

Version 1.0



Erfüllung der Leitlinien des Hiroshima AI Process (HAIP)

INHALTSVERZEICHNIS

EINLEITUNG	3
ABSCHNITT 1 - RISIKOBESTIMMUNG UND -BEWERTUNG	3
ABSCHNITT 2 - RISIKOMANAGEMENT UND INFORMATIONSSICHERHEIT.....	6
ABSCHNITT 3 - TRANSPARENZBERICHTERSTATTUNG ÜBER FORTSCHRITTLICHE KI-SYSTEME ...	8
ABSCHNITT 4 - ORGANISATORISCHE GOVERNANCE, VORFALLMANAGEMENT UND TRANSPARENZ	9
ABSCHNITT 5 - MECHANISMEN ZUR AUTHENTIFIZIERUNG UND HERKUNFT VON INHALTEN..	10
ABSCHNITT 6 - FORSCHUNG UND INVESTITIONEN ZUR FÖRDERUNG DER KI-SICHERHEIT UND ZUR MINDERUNG GESELLSCHAFTLICHER RISIKEN	11
ABSCHNITT 7 - FÖRDERUNG MENSCHLICHER UND GLOBALER INTERESSEN.....	12

EINLEITUNG

Der Hiroshima AI Process (HAIP) ist eine Initiative der G7 aus 2023, die darauf abzielt, internationale Leitlinien und einen freiwilligen Verhaltenskodex für den sicheren, vertrauenswürdigen Einsatz von (generativer) KI zu entwickeln.

Für Unternehmen ist eine Teilnahme bzw. Orientierung daran sinnvoll, weil sie sich frühzeitig an international anerkannten Standards für sichere und verantwortungsvolle KI ausrichten können, was Vertrauen bei Kunden und Partnern stärkt und regulatorische Risiken reduziert. Gleichzeitig ermöglicht der Prozess den Zugang zu Best Practices, Austausch mit Regulatoren und Mitgestaltung zukünftiger Governance-Regeln.

Buildsimple orientiert sich am Hiroshima AI Process und legt dieses in den nachfolgenden Kapiteln offen.

ABSCHNITT 1 - RISIKOBESTIMMUNG UND -BEWERTUNG

a. Wie definiert und/oder klassifiziert Ihre Organisation verschiedene Arten von Risiken im Zusammenhang mit KI, wie etwa unangemessene Risiken?

Buildsimple ist sowohl Entwickler als auch Anbieter von KI. Eigene KI-Modelle, frei verfügbare Foundation-Modelle und auch via AWS Bedrock bezogene LLM-Modelle werden Kunden über die Cloudplattform "Buildsimple" zur Benutzung zur Verfügung gestellt. Dieser Bericht konzentriert sich auf die bereitgestellten selbstentwickelten KI-Modelle sowie auf die Foundation-Modelle ("Buildsimple-KI"), beide können sowohl durch die Kunden selbst als auch durch Buildsimple (weiter-)trainiert werden.

Die Risikoklassifizierung der Buildsimple-KI erfolgt über eine Matrix, bei der Eintrittswahrscheinlichkeit und Auswirkung eines Risikos mit Punkten belegt und addiert werden. Die resultierenden Werte werden in einer Scorecard in niedrige Risiken, mittlere Risiken und hohe Risiken eingestuft, wobei niedrige Risiken akzeptiert werden können, mittlere mindestens gemanaged werden und hohe Risiken zwingend vermieden werden müssen, bevor ein KI-Modell eingesetzt werden kann.

Risiken können Auswirkungen auf die Integrität, Verfügbarkeit, Vertraulichkeit oder die Authentizität von Daten haben. Die Risikobewertung bei Buildsimple erfolgt nach Risiko-Typ (operativ, Reputation, Compliance, Sicherheit, finanziell), wobei einem Risiko mehrere Risikotypen zugeordnet werden können.

Bei der Ermittlung potenzieller Risiken wurden gängige Bedrohungen u.a. aus den OWASP Top 10 und typische Angriffstaktiken u.a. aus dem MITRE ATT&CK-Framework berücksichtigt, ebenso findet das AWS Well-Architected Framework Berücksichtigung bei der Entwicklung von KI-Lösungen.

Die bei Buildsimple eingesetzte KI wird im Sinne der europäischen Verordnung über künstliche Intelligenz (KI-Verordnung) als KI mit minimalem Risiko eingestuft, da es sich um keine KI handelt, die eigenständig Entscheidungen trifft, sondern lediglich Informationen aus Dokumenten extrahiert, welche dort bereits vorhanden sind. Ebenso generiert die Buildsimple-KI keine neuen Informationen, weshalb bei ihrem Einsatz keine unangemessenen Risiken erwartet werden können.

b. Welche Praktiken verwendet Ihre Organisation, um Risiken wie Schwachstellen, Vorfälle, neue Risiken und Missbrauch im gesamten KI-Lebenszyklus zu identifizieren und zu bewerten?

