MONTHLY SECURITY SUMMARY

)SCIP(

AUSGABE NOVEMBER 2022
BROWSER REQUEST-SMUGGLING UND TRUST

BROWSER-GESTÜTZTES REQUEST SMUGGLING

Neue Recherchen zeigen, dass Request Smuggling nicht nur zwischen Ketten von Servern funktionieren. Wir erklären in unserem Beitrag die Möglichkeiten eines Angriffs.

PARADOXE VERHALTENS-WEISEN BEI TRUST

Die Forschung hat die Bedeutung von Vertrauen zwischen Mensch und KI gezeigt: Kein Vertrauen, keine Nutzung. Ausgehend vom Datenschutzparadoxon untersuchen wir ein potenzielles Vertrauensparadoxon.

November 2022: Die Welt ist klein

Was ich an der *Schweizer Cybersecurity-Branche* besonders mag ist, dass sie klein ist. Die Chance, dass man sich bei einem Projekt oder der nächsten Konferenz über den Weg läuft, ist gross. Es freut mich, wenn ich *alte Bekannte* wieder sehen kann.

Ich habe vergangene Woche neue Nachbarn gekriegt. Per Zufall hat sich herausgestellt, dass die Ehefrau schon einmal an einem Vortrag von mir zum Thema *Darknet* gewesen ist. Und vergangenen Samstag war ich an einem Weiterbildungsseminar ausserhalb der IT. Dort stellte ich fest, dass die Seminarleiterin eine ehemalige Studentin von mir im Bereich *Cybercrime* war.

Ich finde sowas toll, da man sofort ein Gesprächsthema hat. Man muss ja nicht bei Beruf, Cybersecurity und IT bleiben. Aber es ist ein guter Einstieg, um sich unkompliziert austauschen zu können.

Zeitgleich setzt das natürlich auch voraus, dass man anständig sein und sich eine gute Reputation aufbauen muss. Viel zu schnell spricht sich sonst herum, dass man nicht vertrauenswürdig oder gar hinterhältig ist. So etwas käme mir in keinster Weise in Frage. Manch anderer nimmt aber solche Unpässlichkeiten in Kauf. Meines Erachtens eine schlechte Entscheidung, denn Böswilligkeiten holen einem im Leben immer ein. Vor allem in der kleinen Schweiz.

Marc Ruef
Head of Research

NEWS

WAS IST BEI UNS PASSIERT?

PODCAST ZU KÜNSTLICHER INTELLIGENZ IN DER MEDIZIN

Dr. Alexandra Widmer hat sich im Podcast *docsdigital – Der Podcast für innovative Ärzte* mit Marisa Tschopp zum Thema künstlicher Intelligenz in der Medizin unterhalten. Die Sendung ist unter anderem auf Spotify verfügbar und behandelt Fragen rund um Chancen und Nutzen der KI aus psychologischer Perspektive mit Fokus auf die komplizierte Vertrauensbeziehung zwischen Ärzten, Patienten und KI.

Weitere News zu unserem Unternehmen finden Sie auf unserer Webseite.

SCIP BUCHREIHE

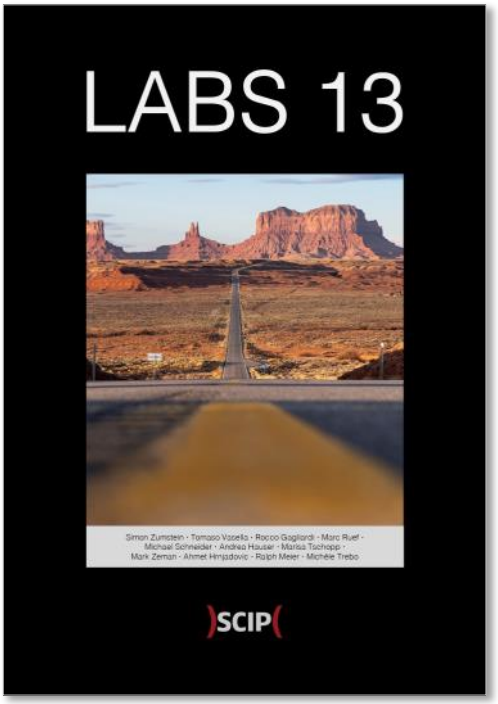
UNSER NEUES JAHRBUCH

Erneut veröffentlichen wir die aktuelle Ausgabe unseres Jahrbuchs. Bereits zum 13ten Mal fassen wir in diesem die Fachbeiträge von einem Jahr Forschung im Bereich Cybersecurity zusammen.

Das Buch ist wiederum sowohl in deutscher (ISBN 978-3-907109-30-4) als auch in englischer Sprache (ISBN 978-3-907109-31-1) verfügbar. Das Vorwort wurde von Dr. iur. David Vasella verfasst und setzt sich mit den rechtlichen Rahmenbedingungen der Informationssicherheit auseinander.

