Richtlinie zum „Software-Sprint“

Spot The Bot

Schlussbericht

Zuwendungsempfänger:

Mark Wernsdorfer

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS23S46 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Motivation

Die Motivation hinter dem Projekt "Spot The Bot" speist sich aus einer Besorgnis über die zunehmende Verbreitung KI-generierter Fehlinformationen und deren potenzielle Auswirkungen auf die öffentliche Meinung und das Vertrauen in digitale Medien. Als Entwickler mit einer starken Grundlage in KI-Technologien und einem akademischen Hintergrund in Philosophie, bin ich zutiefst interessiert an der Schnittstelle zwischen Technologie, Ethik und Gesellschaft. Die negativen Auswirkungen von Fehlinformationen sollten bekämpft und gleichzeitig die digitale Kompetenz und das kritische Denkvermögen der Öffentlichkeit gestärkt werden. Durch "Spot The Bot" möchte ich eine Plattform schaffen, die es NutzerInnen ermöglicht, die Fähigkeiten zu entwickeln, die notwendig sind, um KI-generierte Inhalte von authentischen Informationen zu unterscheiden.

Problemstellung

Das Kernproblem, das "Spot The Bot" angeht, ist die Schwierigkeit, KI-generierte Fehlinformationen von echten Informationen zu unterscheiden. Diese Fähigkeit ist entscheidend, um das öffentliche Vertrauen in digitale Medien zu wahren und die negativen Auswirkungen von Desinformation zu minimieren. Die rasante Entwicklung von KI-Technologien hat die Erzeugung überzeugender und realistischer Inhalte ermöglicht, was dieses Problem noch verstärkt.

Das Projekt zielt darauf ab, diese Herausforderung durch ein interaktives und bildendes Spiel anzugehen, das die NutzerInnen in die Lage versetzt, kritisch über die Inhalte nachzudenken, mit denen sie täglich konfrontiert werden.

Geplante Vorgehensweise und Meilensteine

Die Umsetzung von "Spot The Bot" folgte einem zweistufigen Ansatz: Zuerst die Entwicklung der technischen Infrastruktur, gefolgt von der Schaffung des eigentlichen Spiels. Die wichtigsten Meilensteine waren:

1. **Technische Grundlagen:** Etablierung einer robusten Backend-Infrastruktur mit Python und FastAPI, sowie die Einbindung von KI-Sprachmodellen für die Texterzeugung. Das Frontend wird unter Verwendung von NiceGui entwickelt, um eine intuitive und zugängliche Benutzeroberfläche zu gewährleisten.
2. **Spielmechanik und Benutzererfahrung:** Ausarbeitung der Spielregeln, des Bewertungssystems und der Nutzerinteraktion, um eine engagierte und lehrreiche Erfahrung zu bieten. Dazu gehört die Entwicklung eines Bewertungssystems, das SpielerInnen nach ihrer Fähigkeit, KI-generierte Inhalte zu identifizieren, einstuft und ein kollaboratives Umfeld fördert, in dem Spieler voneinander lernen können.
3. **Integration und Tests:** Umfassende Tests der Spielmechanik und der Systemintegration, um sicherzustellen, dass die Plattform stabil läuft und eine ansprechende Nutzererfahrung bietet. Dies umfasst auch die Einbindung des Feedbacks von NutzerInnen zur kontinuierlichen Verbesserung des Spiels.
4. **Launch und Iteration:** Die Veröffentlichung des Spiels als progressive Webapp und die Durchführung von Nutzertests, um Feedback zu sammeln und das Spiel basierend auf Nutzererfahrungen zu verfeinern.

Durch diesen Ansatz wurde ein Werkzeug geschaffen, das nicht nur bildet, sondern auch unterhält, und damit einen nachhaltigen Beitrag zur Bekämpfung von KI-generierten Fehlinformationen leistet.

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

Zielgruppe und Benefit

Die ursprüngliche Zielgruppe von „Spot The Bot“ umfasste hauptsächlich Fachleute wie Journalisten, Content-Moderatoren, Social-Media-Manager und Public-Relations-Experten, die beruflich mit der Identifizierung von Fehlinformationen konfrontiert sind. Erkenntnisse aus den Coachings und durch die Umsetzung und Anwendung der App haben den Fokus mehr auf Communities von Infotainment-Kanälen, Gruppen, die nach spezifischen Informationen suchen, und generell Menschen mit gemeinsamen Informationsinteressen gerichtet.