Eine Risikobewertung neuer KI-Modelle findet entwicklungsbegleitend statt. Vor der Auslieferung neuer KI-Modelle werden diese automatisch auf Schwachstellen geprüft. Zum Einsatz kommen u.a. Tools, die in die Buildpipeline integriert sind und die genutzten Komponenten mit Schwachstellen abgleichen, die in der National Vulnerability Database (NVD) gemeldet sind. Software Bill Of Material-Files (SBOMs) werden automatisch erzeugt und zyklisch zusätzlich manuell geprüft. Weiterhin kommen Tools wie AWS Guard Duty (intelligente Bedrohungserkennung, kontinuierliche Überwachung, automatisierte Sicherheitswarnungen), IAM Access Analyzer (kontinuierliche, automatische Überprüfung von Berechtigungen), AWS Security HUB (kontinuierliche Überwachung der Sicherheitslage der genutzten Komponenten), AWS Trusted Advisor (Analyse der genutzten Ressourcen bezüglich Sicherheit, Leistung, Ausfallsicherheit) und Amazon Inspector (Schwachstellen-Management-Service, der Workloads automatisch erkennt und sie kontinuierlich auf Software-Schwachstellen und unbeabsichtigte Netzwerkbedrohungen überprüft) zum Einsatz.

Es finden monatlich sowie nach wesentlichen Änderungen automatisierte PEN-Tests gegen die öffentlichen Schnittstellen statt, über die mit den KI-Modellen interagiert werden kann. Jährlich werden manuelle PEN-Tests durch den TÜV Süd durchgeführt.

Bei Auffälligkeiten werden diese im Falle von kritischen Risiken sofort behoben, ansonsten bedarfsgerecht im Sprint eingeplant.

c. Beschreiben Sie, wie Ihre Organisation Tests (z. B. Red-Teaming) durchführt, um die Eignung des Modells/Systems für den Übergang über die Entwicklungsphase hinaus zu bewerten?

Buildsimple verfügt über ein eigenes Trainerteam, welches Kunden dabei unterstützt, die vom Kunden selbst angelernten KI-Modelle auf ihre Eignung zu prüfen. Über Dashboards lassen sich Trainingsdokumente identifizieren, die falsch trainiert wurden oder zum Training weniger gut geeignet sind. Eine Reduktion der Trainingsdaten um diese Dokumente hat einen unmittelbaren positiven Einfluss auf die Qualität der Buildsimple-KI.

Die Buildsimple-Plattform wird einem jährlichen PEN-Tests durch den TÜV Süd unterzogen mit dem Ziel, die Sicherheit der Modelle sicherzustellen.

d. Verwendet Ihre Organisation Vorfallberichte, einschließlich von anderen Organisationen geteilte Berichte, um Risiken zu identifizieren?

Ja.

e. Werden quantitative und/oder qualitative Risikobewertungsmetriken verwendet und falls ja, mit welchen Vorbehalten? Stellt Ihre Organisation Mechanismen zur Meldung von Schwachstellen und Vorfällen einer vielfältigen Gruppe von Interessengruppen zur Verfügung? Hat Ihre Organisation Anreizprogramme für die verantwortungsvolle Offenlegung von Risiken, Vorfällen und Schwachstellen?

Da es sich bei der Buildsimple-KI nicht um eine generative KI bzw. ein Large Language Modell handelt, liegt aus Kundensicht das Risiko vor allem in der Qualität der erzielten Ergebnisse. Die Qualität ist für den Kunden transparent dargestellt, er kann selbst Schwellwerte definieren, bis zu denen Dokumente

"dunkel" verarbeitet werden, und ab wann ein "Human-in-the-loop" zwingend notwendig ist, um das von der KI erzielte Ergebnis zu überprüfen.

Schwachstellen und Vorfälle können von Kunden, Partnern und Mitarbeitern über den Supportweg gemeldet werden. Bei interner Kenntnisnahme einer Schwachstelle, die Auswirkungen auf Kunden haben kann, würden die Kunden proaktiv durch Buildsimple informiert werden.

f. Wird externe unabhängige Expertise für die Identifizierung, Bewertung und Bewertung von Risiken genutzt und wenn ja, wie? Verfügt Ihre Organisation über Mechanismen, um Berichte über Risiken, Vorfälle oder Schwachstellen von Dritten zu erhalten?