In unserem Katalog finden sich ebenfalls themenspezifische Bücher zu Künstlicher Intelligenz (ISBN 978-3-907109-14-4) und Sicherheit in der Hotellerie (978-3-907109-16-8).

Weitere Informationen auf unserer Webseite.



ISBN 978-3-907109-30-4 [de]
ISBN 978-3-907109-31-1 [en]

MAN MUSS SICH NICHT VOR ALLEM SCHÜTZEN

Bildquelle: https://pixabay.com/images/id-6969822/

ANDREA HAUSER

SO FUNKTIONIERT BROWSER-GESTÜTZTES REQUEST SMUGGLING

Browser-gestütztes Request Smuggling ermöglicht neue clientseitige Varianten des Request Smugglings, die nicht nur den Angriff auf eine Kette von mehreren Servern, sondern auch auf einen einzelnen Server ermöglichen. Im Vergleich zu den bisher gezeigten Request Smuggling-Angriffen sind die für diese Angriffe notwendigen Requests korrekt formatiert und können daher über einen Browser gesendet werden. Der Angriff auf einen einzelnen Server ist möglich, da der Angriff so angelegt ist, dass der Browser des Opfers seine eigene Verbindung zu einem verwundbaren Server desynchronisiert.

Dier Artikel basiert stark auf der Recherche von James Kettel, die an der Defcon 30 präsentiert wurde. Basierend darauf wurden in der Portswigger Web Academy Labs zum Thema Client-Side Desync aufgebaut, mit denen diese Angriffe in der Praxis ausprobiert werden können.

CL.0 REQUEST SMUGGLING

Die bisherigen Request Smuggling Angriffe haben alle darauf basiert unterschiedliche Interpretationen mittels Content-Length und Transfer-Encoding Headern zwischen mehreren Servern in einer Kette zu erreichen. Beim CL.0 Angriff geht es darum, dass gewisse Server immer davon ausgehen, dass der Request keinen Body mitschickt und die dementsprechend eine Content-Length von 0 annehmen. Wenn also der Back-End Server immer eine Content-Length von 0 annimmt, der Front-End Server allerdings die mitgegebene Content-Length übernimmt, kann es so zu Diskrepanzen kommen. Das Testen, ob eine CL.0 Schwachstelle vorhanden ist, ist sehr einfach.

```
POST /vulnerable-endpoint HTTP/1.1
Host: example.com
Connection: keep-alive
Content-Type: application/x-www-form-urlencoded
Content-Length: 33 --> gültige Content-Length

GET /shouldBe404 HTTP/1.1
Foo: x
```

Es handelt sich beim oben gezeigten um einen kompletten Request mit einer gültigen Content-Length, der im Body einen zweiten nicht vollständigen Request beinhaltet. Wenn in der Antwort auf einen zweiten gültigen Request kurz nach diesem ersten abgesetzten Request eine 404-Antwort erhalten wird, kann davon ausgegangen werden, dass eine CL.0

Request Smuggling Schwachstelle vorhanden ist. Mit der CL.0 Schwachstelle können die gleichen Schwachstellen ausgenutzt werden, wie sie in den letzten Labs aufgezeigt wurden. Am anfälligsten für CL.0 Schwachstellen sind Endpunkte, die keinen POST-Request erwarten, zum Beispiel statische Dateien oder Redirects.

CLIENT-SIDE DESYNC ANGRIFFE

Bis anhin wurden Request Smuggling Angriffe als ein serverseitiges Problem betrachtet, da es sich um Missverständnisse zwischen mehreren Servern in einer Kette handelt. Diese können nicht aus dem Browser ausgelöst werden, da sie bewusst manipulierte beziehungsweise ungültige Requests verwenden. Mit der CL.0 Schwachstelle gibt es nun allerdings die Möglichkeit Angriffe auch aus dem Browser auszulösen, da es sich bei der CL.0 Schwachstelle um Requests handelt, die nicht ungültig sind, also von einem Browser ausgestellt werden können. Eine Client-Side Desync Angriff ist das Ausnützen einer Schwachstelle, bei dem die Verbindung zwischen dem Browser und einem Server desynchronisiert wird. Dementsprechend eröffnen sich auch neue Angriffswege, da nun nicht mehr nur eine Kette von Servern angegriffen werden können, sondern auch alleinstehende beziehungsweise einzelne Server.

Abstrahiert gesagt besteht ein Client-Side Desync Angriff aus den folgenden Schritten:

1. Das Ziel des Angriffs wird auf eine beliebige Webseite gelockt, auf der durch Angreifer kontrolliertes JavaScript ausgeführt wird.

