Richtlinie zum „Software-Sprint“

Private Prompts

Schlussbericht

Zuwendungsempfänger:

Frank Börncke

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS24S44 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Meine Motivation war der Wunsch nach einer sicheren, lokal nutzbaren Software für den Umgang mit KI-Prompts und Cloud-Diensten, da bestehende Lösungen meine Datenschutzanforderungen nicht erfüllten. Ich habe mich unwohl dabei gefühlt, sensible und private Daten unkontrolliert an Online-Dienste zu übermitteln.

Problemstellung: Der zunehmende Einsatz von generativen KI-Modellen wie ChatGPT oder Webdiensten wie Deepl oder Google Translate birgt das Risiko, bewußt oder unbewußt eigene und anvertraute vertrauliche Daten an Drittanbieter weiterzugeben. Viele Nutzer wünschen sich eine Möglichkeit, KI-gestützte Texteingaben sicherer und datenschutzfreundlicher zu gestalten.

Vorgehensweise: Die Software ermöglicht eine datenschutzfreundliche Nutzung von KI-Tools, indem sensible Daten lokal vor der Verarbeitung pseudonymisiert werden. Nach Abschluß der Verarbeitung kann die Pseudonymisierung wieder rückgängig gemacht werden. Die Anwender können auf diese Weise externe Tools nutzen, ohne dass die sensiblen Informationen jemals den eigenen Rechner verlassen. Private Prompts trägt so dazu bei, Nutzern mehr Kontrolle über ihre Daten zu geben.

Die wesentlichen Meilensteine waren:

- Entwicklung eines funktionsfähigen Prototyps (MVP)
- Zentrale Funktionalität zur Maskierung und Pseudonymisierung von sensitiven Informationen
- Ein Prompt-Managers bietet Mehrwert unabhängig von der Datenschutz-Funktionalität
- Datenhaltung komplett lokal ohne Cloud
- Bereitstellung der Software für mehrere Betriebssysteme (Windows, macOS, Linux)

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Die Hauptzielgruppe von Private Prompts sind Einzelanwender, die KI-Modelle wie ChatGPT oder Webtools wie Google Translate und DeepL nutzen und dabei Wert auf Datenschutz und die Kontrolle über ihre Daten legen. Diese Zielgruppe lässt sich weiter differenzieren: (1) Privatpersonen und Wissensarbeiter: Dazu gehören Autoren, Wissenschaftler, Lehrer, Menschen in der Ausbildung und kreative Köpfe, die KI-Tools für ihre Arbeit nutzen und dabei sicherstellen wollen, dass ihre sensiblen Daten geschützt bleiben. (2) Journalisten und Datenschützer haben oft mit vertraulichen Informationen zu tun und können von einer Lösung profitieren, die den sicheren Umgang mit Webdiensten erleichtert.

Die einfache Bedienung ermöglicht es jedem, der ChatGPT bedienen kann, auch Private Prompts zu nutzen. Positives Feedback aus der Community und auch ein mediales Interesse haben gezeigt, dass der Use Case für viele Menschen von Bedeutung ist.

Für den Einsatz im Unternehmen wären noch zusätzliche Anpassungen erforderlich (Verschlüsselung, Einsatz im Netzwerk, Middleware-Betrieb).

Ausführliche Darstellung der Ergebnisse

Viele der angestrebten Ziele wurden erreicht:

- Die Anwendung ist plattformübergreifend verfügbar.
- Die Pseudonymisierung funktioniert zuverlässig und erhöht den Datenschutz.
- Der Prompt-Manager wurde als eigenständiges Modul etabliert.

Herausforderungen

Die Bereitstellung von Binaries für verschiedene Plattformen war technisch anspruchsvoller als erwartet. Eine noch nicht umgesetzte Signierung führte bei manchen Nutzern zunächst zu Problemen bei der Installation oder mit Virenscannern. Die Idee eines Browser-Plugins musste verworfen werden, da zentrale Funktionen der UI im Plugin-Kontext nicht umsetzbar waren.

Begleitung durch die Open Knowledge Foundation

Das Coaching war sehr hilfreich, ebenso die Erfahrungsdokumentationen im Wiki. Dadurch konnten typische Herausforderungen besser bewältigt werden.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Durch die Berichterstattung in der Süddeutschen Zeitung wurde das Projekt bekannt und das Thema Datenschutz in KI-Anwendungen thematisiert. Es gab viele Anfragen per Mails von interessierten Nutzern, auch weitere Medien haben Interesse an einem Gespräch und Austausch bekundet.

