

Expertenkreis “KI-Sicherheit”

Der Expertenkreis¹ hat das Ziel einen praxisorientierten Erfahrungsaustausch zum Thema „KI-Sicherheit“ zwischen verschiedenen Stakeholdern zu ermöglichen und gemeinsam praktisch orientierte Best Practices bzw. Handlungsleitfäden für verschiedene Themen rund um das Thema „KI-Sicherheit“ zu erarbeiten. Daneben können Schulungen und Wissensvorträge im Rahmen des Expertenkreises durchgeführt werden.

In einem ersten Schritt soll ein praxisorientierter Handlungsleitfaden entstehen, der Pentests anwendungs-integrierter Large Language Models systematisiert. Dieser Leitfaden wird auch nach seiner initialen Veröffentlichung ein Arbeitsdokument bleiben und konstant weiterentwickelt werden, um adäquat auf neue Bedrohungen reagieren zu können. Mittel- und langfristig können auch weitere, verwandte Themen im Bereich KI-Sicherheit bearbeitet werden. Die Involvierung diverser Stakeholder in dem Kreis hat insbesondere auch das Ziel, die Bedarfe der einzelnen Zielgruppen hinsichtlich Publikationen oder Angeboten zu diskutieren, zu erfassen und durch gemeinsam erarbeitete Angebote zu bedienen.

Ziel des Kreises ist es weiterhin, den praxisorientierten Austausch der folgenden Akteure zu fördern:

- Branchen- und Berufsverbänden
- Behörden, insbesondere dem Bundesamt für Sicherheit in der Informationstechnik
- Herstellern, Betreibern und Dienstleistern im Bereich KI
- Forschungseinrichtungen

Der Mitarbeit liegt eine formlose Erklärung zur Teilnahme am Expertenkreis KI-Sicherheit unter Berücksichtigung des Code of Conduct (CoC) zu Grunde. Die Teilnehmenden treffen sich monatlich in Form von virtuellen Konferenzen. Präsenzveranstaltungen können in begründeten Ausnahmefällen auch durchgeführt werden, sind aber für die Zusammenarbeit nicht notwendig.

Der Expertenkreis KI-Sicherheit betreibt einen Bereich auf der Webseite der Allianz für Cyber-Sicherheit. Dort finden sich weitere Informationen zur Teilnahme am Expertenkreis KI-Sicherheit sowie die Daten zur Kontaktaufnahme mit dem Lenkungskreis: https://www.allianz-fuer-cybersicherheit.de/Webs/ACS/DE/Netzwerk-Formate/Veranstaltungen-und-Austausch/Expertenkreise/KI-Sicherheit/ki-sicherheit_node.html

¹ „Expertenkreis“ wird als feststehender Begriff verwendet, selbstverständlich sind Expertinnen und Experten Mitglieder des Kreises.

Code of Conduct (CoC) zum Expertenkreis KI-Sicherheit

0. Präambel

Zielsetzung und Gründung

Der Expertenkreis (im Folgenden auch Arbeitskreis genannt) hat das Ziel einen praxisorientierten Erfahrungsaustausch zum Thema „KI-Sicherheit“ zwischen verschiedenen Stakeholdern zu ermöglichen und gemeinsam praktisch orientierte Best Practices bzw. Handlungsleitfäden für verschiedene Themen rund um das Thema „KI-Sicherheit“ zu erarbeiten. Daneben können Schulungen und Wissensvorträge im Rahmen des Expertenkreises durchgeführt werden.

In einem ersten Schritt soll ein praxisorientierter Handlungsleitfaden entstehen, der Pentests anwendungs-integrierter Large Language Models systematisiert. Dieser Leitfaden wird auch nach seiner initialen Veröffentlichung ein Arbeitsdokument bleiben und konstant weiterentwickelt werden, um adäquat auf neue Bedrohungen reagieren zu können². Mittel- und langfristig können auch weitere, verwandte Themen im Bereich KI-Sicherheit bearbeitet werden. Die Involvierung diverser Stakeholder in dem Kreis hat insbesondere auch das Ziel, die Bedarfe der einzelnen Zielgruppen hinsichtlich Publikationen oder Angeboten zu diskutieren, zu erfassen und durch gemeinsam erarbeitete Angebote zu bedienen.