2. Das JavaScript löst einen Request auf die Webseite mit der Client-Side Desync Schwachstelle aus, bei der ein nicht abgeschlossener, bösartiger beziehungsweise für den Angreifer interessanter Request in der Pipeline übrig bleibt.

3. Dieser nicht abgeschlossene Request bleibt in der Pipeline zwischen Browser und Request, nach dem der initiale Request beantwortet wurde. Damit ist die Verbindung zwischen Browser und Server desynchronisiert.

4. Mit dem JavaScript wird nun ein weiterer Request an den Server ausgelöst, der den bereits angefangenen Request mit den Cookies des Opfers erweitert.

In Code sieht das ganze wie folgt aus:

```
<html>
<script>
fetch('https://example.com/vulnerable-endpoint',
{
    method: 'POST',
    body: 'GET /thisShouldBe404 HTTP/1.1
\\r\\nFoo: x',
    credentials: 'include' // damit der "with-
cookies" Connection Pool für den
Verbindungsaufbau verwendet wird
}).then(() => {
    location = 'https://example.com/' // hier
wird die desynchronisierte Verbindung genutzt
})
</script>
</html>
```

Nach Besuch dieser Webseite in einem Browser sollten im Netzwerk-Tab zwei Requests ersichtlich sein. Falls der Angriff erfolgreich war, wird der zweite Aufruf, der normalerweise mit 200 OK beantwortet wird, mit 400 Bad Request beantwortet. Dieses Beispiel kann für die Verifikation einer Client-Side Desync Schwachstelle genutzt werden. Für einen tatsächlichen Angriff würde der Angreifer im Body einen Request definieren, der den Inhalt des zweiten, desynchronisierten Request des Opfers im Account des Angreifers abspeichert. Damit würde der Angreifer an die Cookies des Opfers kommen und kann die Session des Opfers übernehmen.

Das oben gezeigte Beispiel funktioniert allerdings nicht in allen Fällen, denn wenn ein serverseitiger Redirect ausgelöst wird, wird der Code-Bereich der im .then() definiert ist, nicht ausgeführt, sondern der Browser folgt dem Redirect. Damit der Redirect verhindert werden kann und der Client-Side Desync Angriff funktioniert, kann folgender Code verwendet werden.

```
<html>
<script>
fetch('https://example.com/redirect-endpoint', {
    method: 'POST',
    body: 'GET /thisShouldBe404 HTTP/1.1
\\r\\nFoo: x',
    mode: 'cors', // wird bewusst gesetzt, damit
ein Fehler geworfen wird und der Redirect
abgebrochen wird
    credentials: 'include'
}).catch(() => {
    // hier wird der CORS-Fehler aufgefangen und
```

```
die desynchronisierte Verbindung genutzt
    fetch('https://example.com/', {
        mode: 'no-cors',
        credentials: 'include'
    })
})
</html>
</script>
```

Nach Besuch dieser Webseite in einem Browser sollten wie im vorherigen Beispiel im Netzwerk-Tab zwei Requests ersichtlich sein. Falls der Angriff erfolgreich war, wird der zweite Aufruf, der normalerweise mit 200 OK beantwortet wird, mit 400 Bad Request beantwortet. Eine umfängliche Schritt für Schritt Anleitung wie auf Client-Side Schwachstellen getestet werden kann ist in der Portswigger Web Academy ersichtlich.

ZUSAMMENFASSUNG

Mit Client-Side Desync Angriffen erweitert sich das Feld für Request Smuggling erneut und es kann zu weiteren spannenden Angriffsszenarien führen, da nun auch der Angriff auf einzelne Server möglich wird. Da die Research von James Kettel durchgeführt wurde, der bei Portswigger angestellt ist, bedeutet dies, dass das Tooling zum Auffinden dieser neuen Schwachstellen-Typen bereits ausgezeichnet ist, da in aktuellen Versionen von Burp die Identifikation dieser Schwachstellen bereits eingebaut ist. Für eine erweiterte Abdeckung wird die HTTP Request Smuggler Erweiterung empfohlen.

Andrea Hauser

Threat Intelligence mit Splunk

Laden Sie einfach und unkompliziert die Daten für das Vulnerability Management Ihrer Umgebung in Splunk. Die frei installierbare Applikation nutzt die offene API von VulDB, um Daten einzulesen, abzulegen und aufzuarbeiten. Noch nie war Vulnerability und Threat Intelligence so einfach. Setzen Sie sich mit uns in Verbindung!

