**Exposé für eine Belegarbeit zum Thema**

**Recognition of Habits and Their Contexts Using Pretrained Language Models: Methodological Approaches and Prototypical Implementation**

|  |  |
| --- | --- |
| **Name:** | **Jingting Hua** |
| **Studiengang:** | **Wirtschaftsinformatik** |
| **Matrikelnummer:** | **4910816** |
| **Betreuer:** | **Felix Reinsch,Marcel Susky** |

# Forschungsthema und Kontext

In everyday life, especially in the domain of health, habits play a crucial role in human behavior. Habits are memory-based associations formed by repeatedly performing behaviors in specific context cues; when encountering the same context cues again, the associated behavior is automatically activated, even without conscious awareness (Wood & Neal, 2007). Traditionally, the formation and change of habits have been primarily studied in psychology. However, in practical applications, the challenge remains of how to recognize and understand individuals’ habitual behaviors from natural language. Recent advances in large language models (LLMs) suggest that these models are not only capable of generating human-like text but can also perform reasoning over complex semantic structures, including behavioral patterns (Minaee et al., 2025).

# Forschungsstand und theoretische Grundlagen

Across behaviorism, cognitive neuroscience, and social psychology, there is a shared understanding of habits: habits are formed in memory through repeated responding, establishing context–response associations and automated habit performance, which is relatively insensitive to changes in the value or contingency of behavioral outcomes (Wood & Rünger, 2016). Context or context cues are among the most important elements in studying habits, including their generation, triggering, and change. During habit formation, individuals tend to automatically repeat previously rewarded behaviors in specific contexts, thereby forming the unconscious outcome of context–response associations (Labrecque, Lee, & Wood, 2024; Wood, Mazar, & Neal, 2022; Wood & Rünger, 2016). Once a habit is formed, exposure to context cues associated with past habitual behavior can automatically activate the habitual response, without the need for conscious reflection on the underlying goal or purpose (Wood & Neal, 2007). For example, only when participants were in a context previously associated with the behavior—such as a movie theater—and ate in a way that allowed the automatic execution of context cue responses—such as eating with their hands—did they continue to eat popcorn out of habit, regardless of its freshness (Neal, Wood, Wu, & Kurlander, 2011). This vividly illustrates the direct relevance of context cues to the automatic cuing of habits.

When it comes to habit change, context also plays a critical role: when familiar context cues disappear, habitual behavior is interrupted, requiring individuals to make new decisions—thus opening a window of opportunity for new behavior formation. People can break bad habits by disrupting the associated context cues or by repeatedly acting in a new environment to form new habits (Carden & Wood, 2018). Therefore, analyzing the characteristics of context cues is crucial. However, there is still a lack of systematic classification and modeling of context cue features. This gap limits our understanding of how different contextual characteristics influence habit formation, habit triggering, and change. A necessary prerequisite for analyzing contextual cues is first to determine whether the behavior qualifies as a habit—only then is context analysis required.

Context is not limited to simple elements such as physical location, other people, internal states, or the previous action in a behavioral sequence; it also encompasses complex composite situations formed by the combination of such elements (Wood, 2017). Large language models (LLMs), with their strong semantic understanding capabilities, are well-suited for handling binary or multi-label classification problems on unstructured text (Peña et al., 2023; Tao et al., 2023; Yang & Menczer, 2025), making them suitable for the task of habit context classification. Due to their large capacity, LLMs are capable of performing a wide range of tasks in zero-shot settings (Brown et al., 2020; Radford et al., n.d.). However, this generalization ability has limitations, such as the lack of domain-specific knowledge and the uniqueness of certain tasks. Fine-tuning is thus a necessary and critical step, bridging the gap between general knowledge and domain-specific requirements (Anisuzzaman, n.d.).

In recent years, large language models have made significant advances in the field of medical applications, particularly in areas such as medical queries, medical examinations, and medical assistants (Chang et al., 2024). Numerous LLMs have been developed for specialized tasks—for example, AquliaMed, a bilingual LLM that applies RLHF for fine-tuning to improve its performance in medical dialogues and multiple-choice question answering (Zhao et al., 2024). Although the intersection between large language models and habit research remains a gap, the progress of LLMs in the medical applications domain can serve as a valuable point of reference.

Therefore, testing large language models on the classification of habits and their contexts is not only of theoretical interest but also of significant practical relevance.

Current research on “habit context” often remains at the level of structured variables (such as location, time, etc.), lacking systematic extraction and representation of complex context cues in natural language. Leveraging the advantages of LLMs in handling semantic complexity and context sensitivity, it becomes possible to automate the classification, encoding, and modeling of contextual cues for habitual behavior.

Especially when combined with supervised fine-tuning or Reinforcement Learning From Human Feedback (RLHF), LLMs can maintain their broad generalization ability while responding more precisely to domain-specific tasks (Naveed et al., 2024). We can observe whether the model is capable of distinguishing whether a behavior qualifies as a habit in different contexts, and whether it can identify the key contextual factors of habits. This not only provides methodological support for the digital modeling of habitual behavior but also lays the foundation for the development of intelligent systems with behavioral intervention capabilities.

