Komm ins #TeamBSI

Wir sind die Cyber-Sicherheitsbehörde des Bundes. Gemeinsam gestalten wir mit bislang rund 1.500 Beschäftigten eine sichere digitale Zukunft für Deutschland. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) bietet Studierenden von Fachhochschulen, Hochschulen und Universitäten die Möglichkeit der Betreuung von Abschlussarbeiten, auch in Verbindung mit einem Praktikum/Praxisprojekt.

Masterarbeit
"Cybersicherheitsaspekte im neuroexpliziten KI-System HyLEAP"

Hintergrund:

Von klassischen Anwendungen wie Bilderkennung, bis hin zu hochaktuellen Themen, wie Sprachmodellen oder autonomem Fahren: künstliche Intelligenzen sind aus der heutigen Zeit nicht mehr wegzudenken. Im Zentrum der Betrachtung stehen derzeit oft neuronalen Netze, die in Anwendungen beachtliche Leistungen zeigen, jedoch mehrere Schwächen, wie z.B. eine fehlende Nachvollziehbarkeit oder besondere Anfälligkeit für neuartige Angriffsvektoren, aufweisen. Um diesen Schwächen zu adressieren, werden derzeit in der Forschung neuroexplizite Modelle diskutiert, die neuronale Netze mit anderen KI-Techniken kombinieren, die Wissen explizit darstellen.

1. PLATZ
Attraktivste Arbeitgeber 2022
Für Absolventen
Öffentlicher Sektor
trendence

Bundesamt für Sicherheit in der Informationstechnik

Deutschland Digital•Sicher•BSI•

HyLEAP ist ein Beispiel für ein solches neuroexplizites oder hybrides KI-System, in dem ein neuronales Netz für tiefes bestärkendes Lernen (Deep Reinforcement Learning) einen automatischen Aktionsplaner unterstützt. Ein mögliches Anwendungsgebiet ist hierbei autonomes Fahren, genauer, die kollisionsfreie Steuerung von autonomen Fahrzeugen. Das hybride System wurde 2019 am DFKI und der Universität des Saarlandes entwickelt und im genannten Anwendungsgebiet simulativ eingesetzt.

In der Masterarbeit soll HyLEAP in Hinblick auf Aspekte der Cybersicherheit analysiert werden. Mögliche Angriffe und Verteidigungen sollen getestet, und die Widerstandsfähigkeit des Systems untersucht werden. Der zentrale Fokus soll hierbei auf KI-spezifischen Angriffen liegen. Von besonderem Interesse ist die hybride Natur des Systems, und wie sich diese auf die Sicherheitsaspekte auswirkt. Ziel ist es, ein Verständnis der speziellen sicherheitsrelevanten Eigenarten des Systems zu entwickeln, diese systematisch darzustellen und praxisnah zu testen und zu bewerten.

Ziele: Im Folgenden werden mögliche Ziele für Untersuchungen im Rahmen dieser Arbeit aufgeführt; die Auswahl und Gestaltung von Schwerpunkten aus dieser Liste werden für die konkrete Aufgabenstellung der Arbeit individuell mit Ihnen abgestimmt.

- Untersuchung des Systems HyLEAP hinsichtlich Anfälligkeiten für **KI-spezifische Angriffe & Verteidigungen (**Beispiele solcher Angriffe sind Adversarial Attacks oder Data Poisoning)
- Recherche & Analyse, welche Angriffe & Verteidigungen für die einzelnen Systeme relevant sind
- Konzeptionelle Analyse, inwiefern sich Angriffe & Verteidigungen auf das neuroexplizite Gesamtsystem übertragen lassen, bzw. Bewertung der Relevanz in dem konkreten Fall
- Verallgemeinerung der Aussagen: Gibt es Angriffe und Verteidigungsmaßnahmen, die spezifisch für die Klasse der neuroexpliziten Systeme sind?
- Bewertung der Robustheit und Widerstandsfähigkeit (Resilience) des Systems in Bezug auf KI-spezifischen Angriffe, insbesondere
    - Mögliche Quantitative Bewertung von Eigenschaften
    - Systematische Analyse von Sicherheitsgewinnen oder -verlusten durch die hybride Architektur
- Prüfung inwiefern für das **konkrete Szenario** Erklärbarkeitsaspekte relevant sind und wie sich die neuroexplizite Architektur auf diesen Aspekt auswirkt
- Bewertung der Praxisrelevanz und Wirksamkeit von Angriffen und Verteidigungen, insbesondere
    - Erarbeitung eines Bewertungsrahmens für Angriffe & Verteidigungen
    - Aufbau einer geeigneten Testumgebung
    - Durchführung von empirischen Versuchen

Nützliche Vorkenntnisse / Anforderungen:

- Masterstudentin / Masterstudent im Bereich Informatik oder Data Science and AI an der Universität des Saarlandes (UdS), Fakultät für Mathematik und Informatik.
- Interesse und grundlegende Kenntnisse im Bereich Cybersicherheit und KI, insbesondere im Bereich Deep Learning und automatisches Planen.
- Programmiererfahrung in Python.
- Interesse an der Nutzung des Fahrsimulators CARLA.

Sie haben Interesse?

- Bitte senden Sie Ihre Bewerbung (Motivationsschreiben, Lebenslauf und aktuelle Notenübersicht) per E-Mail an hochschulen@bsi.bund.de
- Fachliche Ansprechpartner bei Rückfragen sind Dr. Thomas Jung (BSI Saarbrücken, thomas.jung@bsi.bund.de) und PD Dr. Matthias Klusch (DFKI und UdS, matthias.klusch@dfki.de)
- Wir betrachten Diversität und die geschlechtsunabhängige berufliche Gleichstellung als wichtige Bestandteile unserer Personalpolitik, daher freuen wir uns über Bewerbungen von Menschen aller geschlechtlichen Identitäten, jeden Alters oder Herkunft. Frauen möchten wir ausdrücklich ermutigen, sich zu bewerben.

1. PLATZ
Attraktivste Arbeitgeber 2022
Für Absolventen
Öffentlicher Sektor
trendence

Bundesamt für Sicherheit in der Informationstechnik

Deutschland Digital•Sicher•BSI•