)SCIP(
official data provider

> https://vuldb.com

MARISA TSCHOPP & PIERRE RAFIH

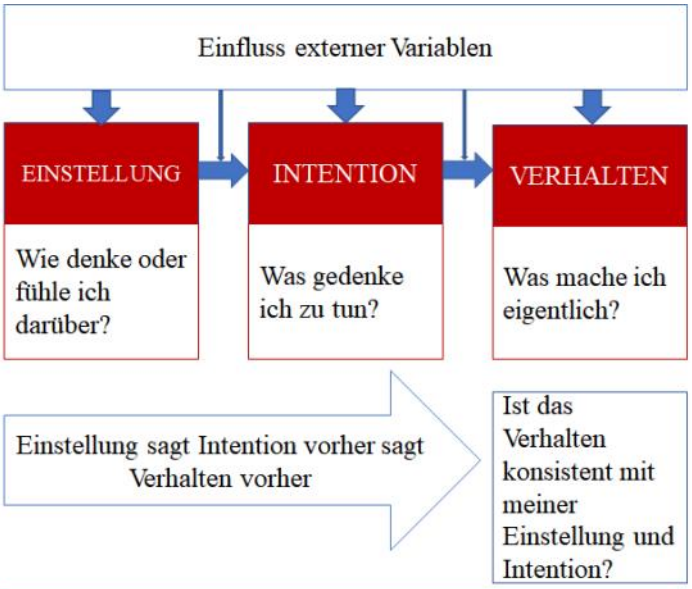
PARADOXE VERHALTENSWEISEN IN DER MENSCH-KI-INTERAKTION

Kein Vertrauen, kein Nutzen! Zahlreiche Studien aus dem Bereich der Mensch-Maschine-Interaktion haben die Hypothese “No trust, no use” bestätigt. Kurz gesagt, haben diese Studien gezeigt, dass Vertrauen ein wichtiger Prädiktor für Vertrauen ist, d.h. dafür, wie Menschen Technologie nutzen. Unsere Forschung konzentriert sich auf konversationelle KI, auch digitale, sprachgesteuerte oder intelligente Assistenten genannt, weil wir glauben, dass konversationelle Benutzerschnittstellen (CUI) sich von traditionellen Technologien unterscheiden und neue Forschungsansätze bieten. Sie erfordern nicht, dass sich die Nutzer direkt auf das Gerät konzentrieren und mit ihm interagieren, z. B. durch Berührung, so dass sie eher intuitiv sind. Gleichzeitig ruft das Gerät oft eine Art von Präsenz hervor und ist Teil der privatesten Bereiche des sozialen Lebens der Nutzer. Genauer gesagt wollen wir die Rolle des Vertrauens in der Konversations-KI oder der Mensch-KI-Interaktion besser verstehen. Die Ergebnisse einer unserer Pilotstudien, die auf einem AI4EU-Workshop vorgestellt wurden, stützen die Hypothese “kein Vertrauen, keine Nutzung”: Wir fanden heraus, dass tatsächliche Nutzer von konversationeller KI ein signifikant höheres Mass an Vertrauen haben (Tschopp et al., 2021).

Dies ist jedoch nur ein Teil des Ganzen, um die Rolle des Vertrauens in der Mensch-KI-Interaktion besser zu erklären. Es könnte sich lohnen, mit dem witzigen Slogan no trust, no use genauer zu sein: Die negativ gerahmte Hypothese impliziert eine positive, kongruente Vertrauens-Nutzungs-Beziehung: Je mehr die Nutzer einer Technologie vertrauen, desto eher nutzen sie sie, ihr Verhalten folgt ihrer Einstellung. Dieses Grundprinzip geht jedoch nicht auf inkongruentes Verhalten ein. Kein Vertrauen, keine Nutzung reicht nicht aus, um andere Szenarien zu erklären, z. B. dass einige Menschen, die einer bestimmten Technologie nicht vertrauen, sie dennoch nutzen. Wir wissen nur wenig darüber, wie das tatsächliche Vertrauensverhalten von Menschen aussieht, die kein oder wenig Vertrauen haben. Aufbauend auf den Erkenntnissen der Datenschutzforscher über das Datenschutzparadoxon wollen wir die Beziehung “kein Vertrauen, aber Nutzung” besser verstehen, die als tatsächliches Vertrauensverhalten interpretiert werden kann, das nicht mit der jeweiligen Vertrauenseinstellung übereinstimmt.