Im Rahmen des AWS Foundational Technical Reviews wird nachgewiesen, dass die Buildsimple-KI über ein durch AWS gefordertes Mindestmaß an Risikomitigationsmaßnahmen verfügt, welches im AWS Well-Architected Framework festgehalten ist. Weiterhin wird eine Risikoüberprüfung durch mehrfache unterjährige Audits dokumentiert durchgeführt. Diese Audits werden u.a. durch Wirtschaftsprüfer, Information-Risk-Manager, externe Datenschutzbeauftragte sowie ISO27001 Lead-Auditoren statt.

Berichte über Risiken, Vorfälle und Schwachstellen Dritter werden Buildsimple automatisch durch AWS übermittelt. Während der Entwicklung werden eingesetzte Komponenten u.a. gegen die National Vulnerability Database abgeglichen, um Risiken erkennen zu können, bevor sie ins Produkt gelangen.

g. Trägt Ihre Organisation zur Entwicklung von und/oder zur Anwendung internationaler technischer Standards oder Best Practices zur Identifizierung, Bewertung und Bewertung von Risiken bei?

Buildsimple ist Mitglied im Wirtschaftsrat und der Bundesfachkommission für Künstliche Intelligenz und arbeitet in dieser Funktion auch aktiv an der Gesetzgebung mit. Buildsimple war u.a. direkt am Entwurf des AI-Acts (KI-Verordnung) der europäischen Union beteiligt, mit dessen Hilfe Risiken im Umgang mit KI identifiziert und bewertet werden können.

h. Wie arbeitet Ihre Organisation mit relevanten Interessengruppen aus verschiedenen Sektoren zusammen, um Risikominderungsmaßnahmen zur Bewältigung von Risiken, insbesondere systemischen Risiken, zu bewerten und einzusetzen?

Anforderungen für Risikomitigationsmaßnahmen bekommt Buildsimple vor allem von BaFin-regulierten Kunden, hier findet vor allem mit potenziellen Neukunden ein regelmäßiger Austausch statt. Auch aus anderen Kundenkreisen wie z.B. der Luftfahrtindustrie werden Risikoanforderungen an Buildsimple gestellt, um interne und externe Compliance-Anforderungen zu erfüllen.

Buildsimple ist Mitglied im Wirtschaftsrat und der Bundesfachkommission für Künstliche Intelligenz, Gründungsmitglied der AWS Usergroup Münsterland und in vielen weiteren Netzwerken aktiv, innerhalb derer es einen regen Austausch zu möglichen Bedrohungslagen gibt. Beispielhaft genannt seien hier die Versicherungsforen Leipzig, die Initiative Cloud Services Made in Germany, die Allianz für Cyber-Sicherheit sowie die aktive Teilnahme im Netzwerk LinkedIn, um sich dort mit Experten auszutauschen.

ABSCHNITT 2 - RISIKOMANAGEMENT UND INFORMATIONSSICHERHEIT

a. Welche Schritte unternimmt Ihre Organisation, um Risiken und Verwundbarkeiten im gesamten KI-Lebenszyklus zu adressieren?

Buildsimple hat einen umfassenden Ansatz zur Risikominderung implementiert, der den gesamten KI-Lebenszyklus abdeckt – von der Entwicklung über die Einführung bis hin zur fortlaufenden Überwachung. Das Rahmenwerk umfasst technische Schutzmaßnahmen, Governance-Kontrollen und kontinuierliche Evaluationsprozesse, die darauf ausgelegt sind, Schwachstellen zu identifizieren und zu beheben und gleichzeitig eine sichere und verantwortungsvolle KI-Entwicklung zu fördern. Näheres hierzu wurde bereits in Abschnitt 1b beschrieben.

b. Wie beeinflussen Testmaßnahmen Maßnahmen zur Bewältigung identifizierter Risiken?

Die bei Buildsimple eingesetzten umfangreichen Testmaßnahmen umfassen sowohl manuelle als auch automatisierte Tests, die im Rahmen der Entwicklungspipeline ausgelöst werden.

- Codeanalysetests warnen u.a. vor veralteten Bibliotheken, Bugs und Vulnerabilities
- Funktionstests sind in den genutzten Frameworks integriert
- PACT-Tests stellen sicher, dass Nachrichten zwischen Integrationspunkten einem gemeinsamen Verständnis entsprechen
- Integrationstests testen das Zusammenspiel der einzelnen Services
- Selenium-Tests testen die Funktionalität der Benutzeroberfläche

Deployments erfolgen grundsätzlich zuerst in eine Entwicklungsumgebung. Erst nach durchlaufenen erfolgreichen Tests erfolgt das Deployment in die Produktivumgebung.

Die durchgeführten Tests liefern direkt umsetzbare Erkenntnisse bezüglich potenzieller Risiken und sind somit geeignet, identifizierte Risiken zu bewältigen, bevor diese in das Produktivsystem gelangen.

c. Wann finden Tests in sicheren Umgebungen statt, falls überhaupt, und wenn ja, wie?