Ziel des Kreises ist es weiterhin, den praxisorientierten Austausch der folgenden Akteure zu fördern:

- Branchen- und Berufsverbänden
- Behörden, insbesondere dem Bundesamt für Sicherheit in der Informationstechnik
- Herstellern, Betreibern und Dienstleistern im Bereich KI
- Forschungseinrichtungen

Die offizielle Gründung des Expertenkreises KI-Sicherheit wird nach einem ersten Auftakttreffen im nachfolgenden Treffen beschlossen. Der Beschluss muss einstimmig sein.

1. Teilnehmende

Der Expertenkreis KI-Sicherheit steht Teilnehmenden folgender Institutionen offen:

- Branchen- und Berufsverbänden
- Behörden, insbesondere dem Bundesamt für Sicherheit in der Informationstechnik
- Herstellern, Betreibern und Dienstleistern im Bereich KI
- Forschungseinrichtungen

Initial entscheiden die Initiatoren des Expertenkreises über die Aufnahme interessierter Institutionen. Im Rahmen des Auftakttreffens wird eine Teilnehmendenliste erstellt, der eine formlose Erklärung zur Teilnahme am Expertenkreis KI-Sicherheit unter Berücksichtigung des Code of Conduct (CoC)

² Die Initiative für den Arbeitskreis wurde seitens sequire technology GmbH ergriffen. Anlass war die Anfrage eines Kunden im April 2024 nach einem Pentest eines anwendungs-integrierten LLMs. Da hierfür bisher kein standardisiertes Verfahren existierte, bestand der Wunsch einen systematischen Leitfaden zu erstellen.

zugrunde liegt. Teilnahme und Mitarbeit erfolgen erst nach vorhergehender Einladung. Im Folgenden entscheidet der Lenkungskreis über die Aufnahme weiterer Institutionen.

Teilnehmende können bei Vorliegen konkreter Gründe auf Gesuch eines anderen Teilnehmenden ausgeschlossen werden. Dazu ist ein Gesuch an den Lenkungskreis zu richten, welcher den Ausschlussvorschlag in den Kreis einbringt. Zum Ausschluss ist eine 2/3 Mehrheit (in absoluten Zahlen) erforderlich. Ob die Abstimmung beim nächsten Treffen oder vorher erfolgt, bleibt dem Lenkungskreis überlassen.

Es können bis zu zwei Vertreterinnen und Vertreter einer teilnehmenden Institution aktiv an den Treffen des Expertenkreises mitwirken. Generell ist es möglich, dass sich mehr als zwei Personen einer teilnehmenden Institution als Teilnehmende melden, in diesem Fall wird durch interne Absprachen sichergestellt, dass maximal zwei Personen der teilnehmenden Institution pro Treffen anwesend sind. Die Vertreterinnen und Vertreter können mit Blick auf verschiedene Themen, die im Arbeitskreis bearbeitet werden, zwischen einzelnen Terminblöcken variieren. Ein Personenwechsel ist dem Lenkungskreis unter Zusendung der formlosen Erklärung (zur Teilnahme am Expertenkreis KI-Sicherheit unter Berücksichtigung des CoC) der neuen Teilnehmenden, sowie durch eine ebenfalls formlose Erklärung zur Beendigung der Teilnahme durch die ehemaligen Teilnehmenden, anzuzeigen. In begründeten Ausnahmefällen können mehr Personen einer Institution (temporär) an den Treffen mitwirken. Über die Ausnahmen entscheidet der Lenkungskreis.

Um ein produktives Arbeiten zu ermöglichen, werden maximal 60 Personen pro Treffen zugelassen. Der Lenkungskreis kann diese Zahl später noch korrigieren.

Der Kreis soll einen vertraulichen Austausch ermöglichen. Die Treffen sind nicht öffentlich.

2. Die Allianz für Cyber-Sicherheit

Mit der Allianz für Cyber-Sicherheit (ACS) steht das Bundesamt für Sicherheit in der Informationstechnik (BSI) als die nationale Sicherheitsbehörde Unternehmen und Institutionen zur Seite. Bereits seit 2012 arbeitet das BSI intensiv mit Partnern und Multiplikatoren aus Wirtschaft und Forschung zusammen, um strategische und praktische Hilfestellung zur Umsetzung von Informationssicherheit in den Unternehmen zu leisten und so die Cybersicherheit am Wirtschaftsstandort Deutschland zu fördern.