# Zielsetzung

In this Belegarbeit, the objective is to identify habits in unstructured texts by means of pre-trained Large Language Models (LLMs) and to classify them with regard to their contextual characteristics. The aim is for the resulting prototypical end-to-end system to automatically determine whether a behavior qualifies as a habit and to classify the context of the habit accordingly.  
In this process, the following two research questions are of particular importance:

1. How effectively can pre-trained Large Language Models (LLMs) be used to identify habits in unstructured texts?
2. How accurately can the contextual characteristics (e.g., time, location, preceding actions) of identified habits be automatically classified by LLMs?

# Forschungsdesign

Here is the English translation of your paragraph, staying true to your requirement for a direct translation without stylistic embellishments:

This study will adopt a systematic approach aimed at utilizing large language models (LLMs) to automatically identify and classify habits and their contextual characteristics from unstructured textual data.  
First, a multilingual dataset (FGDH-HabitDataset) will be constructed, containing positive examples, negative examples, and misclassified examples of habitual behavior, which will serve as the foundation for the subsequent research steps.  
Based on this, the first LLM-based architectural component (Component K1) will be developed for the automatic identification of habitual behaviors in text. The performance of this component will be evaluated using standard metrics such as accuracy, precision, recall, and F1-score.  
Next, a second LLM-based architectural component (Component K2) will be implemented, which will systematically classify the previously identified habitual behaviors according to predefined contextual characteristics (such as time, physical environment, preceding actions, participants, internal states, or the behavior itself). The effectiveness of Component K2 will be validated by comparing its output with human annotations, using standardized evaluation metrics.  
Finally, these two main components will be integrated into a prototype-level end-to-end system that is capable of recognizing habitual behaviors and classifying their context, and the extracted structured information will be stored in a graph database to enable efficient access and querying of habit data.

# Vorläufige Gliederung

Introduction

* Problem statement and motivation
* Objective of the work

Theoretical Background

* Concept and significance of habits
* Contextual characteristics of habits
* Application possibilities of Large Language Models (LLMs)

Methodology

* Construction of a ground-truth dataset (FGDH-HabitDataset)
  + Development of component K1 for habit recognition
* Evaluation of component K1 (e.g. accuracy, precision, recall, F1-score)
  + Development of component K2 for context classification
* Evaluation of component K2
* Prototypical implementation of a pipeline

Implementation and Execution

* Present the concrete development process for each step

Results

* Present the results obtained for each step

Discussion

* Evaluation of the results
* Limitations of the developed components

Conclusion and Outlook

* Summary of the findings
* Preparation for the research seminar

# Literaturverzeichnis:

Anisuzzaman, D. M. (n.d.). Fine-Tuning Large Language Models for Specialized Use Cases. *MAYO CLINIC PROCEEDINGS*.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020, July 22). *Language Models are Few-Shot Learners*. arXiv. https://doi.org/10.48550/arXiv.2005.14165

Carden, L., & Wood, W. (2018). Habit formation and change. *Current Opinion in Behavioral Sciences*, *20*, 117–122. https://doi.org/10.1016/j.cobeha.2017.12.009

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., … Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

Labrecque, J. S., Lee, K. M., & Wood, W. (2024). Measuring context–response associations that drive habits. *Journal of the Experimental Analysis of Behavior*, *121*(1), 62–73. https://doi.org/10.1002/jeab.893

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025, March 23). *Large Language Models: A Survey*. arXiv. https://doi.org/10.48550/arXiv.2402.06196

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., … Mian, A. (2024, October 17). *A Comprehensive Overview of Large Language Models*. arXiv. https://doi.org/10.48550/arXiv.2307.06435

Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The Pull of the Past: When Do Habits Persist Despite Conflict With Motives? *Personality and Social Psychology Bulletin*, *37*(11), 1428–1437. https://doi.org/10.1177/0146167211419863

Peña, A., Morales, A., Fierrez, J., Serna, I., Ortega-Garcia, J., Puente, I., … Cordova, G. (2023). *Leveraging Large Language Models for Topic Classification in the Domain of Public Affairs*. https://doi.org/10.1007/978-3-031-41498-5\_2

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (n.d.). *Language Models are Unsupervised Multitask Learners*.

Tao, Z., Jin, Z., Bai, X., Zhao, H., Feng, Y., Li, J., & Hu, W. (2023, May 24). *EvEval: A Comprehensive Evaluation of Event Semantics for Large Language Models*. arXiv. https://doi.org/10.48550/arXiv.2305.15268

Wood, W. (2017). Habit in Personality and Social Psychology. *Personality and Social Psychology Review*, *21*(4), 389–403. https://doi.org/10.1177/1088868317720362

Wood, W., Mazar, A., & Neal, D. T. (2022). Habits and Goals in Human Behavior: Separate but Interacting Systems. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(2), 590–605. https://doi.org/10.1177/1745691621994226

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, *114*(4), 843–863. https://doi.org/10.1037/0033-295X.114.4.843

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(Volume 67, 2016), 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Yang, K.-C., & Menczer, F. (2025, February 14). *Accuracy and Political Bias of News Source Credibility Ratings by Large Language Models*. https://doi.org/10.1145/3717867.3717903

Zhao, L., Zeng, W., Shi, X., Zhou, H., Hao, D., & Lin, Y. (2024, June 18). *Aqulia-Med LLM: Pioneering Full-Process Open-Source Medical Language Models*. arXiv. https://doi.org/10.48550/arXiv.2406.12182