Datenschutz & Datensicherheit bei Chatbots

Masterseminar: Conversational AI. Dialogsysteme, Bots, Assistenten.
Wintersemester 2021/22

Dzmitry Kirylenka

13.01.2022

Gliederung

Datenschutz vs. Datensicherheit

Datenschutz

- Schutz von **personenbezogenen** (pb) Daten
- Bsp. für pb-Daten: Name, Geburtsdatum, Anschrift, Herkunft, Religion, Gesundheit, etc.
- **Ziel**: Schutz des Rechts auf informationelle Selbstbestimmung
- Umfasst rechtliche Fragen zur Verarbeitung von pb-Daten
- Verankerung von entspr. Normen auf nationaler (z.B. BDSG) und/oder internationaler (z.B. DSGVO) Ebene möglich

Datensicherheit

- Schutz von Daten **allgemein**
- **Ziel**: Gewährleistung der *Vertraulichkeit*, *Integrität* & *Verfügbarkeit* von Daten
- Umfasst technische Maßnahmen und Ansätze zum Schutz von Daten

DSGVO

- EU-Datenschutz-Grundverordnung
- Seit 27. April 2016 (informell) bzw. 25. Mai 2018 (formell) in Kraft
- **Ziel:** einheitlicher Schutz von pb-Daten aller EWR-Bürger
- Zusammenspiel zwischen Richtlinie und Verordnung
- Verbindlich für alle EWR-Mitgliedstaaten
- Verknüpft Datenschutz & Datensicherheit
- Das strikteste Datenschutzgesetz der Welt

DSGVO-Grundsätze

Quelle: `https://mycitykids.de/wichtigsten-punkte-der-dsgvo-fuer-start-ups-blogger-unternehmen/` (Abfragedatum: 08.01.2022)

Betroffenenrechte nach DSGVO

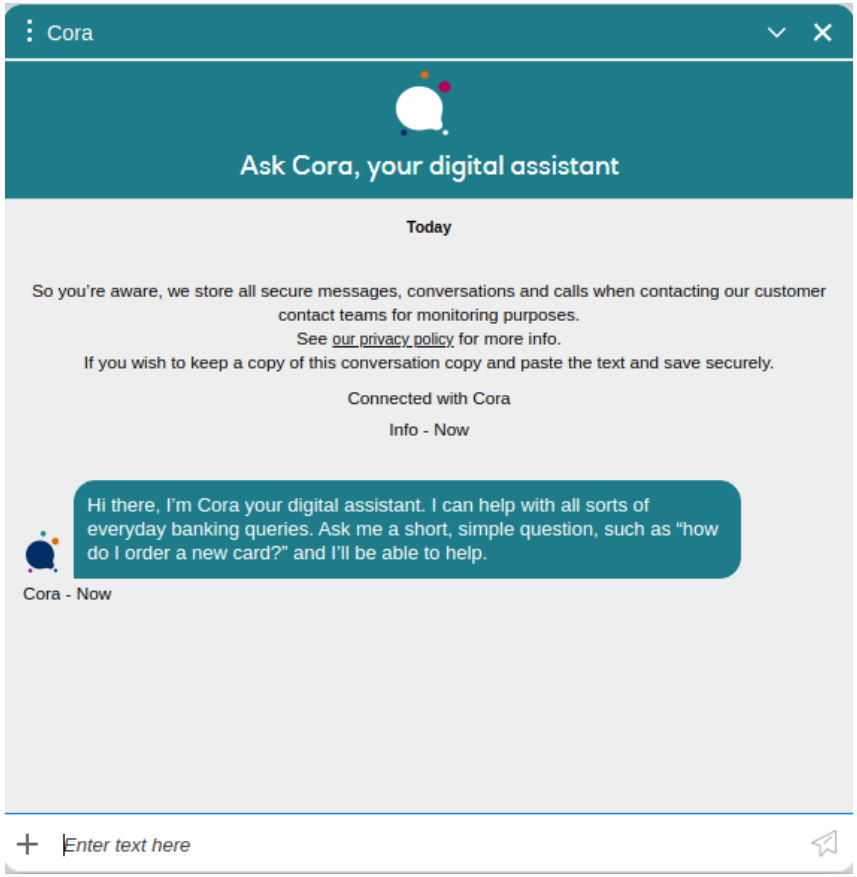
Quelle: `https://easygdpr.eu/de/2019/04/recht-auf-berichtigung/` (Abfragedatum: 08.01.2022)

Rechtmäßigkeit & Transparenz (1/5)

- Mangelnde Hinweise & Einholung der Nutzereinwilligung (z.B. als Opt-In) zur Datenverarbeitung

