Richtlinie zum „Software-Sprint“

Human Rights Predictor (HRP)

Schlussbericht

Zuwendungsempfänger:

Viraaj Akuthota

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS24S07 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Was war Deine Motivation? Welches Problem wolltest Du mit Deinem Projekt lösen? Wie war die geplante Vorgehensweise zur Problemlösung (auch Angabe der wichtigsten Meilensteine)?

Menschenrechtsverletzungen nehmen zu: Die Zahl der Flüchtlinge steigt, in der Ukraine werden Kriegsverbrechen begangen. Dennoch gibt es kaum Wiedergutmachung. Ein Grund dafür sind strukturelle rechtliche Hürden wie Kosten, Zeit und Kompetenz, die für die Beurteilung von Menschenrechtsverletzungen erforderlich sind. Für die meisten Menschen, insbesondere für marginalisierte Gruppen, ist es daher mangels Ressourcen, Sprachkenntnissen und Kenntnis des lokalen Systems praktisch unmöglich, eine angemessene Beurteilung vorzunehmen. Die Opfer sind daher auf Rechtsbeistände angewiesen. Diese erhalten jedoch immer weniger öffentliche Mittel oder Prozesskostenhilfe, was eine enorme Belastung für sie darstellt und dazu führt, dass eine beträchtliche Anzahl von Personen überhaupt keine Unterstützung erhält - und gleichzeitig eine sinnvolle Beteiligung der Opfer verhindert.

Ziel des HRP ist es, die Wissens- und Finanzlücken zu schließen, mit denen Einzelpersonen konfrontiert sind, die mögliche Verletzungen der Europäischen Menschenrechtskonvention (EMRK) beurteilen wollen. Dazu geben die NutzerInnen ihre persönliche Situation in einfacher Sprache ein und erhalten eine Neuformulierung ihres Problems in juristischer Sprache (Issue Identifier), eine nach Relevanz geordnete Liste verletzter Menschenrechtsartikel (Predictor) sowie die wichtigsten Aspekte von Fällen vor dem Europäischen Gerichtshof für Menschenrechte (EGMR), die für das Problem kontextabhängig relevant sind (Search). Für Opfer von Menschenrechtsverletzungen erleichtern diese Funktionen und Ergebnisse den Zugang zum Recht erheblich. Für Experten dienen sie als differenzierter Ausgangspunkt für Recherchen und schaffen damit zusätzliche freie Kapazitäten.

Die relevanten Meilensteine bei der Fertigstellung des HRP waren:

1. Erweiterung des Datensatzes (11.000 Fälle) um die über 30.000 Fälle von My-Rights-info;
2. Bereinigung der Daten, um inkorrekte Referenzen für den Predictor zu entfernen;
3. Tests verschiedener Encoder Models und Entwicklung von Strategien zur Maximierung der Sequenzlänge, um die Verbesserungen des Predictor zu erkennen;
4. Erstellung von sinnvollen Eingabepaaren, die um unterschiedlichen Niveaus bei der Eingabe gerecht zu werden;
5. Identifizierung von Verbesserungen des FAISS-Modells durch die Verwendung von verschiedener Vektorisierungsmethoden;
6. Einführung von automatischer Übersetzung;
7. Tests mit Anwälten und marginalisierten Bevölkerungsgruppen;
8. Umsetzung der user-centered Design-Prinzipien und der Barrierefreiheit;
9. Veröffentlichung einer Webanwendung, die alle Meilensteine erfüllt.

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

Welche Bezüge gibt es zu den Themenfeldern „Civic Tech“ und „Data Literacy“ des Software Sprints oder zu weiteren gesellschaftlich relevanten Zielen bzw. Lösungsansätzen?

