**Merkblatt: KI-Einsatz zum Generieren von Sozialberichten**

Dieses Merkblatt ist am zweitägigen hack4socialgood in einem interdisziplinären Team aus Fachpersonen des Sozialwesens und Ingenieurwesen entstanden. Es dient als eine erste Ausgangslage für die Anwendung von KI für Sozialberichte.   
  
Wenn man KI nutzen möchte, um aus vorliegenden Daten (z.B. Akten, Aktennotizen, etc.) einen Sozialbericht zu generieren, stellen sich Fragen in den folgenden drei Bereichen:

* Technische Lösung
* Compliance: Datenschutz, etc. (Was darf man im Zusammenhang mit hoch sensitiven Daten?)
* Sozialarbeiterische Aspekte

1. **Technische** **Lösung**
2. Prüfen, ob ChatGPT/oder eine andere Open Source KI genutzt werden können, um z.B. eine neutrale Textvorlage zu erstellen (ohne spezifische Angaben des Falls). Diese nachher mit den eigenen sensitiven Daten füllen.
3. Abklären: Können andere Daten als Basis genutzt werden, um als Vorlage für Berichte zu dienen, wenn die eigenen Berichte nicht genutzt werden können (weil zu wenige vorhanden sind oder weil sie wegen sensitiven Daten nicht hochgeladen werden können)?
4. Prompts so gestalten, dass die KI die richtigen Inhalte aus den Datensätzen auswählt für den neuen Sozialbericht (z.B. Welche Inhalte müssen nicht in den Bericht übertragen werden? Welche Inhalte sind besonders wichtig und sollten hervorgehoben werden? Etc.)
5. Prompts so gestalten, dass aus unterschiedlichen Textsorten und Stilen (z.B. medizinische, juristische, sozialarbeiterische Texte) der neue Sozialbericht im gewünschten Stil generiert wird
6. Sicherstellen, dass das Tool, das man nutzt, sich nicht laufend (durch Updates der Firma, durch neue Datensätze, etc.) so verändert, dass die eigenen Prompts nicht mehr funktionieren, bzw. nicht mehr die Art Text generiert wird, den man braucht.

**Bestandteile des zukünftigen LLM-Programms**

* Aufbau des Berichts;
* Schreibstil (soll sachlich sein, nicht bewertend)
* Kürzel / Abkürzungen erfassen, damit dies KI nachvollziehen kann bzw. der gleichen Person zuordnen kann (z.B. Mutter, Fr. M. = KM)
* Möglichkeit einzelne Abschnitte nochmals neu zu generieren (mit zugefügten Informationen)
* Rechtschreibkorrektur
* Halluzinieren minimieren

1. **Compliance**
2. Klären: Welche rechtlichen Grundlagen sind für die betroffene Organisation zu beachten bezüglich Datenschutz?
3. Klären: Wer kann für die betreffende Organisation Auskunft geben, was welche rechtlichen Möglichkeiten bzw. welche Grenzen bestehen, wenn KI genutzt werden soll?
4. Sicherstellen: Es dürfen keine sensitiven Daten in falsche Hände geraten (z.B. ins Ausland oder in Datensätze der genutzten KI, etc.)?

**Inputs zum Thema Datenschutzanforderungen**

datenschutz.ch

[Datenschutz im Sozialbereich](https://docs.datenschutz.ch/u/d/publikationen/leitfaeden/leitfaden_datenschutz_im_sozialbereich.pdf)

[Entwicklung datenschutzkonformer Apps](https://docs.datenschutz.ch/u/d/publikationen/leitfaeden/leitfaden_entwicklung_datenschutzkonformer_apps.pdf)

[Datenschutz im Kindes- und Erwachsenenschutzrecht](https://docs.datenschutz.ch/u/d/publikationen/uebersichten/uebersicht_datenschutz_im_kindes_und_erwachsenenschutzrecht.pdf)

[Nutzung externer Cloud-Dienste (datenschutz.ch)](https://docs.datenschutz.ch/u/d/publikationen/leitfaeden/leitfaden_nutzung_externer_cloud_dienste.pdf)