ATTITUDE BEHAVIOR GAP IN DER PRIVACY FORSCHUNG: DAS PRIVACY PARADOX

Ein Grossteil der Verbraucherpsychologie befasst sich mit Verbraucherentscheidungen: Die Forscher versuchen zu verstehen, wie Einstellungen, Werte, Überzeugungen oder Wissen gebildet werden und wie diese Variablen das Verbraucherverhalten beeinflussen. Die meisten Menschen würden es für normal halten, dass Menschen, die eine bestimmte Einstellung haben, nach dieser Einstellung handeln. Dies ist jedoch nicht immer der Fall. Das Phänomen, dass Menschen sich entgegen ihrer Einstellung verhalten, wird als Einstellungs-Verhaltens-Lücke bezeichnet und hat in der Datenschutzforschung besonderes Interesse gefunden. In zahlreichen Studien wurde festgestellt, dass die Nutzung der Technologie durch das so genannte Privatsphärenparadox gekennzeichnet ist (Norberg, 2007): Die Menschen sagen, dass sie sich um die Privatsphäre kümmern oder Bedenken haben, aber sie sind unvorsichtig, wenn es um ihr tatsächliches Verhalten geht. Die nachstehende Abbildung zeigt in vereinfachter Form, dass eine Einstellung gebildet wird, die die Absicht beeinflusst, die dann zu einer Handlung führt.



Attitude Behavior Gap

Einstellung: Ich habe Privacy Bedenken.
Intentions: Ich gedenke nicht meine Location data und alter zu teilen.
Szenarien für Verhalten:

1) Ich gebe meine Standortdaten und mein Alter nicht weiter (kein Unterschied im Verhalten)
2) Ich gebe meine Standortdaten und mein Alter weiter (klassisches "Datenschutzparadoxon")
3) Ich teile meinen Standort, mein Alter, meine Kontaktliste und ... alles, was nötig ist (super "Paradoxon")
4) Ich gebe meine Standortdaten weiter, aber nicht mein Alter (ausgewähltes "Paradoxon")
5) Ich gebe (alle oder ausgewählte) Daten an ausgewählte Organisationen weiter
6) ...?

Wie in der Abbildung zu sehen, ist es wichtig zwischen Absicht und Verhalten zu unterscheiden. Obwohl die Absicht im Allgemeinen ein starker Prädiktor für das Verhalten ist, haben Studien über das Datenschutzparadoxon gezeigt, dass die Absicht nicht genau genug ist, um das tatsächliche Verhalten zu erklären. Leider wird die Absicht meist als Ergebnisvariable untersucht, da sie viel einfacher und kostengünstiger zu operationalisieren ist als tatsächliche Nutzungsvariablen (Sheeran, 2002. Kokalakis (2015) schlägt vor, zwischen Privatsphärenbedenken (z. B. Stalking, sekundäre Nutzung durch Dritte oder missbräuchlicher Zugriff durch Arbeitgeber) und Privatsphäreneinstellungen (z. B. "Ich lege Wert auf Privatsphäre") zu unterscheiden. Ausserdem kann man zwischen der Intention und dem Verhalten zum Schutz der Privatsphäre unterscheiden: Annahme, z. B. Kauf, oder Verhalten zum Schutz der Privatsphäre, z. B. die Verweigerung von Informationen oder die Angabe falscher Informationen (Son & Kim, 2008).

DAS PARADOX DER PRIVATSPHÄRE IST KONTEXTABHÄNGIG

Die meisten Studien konzentrieren sich auf soziale Netzwerke und den elektronischen Handel, wobei die Literatur über Smartphones zunimmt. In der Forschung im Bereich der Smartphone-Nutzung wird häufig versucht, Fragen im Zusammenhang mit dem Datenschutzparadoxon im Kontext der Uses and Gratification Theory (UGT) oder der Information Boundary Theory (IBT) zu erklären (Sutanto 2013). Andere Forschungsansätze befassen sich mit Emotionen. In einer Studie, die kurz nach dem Facebook-Skandal um Cambridge Analytica durchgeführt wurde, stellen Sarabia-Sanchez et al. (2019) fest, dass es keinen Zusammenhang zwischen der Intensität der berichteten Emotionen und der Handhabung der Privatsphäre-Einstellungen durch die befragten Facebook-Nutzer gibt. Ein grosser Teil der Forschung über das Paradox der Privatsphäre im digitalen Bereich befasst sich mit der Cybersicherheit. Jenkins et al. (2021) stellen fest, dass der Wunsch der Menschen, den erforderlichen Aufwand zu minimieren, ihr tatsächliches Sicherheitsverhalten negativ moderiert. In ihrer Studie zum Sicherheitsverhalten von Smartphone-Nutzern stellen Das und Kahn (2016)

fest, dass die konsistentesten Prädiktoren für das Sicherheitsverhalten die wahrgenommene Wirksamkeit und die Kosten der Annahme von Sicherheitsmassnahmen sind. Ein anderer Ansatz zur Erklärung des Datenschutzparadoxons im digitalen Kontext bezieht sich auf den rationalen Fatalismus, demzufolge der Grad der fatalistischen Überzeugung über Technologien und Unternehmen die Wahrscheinlichkeit beeinflusst, dass Nutzer ihre Privatsphäre im Internet im Allgemeinen schützen (Xie et al. 2019).

