semantic Understanding: Many attribute manipulation methods lack a deep understanding of image semantics. For instance, changing the color of an object might not consider its real-world plausibility, leading to unnatural results.

5. Limited Resolution and Detail: Some methods struggle to preserve fine details when modifying attributes. This can result in loss of image quality or artifacts in the edited regions.

6. Dependency on Training Data: Attribute manipulation models are highly dependent on the quality and diversity of the training data. If the training data lacks certain attributes or has biases, the model may not perform well for those attributes.

7. Generalization to Complex Scenes: Attribute manipulation is often easier on simpler scenes and objects. Handling complex scenes with multiple objects and interactions can lead to errors and unrealistic results.

8. Limited Domain Transfer: Models trained on one domain (e.g., human faces) may not generalize well to other domains (e.g., fashion items). Domain-specific models are often needed for accurate attribute manipulation.

In this project, we will survey the recent advances in controllable attribute manipulation and discuss the challenges and limitations of existing methods. We will also explore the future directions for research in this area.

Data	Several publicly available datasets

- **III. Dates**

March 20, 2024	Online Application via ILIAS
April 8, 2024	9:00 a.m s.t. – 1:00 p.m. – Seminar Room 236 Neue Aula Kick-off and topic assignment Workshop „Academic Writing"
May 2, 2024	9:00 a.m. s.t. – 1:00 p.m. – Seminar Room 063 Neue Aula Workshop "Presentation Skills"
May 17, 2024	All day – Seminar Room 236 Neue Aula Research plan presentation
June 17, 2024 2023	Term paper is due by noon (12 p.m. s.t.) (You can drop your term paper in the letterbox outside the faculty (addressed to Chair of Marketing - Nauklerstr. 47) or send it by post (postmark date is relevant).) Containing 2 versions of the term paper with a filing clip (https://de.wikipedia.org/wiki/Heftstreifen) Submit the electronic version (pdf) of the term paper incl. analysis scripts as file upload in ILIAS.
July 1, 2024	All day (dates will be coordinated individually) Feedback Session
July 17, 2024	8:00 p.m. Upload Presentation in ILIAS
July 18, 2024 (tentative, subject to change)	All day Seminar – Seminar Room 225 Neue Aula

IV. Course credits

Students can obtain course credit (9 ECTS). To obtain course credit students must meet the following criteria:
- Students participate in all meetings listed above
- Students submit their 12-page thesis on time
- Students present their thesis during the seminar
- Students actively participate during the seminar

Approx. 50% of the final grade will be the thesis, and 50% of the final grade will be the presentation and the participation in the seminar.

Please note:
Topics are subject to change - Students are invited to propose their own topics that fit under the general theme of the seminar.

Tübingen, February 2024