Grundsatz  
Die Daten, welche in diesem Fall bearbeitet werden sind üblicherweise «besonders schützenswerte» (Gesundheitsdaten, Religionszugehörigkeit etc.), je nach anwendbarem Recht heissen diese «besonders sensible». Anwendbares Recht ist üblicherweise das Datenschutzgesetz und -verfügung und das jeweilige kantonale Recht (im Kt. ZH ist es das IDG/IDV)  
  
Anforderungen  
Grundsätzlich dürfen besonders schützenswerte Personendaten nicht bekanntgegeben werden und sollen so sparsam wie möglich gesammelt werden (Privacy by Default). Eine Datenbekanntgabe findet statt, wenn die Daten z.B. auf einen Cloud-Service geladen oder in einen Chat-Prompt (ChatGPT) eingegeben werden.

Wenn nun die Daten bekanntgegeben werden, müssen strenge Regeln befolgt werden. Diese wurden von der kantonalen Datenschützerin (Kt. ZH) in Zusammenarbeit mit anderen kant. Datenschutzstellen gut aufbereitet (unter [www.datenschutz.ch](http://www.datenschutz.ch)).

«Einfache» Lösung 1

Am einfachsten ist es deshalb, wenn die Daten überhaupt nicht ausgelagert werden müssen. Dazu müsste innerhalb eines Betriebes eine LLM-Instanz laufen (vortrainiertes Modell à la Llama).

Bessere Handhabung wäre es vermutlich, auf einer Schweizer-Cloud-Lösung (e.g. Exoscale) auf einem geeigneten Server diese Instanz aufzusetzen. Das wäre vom Datenschutz her relativ einfach umzusetzen, kann auch für mehrere Institutionen genutzt werden und ist variabel für eigene Entwicklungen (Tuning etc.).

«Einfache» Lösung 2

Wenn nur Textbausteine erstellt werden müssen oder generelle Texte, die keinerlei Personendaten (auch nicht pseudonymisierte) enthält (die auch nicht aggregiert zu einer Person deuten können (Singularisierung), kann auch ChatGPT oder datenschutzrechtlich nicht erlaubte Dienste genutzt werden. Der Nutzen ist in diesem Fall aber relativ beschränkt.

1. **Sozialarbeiterische** **Aspekte**
2. Prompts so gestalten, dass sie nicht nur Zusammenfassungen der vorliegenden Dokumente macht, sondern so, dass die sozialarbeiterischen Überlegungen in den Text einfliessen.
3. Sicherstellen: Unterscheidung in deskriptiver und bewertender Bericht (durch KI generiert, vgl. zu predictive analytics. Anwendung KI in einem ersten Schritt ausschliesslich deskriptiv, d.h. ausschliesslich beruhend auf Daten aus den Akten).
4. Sicherstellen: dass alle relevanten Informationen, die im Sozialbericht erscheinen müssen, auch bereits in der Datenbasis (Aktennotizen, etc.) vorhanden sind.
5. Klären: Daten so strukturieren, dass sie als Basis für die KI genutzt werden können, um den Bericht zu generieren (z.B. Vorgaben, wie die Dokumente erstellt werden müssen, damit sie als Grundlage genutzt werden können: Abkürzungen, Namen, Daten). *Idee: Im Aktensystem bereits Unterscheidungen erfassen z.B. nach Zuweisungsgrund/Hintergrundgeschichte, vereinbarte Zielsetzungen, Standortgespräch (Verlauf), Abschluss / Evaluation / Empfehlung oder nach Beobachtung, Hypothese, Beschreib.*

Empfehlungen:

* Die Institution muss sich im ersten Schritt überlegen, was möchte sie jeweils im Bericht erfassen und anhand derer ein Aktensystem führen. Allenfalls im Fallführungssystem unterscheiden in berichtsrelevant und nicht berichtsrelevant. Somit können sensiblere (nicht berichtsrelevante) Daten dennoch erfasst werden.
* Durch Risiken wie Halluzinieren (Ergänzungen von Informationen durch KI, welche nicht stimmen) wird ein erweitertes Sicherheitskonzept benötigt. Wie können Falschinformationen verhindert werden? *Idee: Entwurf KI mit Wasserzeichnen generieren, Sozialarbeitende muss dies bewusst „freischalten“; 4-Augen-Prinzip*.