Vor dem Release neuer KI-Modelle werden diese in einer Testumgebung getestet. Es handelt sich hierbei um einen eigenen AWS-Account ohne Bezug zum Produktivsystem. Der Zugang zur Testumgebung ist streng limitiert und nur ausgewählten Personen in unterschiedlichen Rollen möglich, er wird durch ein umfassendes Sicherheitsframework gesteuert.

Die Tests umfassen Maßnahmen, die in 1b näher erläutert wurden.

d. Wie fördert Ihre Organisation die Datenqualität und mindert Risiken schädlicher Verzerrungen, auch bei Schulungen und Datenerhebungsprozessen?

Kunden haben bei der Buildsimple-KI die Möglichkeit, ein durch Buildsimple mit handverlesenen Dokumenten vortrainiertes Modell zu nutzen, ihr Modell mit eigenen Dokumenten selbst zu trainieren oder auf ein Foundation Modell aufzusetzen, welches durch Buildsimple und/oder einen externen Anbieter vortrainiert wurde. Der Kunde hat jederzeit die Möglichkeit einzusehen, mit welchen

Dokumenten er sein KI-Modell trainiert hat. Dokumente, die das Verarbeitungsergebnis negativ beeinflussen, lassen sich identifizieren und aus der Trainingsmenge entfernen.

e. Wie schützt Ihre Organisation geistiges Eigentum, einschließlich urheberrechtlich geschützter Inhalte?

Durch Buildsimple vortrainierte KI-Modelle werden ausschließlich mit handverlesenen Daten trainiert, für die eine Einwilligung des Rechteinhabers vorliegt. Buildsimple-Kunden haben die Möglichkeit, Dokumente ihrer eigenen Trainingsmenge zum Training öffentlich verfügbarer Buildsimple-KI-Modelle zur Verfügung zu stellen oder sogar ihr trainiertes KI-Modell mit anderen Kunden zu teilen, eine Pflicht hierfür besteht jedoch explizit nicht.

f. Wie schützt Ihre Organisation die Privatsphäre? Wie schützt Ihr Unternehmen davor, dass Systeme vertrauliche oder sensible Daten preisgeben?

Wurde ein Dokument zum Training einer KI-Lösung verwendet, kann dieses in der Trainingsmenge durch den Kunden eingesehen werden. Dieses ist automatisch von jeglichem Personenbezug befreit und anonymisiert worden, wodurch im Nachhinein keine personenbezogenen Daten mehr aus den Trainingsdaten eingesehen werden können. In bildbasierter Form werden auf dem Originaldokument alle identifizierten personenbezogenen Daten mit einem weißen Balken überdeckt, auf dem ein neuer Text enthalten ist, welcher aus einer Zufallsfolge von Buchstaben und Ziffern besteht. Dabei wird jeder vorherige Buchstabe wieder durch einen Buchstaben und jede Ziffer durch eine Ziffer ersetzt. In textbasierter Form werden ebenfalls alle personenbezogenen Daten identifiziert und durch eine Zufallsfolge von Buchstaben und Ziffern nach gleichem Prinzip ersetzt. Durch diese Ersetzung mit zufälligen Buchstaben und Ziffern ist der Originalwert auf der Bilddatei und in textbasierter Form dauerhaft nicht mehr rekonstruierbar.

g. Wie implementiert Ihre Organisation KI-spezifische Informationssicherheitspraktiken im Bereich operativer und Cyber-/physischer Sicherheit?

Cybersicherheitsrisiken werden kontinuierlich bewertet die abgeleiteten Maßnahmen werden extern auditiert. Bei der Bewertung wird u.a. das Common Vulnerability Scoring System (CVSS) genutzt, um den Schweregrad eines Risikos abschätzen zu können. Die getroffenen Sicherheitsmaßnahmen werden entsprechend der Bedrohungslage fortlaufend verfeinert. Richtlinien werden wo immer möglich in Software abgebildet, um z.B. harte Deploymentschranken durchzusetzen oder Zugriffe durch böswillige Akteure zu verhindern.

Der Zugang zu schützenswerten Daten ist gemäß Berechtigungskonzept nur einem ausgewählten Personenkreis möglich. Die Umgebung ist in AWS aufgebaut und wurde gemeinsam mit AWS Solution Architekten gemäß dem AWS Well-Architektured-Framework erstellt und durch AWS im Rahmen des AWS Foundational Technical Reviews überprüft.

Das Schwachstellenmanagement erfolgt kontinuierlich und automatisiert, zusätzlich findet ein wöchentlicher manueller Security-Audit statt. Informationen über Risiken und Schwachstellen stammen z.T. aus Datenbanken wie der National Vulnerability Database (NVD), aber auch aus anderen Quellen wie der Allianz für Cyber-Sicherheit oder dem BSI. Identifizierte Schwachstellen werden auf Eintrittswahrscheinlichkeit und Auswirkung hin gemäß der dokumentierten Risikobewertung hin überprüft und entsprechend behoben.