Die wesentlichen Punkte im Hinblick auf den Expertenkreis KI-Sicherheit werden nachfolgend zusammenfassend dargestellt:

- Die Teilnahme an der Allianz für Cyber-Sicherheit erfolgt auf freiwilliger Basis und ist kostenfrei. Sie kann beidseitig jederzeit ohne Einhaltung von Fristen in schriftlicher Form beendet werden.
- Es ergeben sich keine weiteren rechtlichen Verpflichtungen außer der Zustimmung zu den Bedingungen des Traffic Light Protocol (TLP) was ebenfalls im Interesse der Mitglieder des Expertenkreises ist.
- Die im Rahmen der ACS verbreiteten Informationen werden, entsprechend ihrer Sensitivität, gemäß dem Traffic Light Protocol (TLP) eingestuft.
- Die Regelungen zum TLP werden durch das Merkblatt „Behandlung vertraulicher Informationen“ festgelegt und erläutert. Alle Zugangsberechtigten aus den teilnehmenden Institutionen und Unternehmen haben sich persönlich dazu zu verpflichten, Informationen, welche sie durch oder im Zusammenhang mit der ACS erlangen, entsprechend den Regelungen des TLP zu behandeln und diese unbefugten Dritten nicht zugänglich zu machen.³

³ Das Traffic Light Protocol (TLP) ist eine standardisierte Vereinbarung zum Austausch schutzwürdiger aber nicht formell eingestufte Informationen. Alle Dokumente werden in TLP-Stufen eingeteilt, die die Bedingungen für ihre Weitergabe regeln. Weitere Informationen sowie die geltenden TLP-Stufen finden sich hier: <https://www.bsi.bund.de/dok/tlp-merkblatt>

3. Leitlinien der Zusammenarbeit

Folgende Leitlinien bestimmen die Zusammenarbeit im Expertenkreis KI-Sicherheit:

- Die Kooperation ist freiwillig und kann jederzeit beendet werden
- Businessmodelle Einzelner dürfen unter der Kooperation nicht leiden
- Die Vertraulichkeit der Inhalte hat höchste Priorität, wobei die Chatham-House-Regeln Anwendung finden.⁴
- Informationen und Interessen anderer Mitglieder werden geschützt
- Es erfolgt ein regelmäßiger Erfahrungsaustausch und eine kontinuierliche Weiterbildung innerhalb des Kreises
- Durch wechselseitige Beiträge und Informationen werden Erfahrungen und Best Practices zum sicheren Einsatz von KI innerhalb des Kreises geteilt.
- Von allen Teilnehmenden wird eine aktive Mitarbeit erwartet.
- Die eigene Arbeit und die Zusammenarbeit sollen ein Vorbild für KI-Sicherheitsexpertinnen und -experten sein.

4. Kommunikation

Für die Kommunikation unterhält der Expertenkreis KI-Sicherheit neben einer Webseite mit grundlegenden Informationen und einer Kontaktmöglichkeit des Leitungsgremiums eine zentrale Mailingliste aller Teilnehmenden, Newsletter oder Kollaborationsplattform, die für die Kommunikation und Arbeit zwischen den Teilnehmenden eingesetzt werden soll. Die Nachrichteninhalte werden archiviert, wobei der Zugang zum Archiv nur durch die Personen der Liste möglich ist.

Daneben ist über die E-Mail-Adresse xprt-ki-sicherheit@bsi.bund.de der Lenkungskreis des Expertenkreises KI-Sicherheit erreichbar.

5. Abstimmungen und Wahlen

Abstimmungen und Wahlen werden bei den in der Regel während der Treffen durchgeführt und sind nicht geheim. In unkritischen Fällen erfolgen Sie unkompliziert per mündlicher Mitteilung, Handzeichen oder Chatbeitrag. In kritischen Fällen erfolgen sie per direkter Mitteilung (z.B. per Mail) an den Lenkungskreis. In diesem Fall verpflichtet sich der Lenkungskreis alle Informationen bezüglich der Stimmabgabe vertraulich zu behandeln. Der Lenkungskreis legt fest, welche Abstimmungen/ Wahlen kritisch sind.

6. Veranstaltungen

Der Arbeitskreis KI-Sicherheit führt monatlich Arbeitstreffen (im Rahmen dieses Dokuments „Treffen“ genannt) durch. Treffen sind hierbei kein Selbstzweck und werden effizient gestaltet. Eingeladen sind alle Teilnehmenden des Arbeitskreises. Gäste dürfen nach vorheriger Anmeldung beim Lenkungskreis teilnehmen. Die Veranstaltungen sind üblicherweise virtuell; Präsenzveranstaltungen können in begründeten Ausnahmefällen stattfinden. Der Umfang bemisst sich an der Menge der zu besprechenden Punkte und sollte üblicherweise nicht mehr als 30-60 Minuten beanspruchen.

