

RAIDAR
RAPID
AI
BASED
DETECTION
OF
AGGRESSIVE
OR
RADICAL
CONTENT
ON
THE
WEB

RAIDAR

Projekt-Newsletter #3

*RAPID ARTIFICIAL INTELLIGENCE BASED DETECTION OF
AGGRESSIVE OR RADICAL CONTENT ON THE WEB
(RAIDAR)*

FFG KIRAS 2020



Liebe Leserinnen und Leser,

dies ist die dritte Ausgabe des Newsletters zum Forschungsprojekt *RAIDAR* (Rapid Artificial Intelligence based Detection of Aggressive or Radical content on the Web).

In der technologischen Entwicklung wurde in den letzten drei Monaten weiter an der Plattformarchitektur gearbeitet, die aus Crawlern, Datenbanken, Datenverarbeitungs- und Datenanalyseeinheiten, grafischen Benutzeroberflächen und Orchestrierungs-Tools besteht. Die Komplexität der RAIDAR-Architektur ergibt sich daraus, dass einerseits eine Vielzahl an Machine-Learning-Modellen auf eine große Anzahl an Daten angewendet werden muss, was eine logistische Herausforderung hinsichtlich der limitierten Rechenressourcen darstellt. Andererseits resultiert die Kombination von Ansätzen aus dem Bereich der Semantic AI und Machine-Learning-Modellen in einem gesteigerten Aufwand. Fortschritte wurden auch bei der intelligenten Text- und Bildanalyse gemacht, u. a. durch die Entwicklung von verschiedenen Text-Klassifikatoren in Form von Detektoren für Hassrede, Toxizität, Diskriminierung und Sexismus.

Begleitend zur technologischen Entwicklung finden mittlerweile laufend Workshops und Evaluierungen statt, um die rechtlichen und ethischen Anforderungen im Sinne von Law-by-Design und Ethics-by-Design zu berücksichtigen. Die rechtliche und ethische Begleitforschung beschäftigte sich daher in den letzten drei Monaten vor allem damit, die entsprechenden Anforderungen verständlich zu machen. So wurde aus rechtlicher Perspektive analysiert, wie einzelne Rechtsmaterien wie die Europäische Datenschutz-Richtlinie Polizei und Justiz (EU) 2016/680, die Datenschutz-Grundverordnung (EU) 2016/679, das österreichische Datenschutzgesetz und verfassungsrechtliche Anforderungen zusammenwirken und etwaige Einschränkungen sowohl in der Entwicklung als auch in der Anwendung der bei RAIDAR entwickelten Technologien bedingen. In der ethischen Begleitforschung wurde geklärt, welche grundsätzlichen ethischen Fragen rund um den Einsatz von Künstlicher Intelligenz sowie den Umgang mit Hassäußerungen im Internet aufgeworfen werden und wie diesen im RAIDAR-Projekt begegnet werden kann – etwa durch die Implementierung der Ethik-Leitlinien für eine vertrauenswürdige Künstliche Intelligenz, die von der Hochrangigen Expertengruppe für Künstliche Intelligenz der Europäischen Kommission 2019 veröffentlicht wurden.

Über RAIDAR:

RAIDAR wird im Rahmen des österreichischen Sicherheitsforschungsprogramms [KIRAS](#) gefördert, einem nationalen Programm zur Förderung der Sicherheitsforschung in Österreich. Die Programmverantwortung für das KIRAS-Programm liegt beim *Bundesministerium für Landwirtschaft, Regionen und Tourismus (BMLRT)*. Das BMLRT hat die *Österreichische Forschungsförderungsgesellschaft (FFG)* mit dem Programm- und Schirmmanagement für das KIRAS-Programm beauftragt.

Gemeinsam mit dem AIT (Austrian Institute of Technology GmbH) als leitende Organisation erforschen Semantic Web Company GmbH, SCENOR - Verein zur Erforschung aktueller gesellschaftlicher Herausforderungen, Research Institute AG & Co KG und LIQUA - Linzer Institut für qualitative Analysen Methoden der Erhebung und Bewertung von demokratiegefährdenden Inhalten in großen Datenbeständen, einschließlich Hass und Anzeichen von Radikalisierung. RAIDAR will in diesem Kontext nicht nur neue Methoden zur Sondierung dieser Inhalte liefern, sondern eine anwender*innenfreundliche und IT-basierte Plattform zur teilautomatisierten und versatilen Analyse von großen Datenbeständen aus unterschiedlichsten Quellen entwickeln. Ziel dabei ist, das System in die Lage zu versetzen, automatisch und mit Hilfe von Künstlicher Intelligenz relevante Inhalte zu Hass und Anzeichen von Radikalisierung in diesen Datenbeständen zu identifizieren, die aus strafrechtlicher Sicht relevant sein könnten. Als Bedarfsträger fungiert das Bundesministerium für Justiz (BMJ), das durch die Entwicklung dieses teilautomatisierten Assistenzsystems bei seiner juristischen Arbeit entlastet werden soll. Im Rahmen des Forschungsprojekts wird dabei auch eine Technikfolgenabschätzung zu den ethischen Grenzen und rechtlichen Schranken im Kontext von KI-basierter automatisierter Erfassung von Daten durchgeführt.

Weitere Informationen zu den Hintergründen und Zielsetzungen des Projekts finden Sie auf unserer Homepage unter raidar.at/projekt.



Der RAIDAR-Newsletter wird herausgegeben von:

AIT (Austrian Institute of Technology GmbH)
Giefinggasse 4, 1210 Wien, Österreich
Tel.: +43 50 550 - 0
E-Mail: office@ait.ac.at
Website: raidar.at
Autorisierter RAIDAR-Vertreter: Alexander Schindler