Insider-Bedrohungen wird u.a. durch ein Berechtigungskonzept, Logging und entsprechende Alarmierung begegnet, weiterhin ist an Schlüsselstellen wie Codeänderungen, Deployments, Änderungen von Berechtigungen ein Vier-Augen-Prinzip etabliert. Backups erfolgen in separate Umgebungen, zu denen Entwickler keinen Zugriff haben.

h. Wie geht Ihre Organisation mit Schwachstellen, Vorfällen und neuen Risiken um?

Identifizierte Risiken werden in einer GRC-Plattform dokumentiert, den verantwortlichen Eigentümern zugewiesen und bis zur Lösung nachverfolgt. Automatische Erkennungssysteme und Inhaltsfilter sind geeignet, um bisher unbekannten Risiken zu begegnen. Für die Vorfallsbewältigung sind Verfahrensanweisungen, Notfallpläne und Meldewege implementiert, die mindestens jährlich in Form von Notfallübungen getestet und weiterentwickelt werden.

ABSCHNITT 3 - TRANSPARENZBERICHTERSTATTUNG ÜBER FORTSCHRITTLICHE KI-SYSTEME

a. Veröffentlicht Ihre Organisation klare und verständliche Berichte und/oder technische Dokumentationen zu den Fähigkeiten, Einschränkungen und Bereichen der angemessenen und unangemessenen Nutzung fortschrittlicher KI-Systeme?

Ja. Die Berichte unterliegen einem jährlichen Review und werden ggfs. unterjährig aktualisiert und Kunden auf Anfrage zur Verfügung gestellt. Bedeutende neue Veröffentlichungen werden allen betroffenen Kunden proaktiv zur Verfügung gestellt. Bereiche der angemessenen sowie der unangemessenen Nutzung der Buildsimple KI-Systeme sind vertraglich mit den Kunden vereinbart. Informationen zur Bedrohungsanalyse und mitigierende Maßnahmen werden Kunden auf Wunsch zur Verfügung gestellt.

b. Wie teilt Ihre Organisation Informationen mit einer vielfältigen Gruppe von Interessengruppen (andere Organisationen, Regierungen, Zivilgesellschaft und Wissenschaft usw.) über das Ergebnis von Risikobewertungen und Auswirkungen im Zusammenhang mit einem fortschrittlichen KI-System?

Informationen zur Bedrohungsanalyse und mitigierende Maßnahmen werden Kunden und Partnern auf Wunsch zur Verfügung gestellt. Buildsimple ist über viele verschiedene Kanäle mit unterschiedlichen Interessengruppen (IHK, Polizei, BSI, Usergroups, Interessenverbände) vernetzt, innerhalb derer ein reger Austausch zu KI-relevanten Themen stattfindet.

c. Gibt Ihre Organisation Datenschutzrichtlinien zur Nutzung personenbezogener Daten, Nutzereingaben und/oder den Ergebnissen fortschrittlicher KI-Systeme bekannt?

Ja, diese sind über die Buildsimple-Nutzungsbedingungen Bestandteil des Vertragswerkes mit Kunden.

d. Stellt Ihre Organisation Informationen über die Datenquellen bereit, die für die Ausbildung fortschrittlicher KI-Systeme verwendet werden, soweit angemessen, einschließlich Informationen zur Herkunft von Datenannotation und -anreicherung?

Ja, diese Informationen stellt Buildsimple Kunden zur Verfügung.

e. Zeigt Ihre Organisation Transparenz in Bezug auf fortschrittliche KI-Systeme auf andere Weise?

In Bezug auf Fortschrittliche KI-Systeme verweist Buildsimple auf die Berichte der jeweiligen Anbieter selbst.

ABSCHNITT 4 - ORGANISATORISCHE GOVERNANCE, VORFALLMANAGEMENT UND TRANSPARENZ

a. Wie wurde das KI-Risikomanagement in Ihren Governance-Rahmen Ihrer Organisation integriert? Wann und unter welchen Umständen werden Richtlinien aktualisiert?

Das Risikomanagement ist in die gesamte Organisationsstruktur eingebettet und wird kontinuierlich extern begleitet, überprüft und angepasst. In mehreren Governance-Richtlinien sind klare Leitplanken festgelegt, innerhalb derer die Weiterentwicklung der Buildsimple-KI stattfindet. Die Überarbeitung der Richtlinien findet mindestens jährlich dokumentiert und versioniert statt. Sämtliche Richtlinien sind transparent abgelegt und müssen von jedem Mitarbeiter jährlich bestätigt werden.

b. Sind relevante Mitarbeiter in den Governance-Richtlinien und Risikomanagementpraktiken Ihrer Organisation geschult? Wenn ja, wie?