Es gibt eine neue Strömung, die sich mit dem Datenschutzparadoxon im Zusammenhang mit KI im Gespräch befasst. Es hat sich gezeigt, dass die Art der Technologie und die Art der Informationen, die weitergegeben werden (Alter, Gewicht, Einkommen usw.), einen Unterschied machen. Intelligente Lautsprecher sind per se intrusiv, da sie die privateste soziale Sphäre ihrer Nutzer teilen. Darüber hinaus sind die GDPR-Datenschutzempfehlungen aufgrund der Konversationsschnittstelle (CUI) nur unzureichend umgesetzt und daher schwer zugänglich (Brüggemeier & Lalone, 2022). Ein weiterer Faktor, der konversationelle KI von anderen Kontexten unterscheidet, basiert auf dem CASA-Paradigma (Menschen neigen dazu, Computer als soziale Akteure zu behandeln) und Anthropomorphismus (die Tendenz von Menschen, Maschinen zu vermenschlichen). Intelligente Lautsprecher erzeugen wahrscheinlich eine soziale Präsenz, z. B. die Wahrnehmung einer partnerschaftlichen Beziehung, die nachweislich die unerwünschte Weitergabe persönlicher Informationen (unintended nudging) fördert. Diese besonderen Merkmale rechtfertigen eine Untersuchung der Kluft zwischen Einstellung und Verhalten speziell im Zusammenhang mit konversationeller KI, da Schlussfolgerungen aus Studien zu anderen Technologien nicht verallgemeinert werden können.

KRITIK AM PRIVACY PARADOX: WIDERSPRÜCHLICHE ERGEBNISSE UND METHODOLOGISCHE GRENZEN

Die Erforschung des Paradoxons der Privatsphäre hat zu widersprüchlichen Ergebnissen geführt. Die Forscher haben versucht, das Paradoxon der Privatsphäre in verschiedenen Zusammenhängen zu untersuchen. Ein Teil der Forschung konzentrierte sich darauf, Beweise dafür zu finden, dass das Verhalten der Menschen mit ihren Einstellungen und/

oder Absichten zum Schutz der Privatsphäre übereinstimmt, so dass kein paradoxes Verhalten gefunden wurde (z. B. Young und Quan-Haase, 2013 in sozialen Netzwerken; Wakefield, 2013 im elektronischen Handel). Die oben erwähnte Tatsache, dass das Verhalten in Bezug auf den Schutz der Privatsphäre in hohem Masse kontextabhängig ist, dass die Bedenken in Bezug auf den Schutz der Privatsphäre differenziert werden sollten und dass es davon abhängt, um welche Art von Informationen es geht, erschwert die Vergleichbarkeit und erklärt unterschiedliche Interpretationen der Studien. Insgesamt weisen die meisten Studien Mängel in der Methodik auf. Dies ist ein grosses Problem, da es die Interpretation und Verallgemeinerbarkeit in Frage stellt. Diesem Problem sollte höchste Priorität eingeräumt werden, um das Feld voranzubringen. Erhebungen und Experimente (meist auf der Grundlage von Zufallsstichproben) werden am häufigsten verwendet, was Fragen der Gültigkeit und Verallgemeinerbarkeit aufwirft. Das grösste Problem könnte darin bestehen, dass man sich auf Selbstauskünfte über Bedenken, Einstellungen und Absichten verlässt.

Wir glauben, dass die digitale Privatsphäre ein erstrebenswerter gesellschaftlicher Wert ist, was sich in der Einführung der Datenschutz-Grundverordnung widerspiegelt und von Ethikern oder sogenannten Datenschutzaktivisten gefördert wird. Es ist plausibel, dass individuelle Einstellungen und Bedenken zum Datenschutz aufgrund des Phänomens der sozialen Erwünschtheit weitgehend verzerrt sind. Diese Verzerrung besteht darin, dass die Befragten dazu neigen, Fragebögen generell positiv zu beantworten. Die Frage, die sich viele stellen, lautet daher: Wie viel Wert legen die Menschen wirklich auf die Privatsphäre? Könnte man die Einstellung zum Datenschutz wirklich messen, wäre dies der kürzeste Weg, um die Existenz des Datenschutzparadoxons zu widerlegen. Da dies, wie bei allen latenten Variablen, unmöglich ist, müssen die Forscher noch kreativer sein.