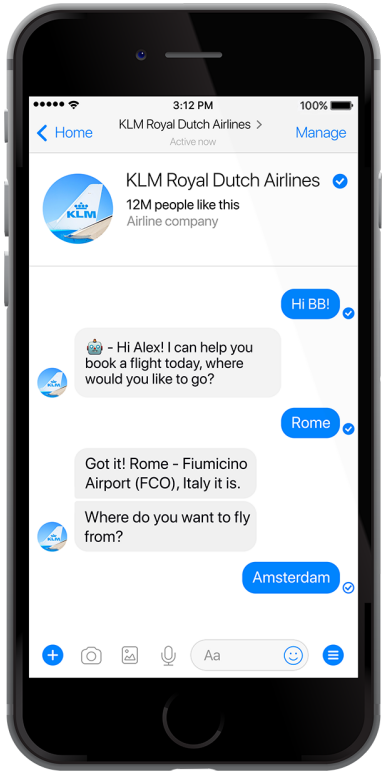
Rechtmäßigkeit & Transparenz (2/5)

Quelle: `https://enso.de/wps/portal/enso/cms` (Abfragedatum: 08.01.2022)

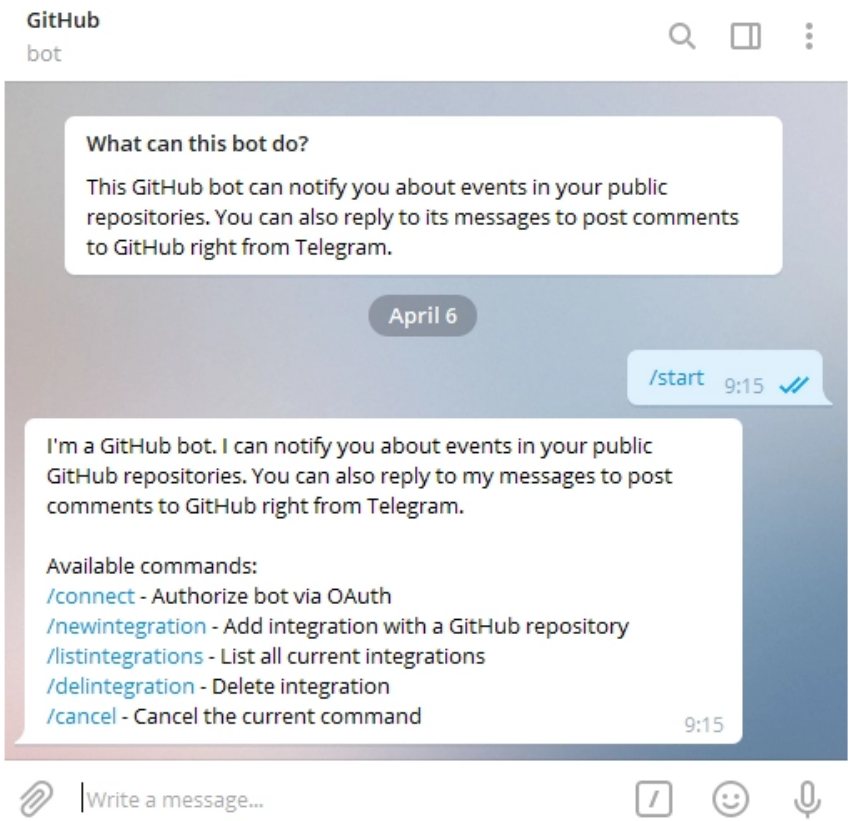


Quelle: `https://www.rbs.co.uk/support-centre/cora.html` (Abfragedatum: 08.01.2022)

Rechtmäßigkeit & Transparenz (3/5)



Quelle: `https://www.altoros.com/blog/klm-handles-2x-more-customer-requests-with-artificial-intelligence/` (Abfragedatum: 08.01.2022)



Quelle: `https://www.affde.com/de/sendpulse-telegram-chatbot.html` (Abfragedatum: 08.01.2022)

Rechtmäßigkeit & Transparenz (4/5)

- Unzureichende Transparenz bzgl. Datenschutz
  - Aktivierung von Sprachassistenten
  - Datensammlung
  - Datenauswertung
  - Datenübermittlung an Drittanbieter
- Stimmaufnahme von Dritten (Gäste, Kinder) $\rightarrow$ Mangel an Rechtsgrundlage zur Datenverarbeitung
- Datentransfer in Drittländer $\rightarrow$ DSGVO gilt nicht mehr
- Zugriff auf EU-Server von US-Unternehmen nach US-Recht

Rechtmäßigkeit & Transparenz (5/5)