Bei Buildsimple finden durch das Unternehmen jährliche Mitarbeiterunterweisungen zu den Themen Datenschutz, Informationssicherheit und Compliance statt. Zusätzlich werden alle Buildsimple-Entwickler jährlich dokumentiert auf die aktuellen Governance-Richtlinien und Risikomanagementpraktiken geschult. Begleitend findet das ganze Jahr durchgehend ein Awarenessstraining als E-Learning statt, welches regelmäßig auf die aktuelle Bedrohungslage angepasst wird.

c. Kommuniziert Ihre Organisation ihre Risikomanagementrichtlinien und -praktiken mit den Nutzern und/oder der Öffentlichkeit? Wenn ja, wie?

Buildsimple bietet Kunden diverse Dokumente an, in denen die Risikomanagementrichtlinien und -praktiken detailliert beschrieben sind. Hierbei kann es sich um Bestandteile der Kundenverträge handeln, um Informationen, um ein Transfer-Impact-Assessment durchzuführen, um Konformitätserklärungen oder um Details, wie bei Buildsimple gesetzliche Anforderungen konkret umgesetzt werden.

d. Werden Maßnahmen zur Behandlung gemeldeter Vorfälle intern dokumentiert und gepflegt? Wenn ja, wie?

Ein Modul des von Buildsimple genutzten ISMS-Tools adressiert das Thema Vorfälle. Sie können direkt im Tool erstellt, nachverfolgt und auch an Kunden kommuniziert werden. Das Vorfallsmanagement ist Bestandteil jährlicher Audits und wird extern überprüft.

e. Wie teilt Ihre Organisation relevante Informationen über Schwachstellen, Vorfälle, neue Risiken und Missbrauch mit anderen?

Intern werden diese Informationen mit den relevanten Sicherheits-, Datenschutz-, Compliance- und Entwicklerteams sowie dem Management geteilt.

Mit Kunden und Partnern teilt Buildsimple diese Informationen über direkte Kommunikationskanäle wie den Kundensupport oder das Partnermanagement.

Vorfälle, die eine Beeinträchtigung der KI-Verarbeitung zur Folge haben, können durch Kunden und Partner im Verfügbarkeitsdashboard innerhalb der Buildsimple-Applikation direkt eingesehen werden.

f. Teilt Ihre Organisation Informationen entsprechend angemessen mit relevanten anderen Interessengruppen zu Vorfällen mit fortgeschrittenen KI-Systemen? Wenn ja, wie? Teilt und meldet Ihre Organisation vorfallbezogene Informationen öffentlich?

Ja. Buildsimple teilt KI-bezogene Vorfälle direkt mit Stakeholdern (Unternehmenskunden und Partnern) und veröffentlicht diese innerhalb der Buildsimple-Plattform. Eine Veröffentlichung erfolgt, sofern diese regulatorisch oder rechtlich erforderlich ist. Meldewege an Behörden bei gesetzlichen Anforderungen sind vorhanden und erprobt.

g. Wie teilt Ihre Organisation Forschung und Best Practices zur Risikobewältigung oder -bewältigung?

Ein regelmäßiger Austausch mit externen Datenschutzbeauftragten, externen Informationssicherheitsbeauftragten, externen Informationssicherheitsberatern sowie den Informationssicherheitskoordinatoren anderer Geschäftsbereiche ist seit Jahren etabliert. Im Monatsrhythmus findet ein Austausch der Information Security Officer sowie der Governance-Verantwortlichen innerhalb der Unternehmensgruppe statt.

h. Verwendet Ihre Organisation internationale technische Standards oder Best Practices für KI-Risikomanagement und Governance-Richtlinien?

Ja. Das Risikomanagement erfolgt auf Basis der ISO/IEC 27005, ein Informationssicherheitsmanagementsystem nach ISO/IEC 27001:2022 und ISO/IEC 27002:2022 ist im Einsatz. Weiterhin erfüllt Buildsimple die an seine Kunden gestellten Anforderungen wie BAIT, VAIT, MaGo und MaRisk und hat die Anforderungen des AI Acts, des Cyber Resilience Acts, des Digital Operational Resilience Acts sowie die der NIS-2-Richtlinie umgesetzt.

ABSCHNITT 5 - MECHANISMEN ZUR AUTHENTIFIZIERUNG UND HERKUNFT VON INHALTEN

a. Welche Mechanismen, falls überhaupt, hat Ihre Organisation eingerichtet, damit Nutzer, wo möglich und angemessen, wissen können, wann sie mit einem fortschrittlichen KI-System interagieren, das von Ihrer Organisation entwickelt wurde?