Kehr et al. (2015) vermuten, dass die “soziale Repräsentation von Privatsphäre von Laiennutzern noch nicht gebildet wird”. Im Einklang mit dieser Aussage fanden Barth et al. (2019) das paradoxe Verhalten in ihrem Experiment (unter Kontrolle von technischem Wissen, finanziellen Ressourcen und Bewusstsein für den Schutz der Privatsphäre), konnten jedoch

keine Darstellung des Schutzes der Privatsphäre in den Bewertungen der verwendeten Technologie durch die Befragten finden. Daraus schliessen sie, dass der Datenschutz nicht als wichtig eingestuft wird. Ausserdem sind Funktionalität, Design und wahrgenommener Nutzen wichtiger als der Datenschutz.

KÖNNEN WIR DAS PRIVACY PARADOX IN EIN TRUST PARADOX ÜBERTRAGEN?

Die EU hat einen Rahmen für vertrauenswürdige KI entwickelt, um eine verantwortliche Nutzung der künstlichen Intelligenz zu fördern. Sie weisen auf eine enge Beziehung zwischen Vertrauenswürdigkeit (als Eigenschaft von KI-Systemen) und der Einstellung zum Vertrauen hin, wobei der Datenschutz eine der Komponenten der Vertrauenswürdigkeit ist. Es ist plausibel, dass die Lehren aus dem Datenschutzparadoxon auf ein potenzielles Vertrauensparadoxon übertragen (oder modifiziert) werden können. Ein theoretischer Rahmen für ein Vertrauensparadoxon existiert nicht und könnte z. B. auf der Grundlage der Überprüfung des mehrdimensionalen Ansatzes für Datenschutz und intelligente Lautsprecher von Lutz und Newlands (2020) entwickelt werden. Unter Berücksichtigung der oben genannten Punkte ist ein Vertrauensparadoxon wahrscheinlich kontextabhängig (weshalb wir uns nur auf konversationelle KI konzentrieren), wobei die zu theoretisierenden Variablen aus der einschlägigen Vertrauensliteratur im Vergleich zur Datenschutzforschung stammen. Die Rolle von Wissen und finanziellen Ressourcen, wie sie im Artikel von Barth et al. im IOT -Kontext (2019) zu finden sind, könnte zum Beispiel sehr relevant sein. Was besonders wichtig sein könnte, ist die Integration der verschiedenen Akteure, die im Rahmen des Vertrauens eine Rolle spielen: Der eigentliche intelligente Lautsprecher (Alexa) gegenüber dem Unternehmen, das dahinter steht (Amazon), und anderen Dritten, die eine Rolle spielen können (z. B. intelligente Glühbirnen von Philips oder der NHS in Grossbritannien, der über Alexa Gesundheitsratschläge erteilt). Betrachtet man die Vielfalt der Erklärungen für paradoxes Verhalten im Datenschutzkontext, ist es plausibel, dass die Interpretationen für ein Vertrauensparadoxon ähnlich sind. Ausgehend von der Literatur zum Datenschutzparadoxon können die Interpretationen ähnlich sein, mit Ausnahme der sozialtheoretischen Interpretation, die wahrscheinlich nur in sozialen Netzwerken

vorkommt (z. B. Offenlegung persönlicher Informationen zur Aufrechterhaltung ihres Online-Lebens).

1. Die Menschen könnten eine Vertrauenskalkulation durchführen, indem sie eine Kosten-Nutzen-Analyse bei der Nutzung des Systems vornehmen.

2. Die Entscheidungsprozesse der Menschen werden durch kognitive Verzerrungen und Heuristiken beeinflusst.

3. Den Menschen fehlt es an bewusster Entscheidungsfindung aufgrund von begrenzter Rationalität und Informationsasymmetrie oder Informationsmangel.

4. Die Menschen geben auf und haben nicht das Gefühl, dass sie die Art und Weise, wie Daten gehandhabt werden, ändern können (siehe Turow et al., 2015).

Die Untersuchung eines potenziellen Vertrauensparadoxons stösst auf dieselben methodischen Herausforderungen wie oben erwähnt, wie z. B. die Selbstauskunft und die mangelnde Detailgenauigkeit, z. B. bei der Unterscheidung zwischen Vertrauensbedenken und Vertrauenshaltung. Vertrauen könnte sogar noch schwieriger sein, da es verschiedene Massstäbe gibt, und im Gegensatz zum Datenschutz gibt es kein offizielles Gesetz oder etwas Vergleichbares, das als Schwellenwert dafür dienen könnte, was "angemessenes" Vertrauen ist und was nicht. Die EU-Leitlinien für vertrauenswürdige KI können als theoretischer Hintergrund dienen. Es besteht die grosse Gefahr, dass sich die gleichen Probleme wiederholen, wenn die Methoden nicht kreativ gewählt werden und sich lediglich auf die Nutzungsabsicht konzentrieren. Andererseits kann die Untersuchung der Nutzungsabsicht gut als Pilotstudie dienen und könnte aufgrund der Neuartigkeit des Problems gut genug sein.