Obwohl es keine direkten Verweise auf „Civic Tech“ und „Data Literacy“ innerhalb der Plattform gibt, befasst sich das gesamte HRP-Projekt mit diesen Themen: Die Plattform fördert Data Literacy, indem sie es den Nutzerinnen und Nutzern ermöglicht, die Daten in HUDOC, der Urteilsdatenbank des Europäischen Gerichtshofs für Menschenrechte, besser zu verstehen. Bisher war diese Datenbank nur für Juristinnen und Juristen oder Personen mit komplexen Datenbankkenntnissen zugänglich. Die HRP-Plattform beseitigt diese Hindernisse, indem sie den Nutzerinnen und Nutzern ein benutzerfreundliches Instrument für den Zugang zur Rechtsprechung im Bereich der Menschenrechte zur Verfügung stellt. Als Civic-Tech-Lösung trägt das Projekt somit zu gesellschaftlich relevanten Zielen bei, indem es das Wissen über Menschenrechte durch den Zugang zur Rechtsprechung fördert.

Darüber hinaus fungiert der HRP als Anwalt in dem Sinne, dass er einen in einfacher Sprache formulierten Sachverhalt in eine Rechtsfrage umwandeln kann (Rechtsrecherche), mögliche Verstöße feststellen kann (Urteil) und eine Verbindung zur relevanten Rechtsprechung herstellen kann (Rechtsrecherche). Dies geschieht sofort und ohne Kosten für den Nutzer.

Ausführliche Darstellung der Ergebnisse

Welche konkreten Ergebnisse hast Du erzielt? Konnten alle Meilensteine erreicht werden? Welche zusätzlichen Erkenntnisse hast Du aus der Projektarbeit gewonnen, auch im Hinblick auf die Begleitung durch die Open Knowledge Foundation?

Mit dem Projekt HRP haben wir mehrere konkrete Ergebnisse erzielt. Dazu gehören:

- Erstellung einer Vielzahl von Datensätzen, die es ermöglichten, LLMs auf die Sprache der EMRK abzustimmen, wodurch einfache sprachliche Aussagen in die vom Gericht verwendete Sprache umgewandelt werden können;
- Training eines Multilabel-Klassifikators für jede einzelne Menschenrechtsverletzung, der unglaublich hohe Genauigkeitsraten bei Vorhersageaufgaben erreicht;
- Entwicklung eines LLM-Systems zur Extraktion von Bezeichnungen aus Urteilstexten, wenn die zugrunde liegenden Metadaten nicht vertrauenswürdig sind;
- Effiziente Kombination mehrerer maschineller Lernwerkzeuge und dekolonialer Designprinzipien, um ein effektives und substanzielles Verständnis von kompliziertem Fachwissen zu ermöglichen;
- Identifizierung und Kennzeichnung von Problemen in den Datensätzens in HUDOC gegenüber den zuständigen Behörden;
- Entwicklung einer Website, die mit dem Tool „Colour Check Analyser“ auf Zugänglichkeit für sehbehinderte Nutzer erfolgreich getestet wurde und alle Anforderungen erfüllt.
- Entwicklung einer Website, die für Screenreader geeignet ist;
- Entwicklung einer Website in einer Sprache, die für Kinder unter 12 Jahren geeignet ist; und
- Es wurde eine juristische Bewertung vorgenommen, die bestätigt hat, dass das Projekt keine Rechtsberatung darstellt, und auf deren Grundlage die Disclaimer auf der Website erstellt wurden.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Welcher Nutzen ergibt sich für die Zielgruppe aus den Ergebnissen Deines Projekts? Welche weitergehenden Effekte ergeben sich aus der Open-Source-Stellung der Ergebnisse? Gibt es Ideen für die Weiterentwicklung Deiner Lösung und Pläne zu deren Umsetzung?

Hat die Arbeit in dem Projekt Euch in Eurer persönlichen, fachlichen Weiterentwicklung unterstützt?