Fortschrittliche KI-Modelle sind innerhalb der Buildsimple-Plattform gesondert gekennzeichnet und somit für den Nutzer sofort erkennbar.

b. Verwendet Ihre Organisation Mechanismen zur Erkennung von Inhaltsherkunft, Kennzeichnung oder Wassermarkierung, die es den Nutzern ermöglichen, Inhalte zu identifizieren, die von fortschrittlichen KI-Systemen generiert werden? Wenn ja, wie? Verwendet Ihre Organisation

internationale technische Standards oder Best Practices bei der Entwicklung oder Implementierung von Inhaltsherkunft?

Die Buildsimple-KI selbst ist nicht in der Lage, Inhalte zu generieren, daher kann keine Kennzeichnung von Inhalten stattfinden.

ABSCHNITT 6 - FORSCHUNG UND INVESTITIONEN ZUR FÖRDERUNG DER KI-SICHERHEIT UND ZUR MINDERUNG GESELLSCHAFTLICHER RISIKEN

a. Wie fördert Ihre Organisation Forschung und Investitionen in den folgenden Bereichen: Sicherheit, Sicherheit, Voreingenommenheit und Desinformation, Fairness, Erklärbarkeit und Interpretierbarkeit, Transparenz, Robustheit und/oder Vertrauenswürdigkeit fortschrittlicher KI-Systeme?

Buildsimple fördert Forschung und Investitionen in den o.g. Bereichen durch die Mitgliedschaft in der Bundesfachkommission für Künstliche Intelligenz. Die Bundesfachkommission Künstliche Intelligenz und Wertschöpfung 4.0 im Wirtschaftsrat bringt Entscheider aus Wirtschaft, Technologie und Politik zusammen und hat sich die folgenden Ziele gesetzt:

- Stärkung des KI-Standorts Deutschland und Europa
- Praxisnahe Umsetzung europäischer KI-Regulierung
- Ausbau leistungsfähiger KI- und Cloud-Infrastrukturen
- Verbesserung des Zugangs zu Daten für KI-Anwendungen
- Förderung von KI-Fachkräften und digitalen Kompetenzen

Weiterhin fördert Buildsimple Studenten, die im Rahmen von Bachelor- und Masterarbeiten im KI-Umfeld forschen und betreut diese bei ihren Abschlussarbeiten.

b. Wie arbeitet Ihre Organisation an Forschung zusammen und investiert sie darin, um den Stand der Authentifizierung und Herkunft von Inhalten zu verbessern?

Siehe Abschnitt 6a.

c. Beteiligt sich Ihre Organisation an Projekten, Kooperationen und Investitionen in Forschung, die die Weiterentwicklung von KI-Sicherheit, -Schutz und -vertrauenswürdigkeit sowie von Werkzeugen zur Risikobewertung und -minderung unterstützen?

Siehe Abschnitt 6a.

d. Welche Forschung oder Investitionen verfolgt Ihre Organisation, um sozioökonomische und/oder ökologische Risiken durch KI zu minimieren?

Buildsimple läuft vollständig in AWS-Rechenzentren. AWS ist stets bestrebt, die Rechenleistung zu steigern und gleichzeitig den Energieverbrauch zu verringern. Der PUE-Wert (Power Usage Effectiveness) von AWS liegt bei 1,15 (Branchendurchschnitt 1,25). Wie AWS im Jahr 2022 bekannt gab, soll das Unternehmen bis 2030 "wasserpositiv" sein, also den Gemeinden und der Umwelt mehr Wasser zurückgeben, als der Betrieb der AWS-Rechenzentren verbraucht.

Auch aufgrund des weiteren Engagements von AWS in Bereichen der Nachhaltigkeit fiel die Wahl auf AWS als Hosting-Partner.

ABSCHNITT 7 - FÖRDERUNG MENSCHLICHER UND GLOBALER INTERESSEN

a. Welche Forschung oder Investitionen verfolgt Ihre Organisation, um den sozioökonomischen und ökologischen Nutzen von KI zu maximieren? Bitte geben Sie Beispiele.

Buildsimple läuft vollständig in AWS-Rechenzentren. AWS unterstützt u.a. Start-ups in den Bereichen Klimatechnologie und saubere Energie. AWS hat FlowMS entwickelt – eine KI-gestützte Lösung, die die Energie- und Wassereffizienz in Gebäuden steigert, die Dekarbonisierung von Unternehmen beschleunigt und die betriebliche Effizienz verbessert.

Die "AWS Public Sector Generative AI Impact Initiative" finanziert Projekte, die Behörden dabei unterstützen, ihre Bürgerdienste zu modernisieren. Über die "AWS Health Equity Initiative" werden KI-Projekte durch AWS unterstützt, die Ungleichheiten in der Gesundheitsversorgung abbauen, die Telemedizin ausbauen, die Diagnostik verbessern und den Zugang für unterversorgte Bevölkerungsgruppen erleichtern.