Durch die Open-Source-Stellung haben sich weiterführende Effekte ergeben: Community-Mitglieder haben wertvolle Feature-Wünsche eingebracht. Durch die Veröffentlichung unter einer freien Open-Source Lizenz erhöhen sich Transparenz, Nachvollziehbarkeit und Vertrauenswürdigkeit der Software. Andere Entwickler können auf der bestehenden Code-Basis aufbauen und die Software weiterentwickeln oder für eigene Anwendungsfälle adaptieren. Denkbare Erweiterungen wären:

- Weitere Optimierungen der UI und Benutzerfreundlichkeit.
- Erweiterung der Zielgruppe: Anpassung für den Unternehmenseinsatz, Middleware-Modus
- Verschlüsselung
- Umschaltbare Regel-Profile, die sich Use-Case spezifisch de-/aktivieren lassen
- Integration weiterer Features auf der Basis von Nutzer-Feedback.

Finanzierung und Zukunft des Projekts

Obwohl es viele Ideen und aktive Nutzer gibt, fehlt aktuell eine nachhaltige Finanzierung. Ohne neue Förderoptionen wird Private Prompts in ein reines Privatprojekt übergehen. Eine "Second Stage"-Förderung wäre erforderlich, um die nächsten Entwicklungsschritte zeitnah umzusetzen. Durch die mediale Berichterstattung können sich vielleicht neue Finanzierungsmöglichkeiten ergeben.

Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?

Das Projekt hat meine technische und organisatorische Kompetenz deutlich erweitert:

- Tiefergehendes Wissen über plattformübergreifende Software-Entwicklung gewonnen.
- Erfahrungen mit neuen Technologien und Frameworks gesammelt.
- Kommunikation: Der frühe öffentliche Launch führte zu wertvollen Nutzer-Rückmeldungen.
- Öffentliche Aufmerksamkeit und Medienberichterstattung waren eine neue Erfahrung.

Private Prompts hat gezeigt, dass es eine starke Nachfrage nach datenschutzfreundlichen KI-Lösungen gibt. Das Projekt hat nicht nur dazu beigetragen, ein reales Problem zu lösen, sondern mir auch dabei geholfen, meine fachlichen, strategischen und kommunikativen Fähigkeiten zu verbessern. Die nächsten Schritte hängen wesentlich von der weiteren Finanzierung ab.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Während der Entwicklung gab es einige Ziele, deren Umsetzung nicht bis zum Ende verfolgt wurden:

- Browser-Plugins: Ursprünglich war geplant, eine Browser-Plugin-Variante der Software zu entwickeln, um Nutzern eine direkte Integration in ihre Web-Anwendungen zu ermöglichen. Allerdings erwiesen sich technische Einschränkungen der Browser-Sandbox als Hindernis. Stattdessen wurde der Fokus dann auf eine eigenständige Desktop-Anwendung gelegt.

- Bestimmte plattformabhängige Features: Einige geplante Funktionen, wie eine tiefergehende Integration mit nativen Betriebssystem-Funktionen, mussten verworfen werden, da sie plattformübergreifend schwer zu implementieren waren. Stattdessen wurde auf eine möglichst breite Kompatibilität mit verschiedenen Betriebssystemen gesetzt.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Die Projekthomepage https://www.privateprompts.org (Deutsch) ist die zentrale Anlaufstelle für interessierte Nutzer. Zusätzliche Informationen und technische Hintergründe gibt es auf der GitHub-Seite zu dem Projekt: https://github.com/fboerncke/private-prompts-prototype (Englisch).

Ein Newsletter, der regelmäßig über Updates berichtet, ist in Planung.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Da ich, wie oben dargestellt, von einigen Ideen Abstand genommen habe, konnte ich die dadurch freigewordene Zeit für andere Teilaufgaben nutzen, deren Aufwand ich unterschätzt hatte. Dadurch blieb die Gesamtarbeitszeit im geplanten Rahmen, auch wenn sich einzelne Prioritäten im Laufe des Projekts verschoben haben. An einigen Punkten habe ich auch Funktionen in die Anwendung integriert, die in dieser Form über den ursprünglich geplanten Projektumfang hinaus gehen.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Gab es Entwicklungen anderer Personen oder Institutionen, die Einfluss auf Deine Arbeiten und die Zielsetzung hatten? Wenn ja, worin bestand dieser und wie bist Du damit umgegangen?

Die Zielsetzung an sich wurde nicht beeinflusst. Da ich aber sehr frühzeitig mit meinen Ideen an die Öffentlichkeit gegangen bin, habe ich viel positives Feedback bekommen, was mich darin bestärkt hat, die Umsetzung meiner Zielvorstellung fortzusetzen und zum Abschluss zu bringen.