HRP kann mehr als 830 Millionen Menschen in den Rechtssystemen der Mitgliedsstaaten des Europarates erreichen. Unsere Zielgruppen sind marginalisierte Gruppen, Rechtsorganisationen und ForscherInnen. Statt sich auf zukünftige Entwicklungen zu konzentrieren, liegt der Schwerpunkt jetzt auf einer robusten ETL-Pipeline, die täglich Daten abruft und die zugrundeliegende Funktionalität kontinuierlich aktualisiert.

Durch den Issue Identifier, den Predictor und die Search können NutzerInnen auch ohne juristische Vorkenntnisse relevante Rechtsprechung finden, ihre rechtlichen Probleme verstehen und erkennen, welche Rechte möglicherweise verletzt wurden. Ein einfaches und kostenloses Tool, das Menschen hilft, ihre Rechte besser zu verstehen und selbstständig Informationen zu finden.

Der Open-Source-Status von HRP ermöglicht es NutzerInnen aus anderen Bereichen, die Codebasis zu kopieren und an ihre eigenen Rechtsprechungsdatenbanken anzupassen, um die Ergebnisse

kontextbezogen zu reproduzieren. Wir hoffen, dass Gerichte auf der ganzen Welt auf diese Weise ihre Rechtsprechung und rechtsbasierte Informationen besser zugänglich machen können.

Auf professioneller Ebene bin ich mir nun der diskriminierenden Auswirkungen von LLMs in ihrer Anwendung und des Klassendilemmas bewusst, das sich aus den Kosten und dem Wissen ergibt, das für eine effektive Nutzung und Schulung von LLMs erforderlich ist.

Persönlich sehe ich jedoch ein großes Potenzial in der Nutzung von maschinellem Lernen als Navigationswerkzeug, das durch effektives Design Menschen in die Lage versetzt, ihre Rechte.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Gab es Arbeiten bzw. Lösungsansätze, die nicht weiter verfolgt wurden? Was waren die Hintergründe, und wie bist Du alternativ vorgegangen?

Der ursprüngliche Ansatz bestand darin, ein T5-Modell zu trainieren, um einfache Sprache in Gerichtssprache zu übersetzen, was jedoch zu schlechten Ergebnissen führte. Daraufhin ging man dazu über, ein 8B Llama 3.1 Instruct-Modell zu verfeinern. Dies erwies sich als effizient, da das Feintuning mit LORAX-Adaptern durchgeführt wurde, die sich leicht teilen und auf die meisten LLM-APIs anwenden lassen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Wo können sich Interessenten detailliert über Deine Projektergebnisse informieren (z.B. Webseite, GHitHub, Veröffentlichungen)?

Projekt-Website: www.my-rights.info/prediction/

Github: https://github.com/Viraaj-A/My-Rights

Tech Solutions for Human Rights: www.tshr.io

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Gab es im Projektverlauf Ereignisse, die eine Anpassung der Planung erforderlich machten – z.B. Mehr- oder Minderaufwand bei der Bearbeitung von Teilaufgaben?

Das größte Hindernis waren die bereits beschriebenen Probleme mit der Datenqualität in HUDOC, die einen unverhältnismäßig hohen Zeitaufwand für die Datenbereinigung und den Datenimport erforderten. Die Lösung dieses Problems hat jedoch auch dazu geführt, dass sich die Nutzerinnen und Nutzer von HRP nun auf die dargestellten Informationen verlassen können.

Darüber hinaus war die genaue Identifizierung der automatisierten Systeme zur Extraktion qualitativer variabler Daten aus den Urteilen durch die Verwendung von LLMs sehr zeitaufwändig.

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 Entwicklung wurde durch die Zusammenarbeit mit HUDOC stark beeinträchtigt, da das Management nicht auf erkannte Probleme reagierte, mehrere Monate für Antworten benötigte und auf keine E-Mail inhaltlich antwortete. Da alle Daten von HUDOC stammen, wirkten sich diese Probleme mit der Datenqualität negativ auf die Produktionsgeschwindigkeit aus.