Es bleiben offene Fragen: Was ist das äquivalente Mass für das Vertrauen in die Leistung, den Prozess und den Zweck der konversationellen KI? Welche Art von Vertrauensbewusstsein ist relevant (im Vergleich zu dem relativ klaren Konzept des Datenschutzbewusstseins)? Wie lässt sich technisches Wissen im Kontext der dialogischen KI operationalisieren? Wie lassen sich die verschiedenen Akteure von Vertrauensbeziehungen integrieren? Und während

man die Angemessenheit individueller Datenschutzentscheidungen an bestehende Regeln und Vorschriften anpassen kann, was sind angemessene individuelle Vertrauensentscheidungen?

FAZIT

Wir glauben, dass die respektable Arbeit der Datenschutzforscher eine wertvolle Gelegenheit bietet, Vertrauen in der Mensch-KI-Interaktion besser zu verstehen, genauer gesagt, Vertrauen in konversationelle KI, die unser Thema ist. Die Forschung zum Datenschutz-Paradoxon hat widersprüchliche Ergebnisse hervorgebracht und weist methodische Mängel auf, so dass es ein weit offenes Thema mit einem Mangel an Wissen speziell im Bereich der konversationellen KI bleibt (Kokalakis, 2015; Barth et al., 2019; Sun et al., 2020; Lutz & Newlands, 2021). Diese Herausforderungen ebnen jedoch auch den Weg für innovative Forschungsagenden. Es ist plausibel, dass in der Mensch-KI-Interaktion ein Vertrauensparadoxon zu finden ist, ähnlich dem Datenschutzparadoxon, einschliesslich der beobachteten schwierigen Herausforderungen. Vertrauen ist ein komplexes Konstrukt mit kognitiven und emotionalen Elementen, und obwohl wir erwarten, viele Ähnlichkeiten mit dem Privatsphärenparadoxon zu finden, theoretisieren wir auch, dass es Unterschiede zum Vertrauensparadoxon gibt, in der Hoffnung, nützliche Empfehlungen für Forschung und Praxis zu entwickeln.

Schliesslich müssen die Ergebnisse bezüglich der Einstellungs-Verhaltens-Kluft von den daraus resultierenden Erwartungen oder der Wahrnehmung, die Individuen in Bezug darauf haben, wie die Anbieter dieser Dienstleistungen handeln sollten, getrennt werden. Martin (2016) stellt fest, dass Personen auch nach der Offenlegung von Informationen starke Erwartungen an die Privatsphäre haben, was bedeutet, dass ihre Bereitschaft, E-Commerce-, Social-Media-, Smartphone-, Browser-, Smart-Assistenten- oder anderen Dienstanbietern Zugang zu privaten Daten zu gewähren, nichts an ihrer Einstellung zur Privatsphäre ändert.

WISSEN IST IMMER EIN VORSPRUNG

Bildquelle:https://pixabay.com/images/id-7567437/

SCIP MONTHLY SECURITY SUMMARY

IMPRESSUM

ÜBER DEN SMSS

Das *scip Monthly Security Summary* erscheint monatlich und ist kostenlos.

Anmeldung: smss-subscribe@scip.ch
Abmeldung: smss-unsubscribe@scip.ch

Informationen zum Datenschutz.

Verantwortlich für diese Ausgabe:
Marc Ruef

Eine Haftung für die Richtigkeit der Veröffentlichungen kann trotz sorgfältiger Prüfung durch die Redaktion des Herausgebers, den Redaktoren und Autoren nicht übernommen werden. Die geltenden gesetzlichen und postalischen Bestimmungen bei Erwerb, Errichtung und Inbetriebnahme von elektronischen Geräten sowie Sende- und Empfangseinrichtungen sind zu beachten.

ÜBER SCIP AG

Wir überzeugen durch unsere Leistungen. Die scip AG wurde im Jahr 2002 gegründet. Innovation, Nachhaltigkeit, Transparenz und Freude am Themengebiet sind unsere treibenden Faktoren. Dank der vollständigen Eigenfinanzierung sehen wir uns in der sehr komfortablen Lage, vollumfänglich herstellerunabhängig und neutral agieren zu können und setzen dies auch gewissenhaft um. Durch die Fokussierung auf den Bereich Information Security und die stetige Weiterbildung vermögen unsere Mitarbeiter mit hochspezialisiertem Expertenwissen aufzuwarten.

Weder Unternehmen noch Redaktion erwähnen Namen von Personen und Firmen sowie Marken von fremden Produkten zu Werbezwecken. Werbung wird explizit als solche gekennzeichnet.

scip AG
Badenerstrasse 623
8048 Zürich
Schweiz

+41 44 404 13 13
www.scip.ch