"AWS Disaster Response" kooperiert mit Einsatzkräften und nutzt KI sowie Geodatenanalysen für eine schnelle Schadensbewertung, die Erstellung von Hochwasserkarten und die Ressourcenverteilung. Die "AWS Education Equity Initiative" unterstützt Organisationen dabei, digitale Lernlösungen für unterversorgte Lernende auf der ganzen Welt zu entwickeln. Amazon unterstützt den AI Hub for Sustainable Development, der darauf ausgelegt ist, die Kompetenzentwicklung sowie die Einführung von KI zu beschleunigen, um das industrielle Wachstum in Afrika voranzutreiben.

In Partnerschaft mit der UNESCO und dem International Research Centre on Artificial Intelligence (IRCAI) finanziert Amazon mit dem "Compute for Climate Fellowship" Proof-of-Concept-Lösungen für den Klimaschutz, die fortschrittliches Cloud-Computing und generative KI nutzen, um der Klimakrise zu begegnen.

b. Unterstützt Ihre Organisation Initiativen zur digitalen Kompetenz, Bildung oder Schulung, um das Bewusstsein der Nutzer zu verbessern und/oder Menschen zu helfen, die Art, Fähigkeiten, Einschränkungen und Auswirkungen fortschrittlicher KI-Systeme besser zu verstehen? Bitte geben Sie Beispiele.

Buildsimple-Mitarbeiter nehmen regelmäßig als Sprecher an Fachvorträgen teil und veröffentlichen Publikationen in einschlägigen Medien. Ziel hierbei ist es immer, die Möglichkeiten und Fähigkeiten von KI-Systemen richtig einzuordnen und deren Einschränkungen und Auswirkungen einem breiten Publikum zugänglich zu machen. Weiterhin werden bei Buildsimple regelmäßig Werkstudenten im KI-Bereich ausgebildet.

c. Priorisiert Ihre Organisation KI-Projekte zur verantwortungsvollen Verwaltung vertrauenswürdiger und menschenzentrierter KI im Rahmen der UN-Ziele für nachhaltige Entwicklung? Bitte geben Sie Beispiele.

Die Entwicklung der Buildsimple-KI steht im Einklang mit den UN-Zielen für nachhaltige Entwicklung. Sie unterstützt bei Versicherern im Gesundheitssektor, wird zur Ausbildung von Studenten verwendet, ist

nicht diskriminierend gegenüber Geschlechtern und führt bei Kunden zu Wirtschaftswachstum, da mehr Arbeit in kürzerer Zeit erledigt werden kann.

d. Arbeitet Ihre Organisation mit der Zivilgesellschaft und Gemeinschaftsgruppen zusammen, um KI-Lösungen zur Unterstützung der UN-Ziele für nachhaltige Entwicklung zu identifizieren und zu entwickeln, um die größten Herausforderungen der Welt anzugehen? Bitte geben Sie Beispiele.

Buildsimple arbeitet primär mit regulierten Organisationen (Banken, Versicherungen, zunehmend öffentlicher Sektor) zusammen. Direkte, formalisierte Partnerschaften mit der Zivilgesellschaft/Community-Gruppen sind aktuell (noch) nicht flächendeckend etabliert. Buildsimple bindet jedoch zivilgesellschaftliche Perspektiven dort ein, wo Lösungen Bürger:innen bzw. Endkund:innen unmittelbar betreffen – insbesondere in Projekten mit dem öffentlichen Sektor und in Use Cases mit hoher sozialer Wirkung.

In Projekten rund um digitalen Posteingang und dokumentbasierte Prozesse werden Stakeholder-Workshops durchgeführt, in denen neben Verwaltung auch Bürger-/Community-Perspektiven (z. B. Beratungsstellen, Sozialverbände, Ombudsstellen) in Anforderungen wie Verständlichkeit, Nachweisführung, Fristen und barrierearme Kommunikation einfließen.

In dokumentenintensiven Abläufen (z. B. Leistungsprüfung bei Versicherungen) hilft Buildsimple, Routinetätigkeiten zu reduzieren und Bearbeitungszeiten zu verkürzen. In der Ausgestaltung werden Anforderungen von Kundendienst-Teams und ggf. Interessenvertretungen (z.B. Patienten-/Versichertenperspektive, Beschwerdemanagement) berücksichtigt, um Transparenz und konsistente Behandlung sicherzustellen.

Gerade im öffentlichen Sektor ist der Schutz sensibler Daten zentral. Die Ausrichtung von Buildsimple auf souveränen Betrieb in europäischen Rechenzentren und auf Governance (Versionierung, Staging, Freigaben, Nachvollziehbarkeit) unterstützt Organisationen dabei, KI sicher und regelkonform einzusetzen.