- (Oft) Erzwungene Einwilligung zur Datenverarbeitung → Verletzung der Freiwilligkeit
- Installation von Erweiterungen (bei Sprachassistenten)
  - Keine explizite Zustimmung zur Datenverarbeitung
  - Keine Bereitstellung von Datenschutzrichtlinien
  - → Manuelle Prüfung notwendig
- Änderung von Nutzungsbedingungen jederzeit möglich
- KI-Modelle zum Training auf Nutzerdaten stellen Black Box dar → Transparenz problematisch
- → Permanente Überwachung & Verletzung der Privatsphäre

Zweckbindung, Datenminimierung & Speicherbegrenzung

- Vage Formulierungen in Datenschutzrichtlinien
- (Meist) Unbefristete Speicherung von Sprachaufzeichnungen voreingestellt → Manuelle Anpassung notwendig
- Profiling → Konflikt mit dem Prinzip der Zweckbindung & Datenminimierung
- Stimmaufnahme von Dritten (Gäste, Kinder) → Verletzung des Prinzips der Datenminimierung

Lösung

- Selbstzerstörung von Konversationen
- Lokale Datenverarbeitung

Recht auf Auskunft, Löschung & Widerspruch

- Mangelnde Bereitstellung erhobener Nutzerdaten auf Anfrage
- Kann von manchen Anbietern in bestimmten Fällen missachtet werden (z.B. Apple)
- Tatsächliche Löschung der Daten von Servern nach Anfrage fraglich

Authentifizierung & Autorisierung

- Häufige Missachtung bzw. unzureichende Wahrnehmung
- Zugriff auf sensible (Konto-)Daten durch Dritte als Folge
- Beeinträchtigung der Datenvertraulichkeit
- Trade-off zwischen Bequemlichkeit und Sicherheit
- Smishing

Lösung

- Explizite Anpassung von Standardeinstellungen
- Anlegen eines Stimmprofils
- Temporäre Authentifizierungstoken
- Zwei-Faktor-Authentisierung (2FA)
- Wachsamkeit

Ende-zu-Ende-Verschlüsselung (E2EE)

E2EE

- Verschlüsselung von Daten auf dem gesamten Übertragungsweg
- Nur Sender & Empfänger können Daten entschlüsseln
- Plattformanbieter haben keinen Zugang auf Daten im Klartext

- Bisher nur gering im Einsatz
- Bsp.: Facebook-Chatbots, Telegram-Bots, Amazon Echo
- Beeinträchtigung der Datenvertraulichkeit & -integrität
- $\rightarrow$ Nur bedingt DSGVO-konform (vgl. Art. 32 DSGVO)

Lösung

- VPN
- Homomorphe Verschlüsselung

Aktivierung von Sprachassistenten

- Häufige Verwechslung von Weckrufen mit anderen Wörtern
- Aktivierung auch aus der Ferne möglich (z.B. TV, IoT-Geräte, Anruf)
- Aktivierung durch lautlose Befehle möglich
- $\rightarrow$ Unbeabsichtigte Aufnahme von Gesprächen
- $\rightarrow$ Verletzung der Privatsphäre & Gefährdung menschlicher Sicherheit

Lösung

- Virtual Security Button (VSButton) Tool zur Erkennung von menschlichen Bewegungen mittels WLAN-Signale [Lei et al., 2018]
- Human-to-Bot-Klassifikator (Bsp.: Xiaoice) [Zhou et al., 2020]

Adversarial Attack

- Angriff auf ML/DL-Modell mit dem Ziel, Klassifikationsergebnisse zu beeinflussen
- Attacke mittels manipulierten Dateninputs ins neuronale Netz (= *Adversarial Examples*)
- Kann durch Einsatz vom "umgekehrten Dialoggenerator" umgesetzt werden
- $\rightarrow$ Beeinträchtigung der Datenintegrität

Lösung

- Einsatz eines Hassrede-Detektors [Ye & Li, 2020]
- Neuronaler Klassifikator zur Abgrenzung zwischen Hassrede und Normalsprache [Dinan et al., 2019]

Konklusion

- 100-prozentige Sicherheit gibt es nicht
- Datenschutz relativ selten im Vordergrund der Entwicklung von Chatbots/Sprachassistenten
- Umfassende Implementierung von 'Privacy by Design' durch Tech-Giganten relativ unwahrscheinlich, weil in Konflikt mit dem Geschäftsmodell
- DSGVO-Umsetzung bei Entwicklung von Chatbots verbesserungsbedürftig
- Trade-off zwischen Komfort und Privatsphäre
- DSGVO wichtiger Meilenstein für den Schutz von Privatsphäre und Grundrechten

Quellen I


Apple Inc. (2021). *Apple Datenschutzrichtlinie*. URL: https://www.apple.com/legal/privacy/de-ww/, Abfragedatum: 08.01.2022.


