

TOP-Forschungsprojekte 2025

LANTMARK - Potential und Grenzen von Large Language Modellen für Anwendungen im Watermarking von Text und gesprochener Sprache

Professur: Intelligente Informationssysteme
Prof. Dr. Benno Stein

gemeinsam mit
Algorithm Engineering
PD Dr. Andreas Jakoby

Fakultät Medien

Laufzeit: 1. April 2025 bis 31. März 2028

Drittmittelgeber: BMBF

Fördersumme: 639.133,82 Euro

Beschreibung:

Gemeinsam mit dem Fraunhofer-Institut für Digitale Medientechnologie IDMT in Ilmenau und der Artefact Germany aus Hamburg widmet sich das Projekt dem sogenannten Text-Watermarking. Hierbei handelt es sich um Verfahren, mit dem sich Markierungen als Teil eines Textes versteckt unterbringen lassen, um beispielsweise die Herkunft des Textes nachzuvollziehen oder Textänderungen zu erkennen. Ein weiteres Ziel des Projektes ist es, große Sprachmodelle, sogenannte Large Language Models (LLMs) so zu modifizieren, dass ihre generierten Texte ein Wasserzeichen enthalten, mittels dem sich das LLM identifizieren lässt. Bisher einzigartig ist die Kombination aus Text- und Audioanalyse in Lantmark, bei der geschriebene und gesprochene Inhalte gemeinsam überprüft werden.

Summary:

Together with the Fraunhofer Institute for Digital Media Technology IDMT in Ilmenau and Artefact Germany in Hamburg, the project is dedicated to so-called text watermarking. This is a process that allows markings to be hidden in a text, for example to trace the origin of the text or to detect changes to it. Another goal of the project is to modify large language models (LLMs) so that the texts they generate contain a watermark that can be used to identify the respective LLM. The combination of text and audio analysis in Lantmark is unique to date: the researchers propose a bimodal authentication method in which written and spoken content are verified together.

Weitere Informationen: <https://weimar.webis.de/research.html#bmbftr-lantmark-2025>