Für alle Treffen gilt der nachfolgende Rahmen als Agenda, wobei die Finalisierung mit einem Vorlauf von ca. einer Woche erfolgt und die finale Agenda an die Teilnehmenden verschickt wird:

- Eröffnung und Begrüßung

⁴ Chatham-House-Regeln besagen, dass den Teilnehmenden die freie Verwendung der erhaltenen Informationen unter der Bedingung gestattet ist, dass weder die Identität noch die Zugehörigkeit von Redenden oder anderen Teilnehmenden preisgegeben werden dürfen.

- Festlegung Moderation und Protokollierung
- Neuigkeiten und Ankündigungen der verschiedenen Stakeholder
- Inhaltliche Vorstellungen und Diskussionen, je nach Thema und Bedarf
- Ausblick mit Festlegung eines Termins für das nächste Treffen

7. Gremien (Lenkungskreis)

Als Gremium nimmt der Lenkungskreis eine koordinierende Funktion ein und ist zugleich das Entscheidungsgremium des Expertenkreises KI-Sicherheit. Er besteht aus mindestens fünf und maximal sieben gewählten Personen aus dem Kreis der Teilnehmenden. Davon muss mindestens eine Vertreterin oder ein Vertreter des BSI im Lenkungskreis sein und ist damit für den Lenkungskreis gesetzt. Die übrigen Personen werden für die Dauer von vier Jahren gewählt. Die Aufgaben des Lenkungskreises sind u.a.:

- Steuerung der inhaltlichen Weiterentwicklung des Expertenkreises KI-Sicherheit durch das Setzen thematischer Schwerpunkte
- Unterstützung des Wachstums des Arbeitskreises z. B. durch aktive, positive Darstellung des Arbeitskreises nach außen oder gezielte Ansprache möglicher neuer Teilnehmender
- Entscheidung über Aufnahme von neuen Institutionen (per Mehrheitsentscheidung)
- Feststellung von Verstößen gegen Vereinbarungen und Beschluss von Sanktionen
- Protokollierung und Verteilung aller Entscheidungen
- Organisation und Durchführung von Treffen

Der Lenkungskreis wählt einen Vorsitz und eine Stellvertreterin oder einen Stellvertreter aus seinen Reihen. Die Aufgaben der oder des Vorsitzenden umfassen dabei:

- Sammlung der Tagesordnungspunkte und Erstellung, sowie Versand der Agenda
- Versand der Einladungen für die Treffen
- Leitung der Treffen
- Kontrolle und Versand des Protokolls an die Teilnehmenden

8. Änderungen des CoC

Änderungsvorschläge des Code of Conduct sind schriftlich zu verfassen und mindestens drei Wochen vor dem nächsten Treffen an den Lenkungskreis xprt-ki-sicherheit@bsi.bund.de zu adressieren. Die Änderungsvorschläge werden mindestens eine Woche vor dem nächsten Treffen durch den Lenkungskreis an alle Teilnehmenden über die offizielle Mailingliste versandt.

Die Entscheidung über Änderungen erfolgt durch die Abstimmung beim Treffen, wobei eine 2/3 Mehrheit der möglichen Stimmen ausreicht. Die Anzahl der möglichen Stimmen wird zu Beginn des Treffens festgelegt. Nach erfolgter Zustimmung treten die Änderungen der CoC sofort in Kraft. Die überarbeiteten Grundsätze werden allen Teilnehmenden über die Mailingliste zur Verfügung gestellt.

9. Auflösung des Expertenkreises KI-Sicherheit

Anträge zur Auflösung des Arbeitskreises sind schriftlich zu verfassen und mindestens drei Wochen vor dem nächsten Treffen an den Lenkungskreis xprt-ki-sicherheit@bsi.bund.de zu adressieren. Sie sind mindestens eine Woche vor dem nächsten Treffen über die Mailingliste an alle Teilnehmenden zu versenden.

Die Entscheidung zur Auflösung des Arbeitskreises KI-Sicherheit erfolgt durch die Abstimmung beim nächsten Treffen, wobei eine 3/4 Mehrheit der möglichen Stimmen ausreicht. Die Anzahl der möglichen Stimmen wird zu Beginn der Veranstaltung festgelegt.