Becker, L. (2021). *Siri hört ungewollt mit: Apple wird Datenschutzklage nicht los*. URL: https://www.heise.de/-6181652, Abfragedatum: 08.01.2022.


Dinan, E., Humeau, S., Chintagunta, B. & Weston, J. (2019). Build it Break it Fix it for Dialogue Safety: Robustness from Adversarial Human Attack. In Inui, K., Jiang, J., Ng, V. & Wan, X. (Hrsg.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (S. 4537–4546). Hong Kong: Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/D19-1, Abfragedatum: 08.01.2022.


Donath, A. (2019). *Apple soll DSGVO nicht vollständig erfüllen*. URL: https://www.maclife.de/news/apple-soll-dsgvo-nicht-vollstaendig-erfuellen-100111423.html, Abfragedatum: 08.01.2022.

Quellen II

Europäische Union (2016). *Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)*. URL: `https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679`, Abfragedatum: 08.01.2022.

Fang, H. & Qian, Q. (2021). Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning. *Future Internet, 13*(94), S. 1–20. DOI: 0.3390/fi13040094.

Funke, S. (2020). Datenschutz – Ein Baustein von Utility 4.0. In Doleski, O. (Hrsg.), *Realisierung Utility 4.0* (Band 1, S. 301–315). Wiesbaden: Springer Vieweg. DOI: 10.1007/978-3-658-25332-5_18

Google LLC. (2021). *Datenschutzerklärung & Nutzungsbedingungen*. URL: `https://policies.google.com/privacy`, Abfragedatum: 08.01.2022.

Quellen III

Harkous, H. (2016). *Encryption, AI, and the Myth of Incompatibility*. URL: https://chatbotsmagazine.com/encryption-ai-and-the-myth-of-incompatibility-9afca1ca115?gi=f3e44d19e7e2, Abfragedatum: 08.01.2022.

Hasal, M., Nowaková, J., Saghair, K., Abdulla, H., Snásel, V. & Ogiela, L. (2021). Chatbots: Security, privacy, data protection, and social aspects. *Concurrency and Computation: Practice and Experience, 33*(19), S. 1–13. DOI: 10.1002/cpe.6426.

Lei, X., Tu, GH., Liu, AX., Li, CY. & Xie, T. (2018). The Insecurity of Home Digital Voice Assistants - Vulnerabilities, Attacks and Countermeasures. *2018 IEEE Conference on Communications and Network Security (CNS)*, S. 1–9. DOI: 10.1109/CNS.2018.8433167.

Lentzsch, C., Shah, SJ., Andow, B., Degeling, M., Das, A. & Enck, W. (2021). Hey Alexa, is this Skill Safe?: Taking a Closer Look at the Alexa Skill Ecosystem. *Network and Distributed Systems Security (NDSS) Symposium 2021*, S. 1–18. DOI: 10.14722/ndss.2021.23111.

Quellen IV

Reuters Staff. (2019). *Austrian data privacy activist files complaint against Apple, Amazon, others*. URL: https://www.reuters.com/article/us-europe-privacy/austrian-data-privacy-activist-files-complaint-against-apple-amazon-others-idUSKCN1PC1FA, Abfragedatum: 08.01.2022.

Schaber, F., Krieger-Lamina, J. & Peissl, W. (2019). *Digitale Assistenten. Projektbericht*. Wien: Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften. URL: https://epub.oeaw.ac.at/ita/ita-projektberichte/2019-01.pdf, Abfragedatum: 08.01.2022.

Schwenke, T. (2020). *EuGH zu Schrems II: EU/US-Privacy Shield ist unwirksam – Was Unternehmen jetzt wissen müssen*. URL: https://datenschutz-generator.de/eugh-privacy-shield-unwirksam, Abfragedatum: 08.01.2022.

Quellen V


Xu, H., Ma, Y., Liu, HC., Deb, D., Liu, H., Tang, JL. & Jain, AK. (2020). Adversarial Attacks and Defenses in Images, Graphs and Text: A Review *International Journal of Automation and Computing, 17*, S. 151–178. DOI: 10.1007/s11633-019-1211-x.


Ye, W. & Li, Q. (2020). Chatbot Security and Privacy in the Age of Personal Assistants. *2020 IEEE/ACM Symposium on Edge Computing (SEC)*, S. 388–393. DOI: 10.1109/SEC50012.2020.00057.


Zhou, L., Gao, J. & Heung-Yeung Shum, DL. (2020). The Design and Implementation of Xiaoice, an Empathetic Social Chatbot. *Computational Linguistics, 46*(1), S. 53–93. DOI: 10.1162/coli_a_00368.

Vielen Dank für die Aufmerksamkeit!