Diese Anpassung ermöglicht es, ein breiteres Publikum anzusprechen, das täglich mit Informationen interagiert und dabei oft Schwierigkeiten hat, authentischer von falscher Information zu unterscheiden. Indem diesen Gruppen die Werkzeuge und das Wissen an die Hand gegeben werden, um KI-generierte Inhalte von echten Informationen zu unterscheiden, wird ihre Fähigkeit gestärkt, informiert Entscheidungen zu treffen und daraufhin selbst zu einer besser informierten Öffentlichkeit beizutragen. Dies fördert nicht nur die Medienkompetenz, sondern auch das Vertrauen in die verfügbaren Informationen und stärkt die demokratischen Prozesse durch eine besser informierte Bürgerschaft.

Ausführliche Darstellung der Ergebnisse

Welche konkreten Ergebnisse hast Du erzielt?

Im Rahmen des Projekts "Spot The Bot" wurde ein stabiles Verfahren zur Identifikation und Vermittlung erfolgreicher Indikatoren für KI-generierte Texten entwickelt. Dieses Verfahren ermittelt und publiziert selbständig Indikatoren, die es NutzerInnen ermöglichen, gefälschte Informationen von echten zu unterscheiden.

Besonders hervorzuheben ist die Dynamik des Systems, welches so konzipiert wurde, dass es auch mit Texten umgehen kann, die von zukünftigen Sprachmodellen (LLMs) erzeugt wurden. Durch die Integration eines adaptiven Absatzes ist es gelungen, eine Plattform zu entwickeln, die mit dem zukünftigen Stand der Technologie Schritt halten kann.

Konnten alle Meilensteine erreicht werden?

Aufgrund der rasanten Entwicklung im Bereich der Künstlichen Intelligenz musste die Strategie während des Projekts angepasst werden. Ursprünglich sollte eine breite Palette von KI-Modellen für die Generierung von Spielinhalten eingesetzt werden. Letztendlich wurde jedoch nur ein Modell herangezogen, um eine konsistente Qualität und Vergleichbarkeit der Inhalte sicherzustellen, die repräsentativ zeigt, wofür die App eingetzt werden kann und was durch ihren Einsatz zu erwarten ist.

Ein weiterer wichtiger Schritt war die Fokussierung auf spezifische Communities und deren Vokabular. Statt einer generischen Beispielapp unter spotthebot.de lag der Fokus verstärkt auf der Möglichkeit der Individualisierung durch interessierte Communities. Diese Anpassung führte zu einer modulareren Umsetzung als ursprünglich geplant.

Welche zusätzlichen Erkenntnisse hast Du aus der Projektarbeit gewonnen, auch im Hinblick auf die Begleitung durch die Open Knowledge Foundation?

Die Projektarbeit hat deutlich gemacht, wie herausfordernd es ist, KI-generierte Texte zuverlässig zu erkennen, ohne dadurch gleichzeitig ihre zukünftige Erkennung zu erschweren. Unachtsame Datenerhebung kann letztlich dazu führen, dass aus den erhobenen Daten neue Systeme trainiert werden, die durch die ermittelten Indikatoren eben nicht mehr erkannt werden können. Es wurde daher besonderer Wert darauf gelegt, dass einzelne Texte nicht zusammen mit den auf sie angewandten Indikatoren erfasst werden.

Die ursprüngliche Idee, eine generische Plattform zu entwickeln, erwies sich als zu allgemein. Die Coachings haben gezeigt, dass relevante Inhalte und das zur Auszeichnung KI-generierter Texte eingesetzte Vokabular sich stark zwischen verschiedenen Communities unterscheiden. Diese Erkenntnis führte zu einer Neuausrichtung des Ansatzes, hin zu einer stärkeren Personalisierbarkeit.

Die Begleitung durch die Open Knowledge Foundation war in diesem Prozess unschätzbar wertvoll. Coachings haben nicht nur dazu beigetragen, die technische Ausrichtung zu schärfen, sondern auch die Bedeutung von Öffentlichkeitsarbeit für Open-Source-Projekte hervorzuheben. Durch die

Unterstützung der Foundation konnten nicht nur technische, sondern auch strategische Weichenstellungen vorgenommen werden, die das Projekt maßgeblich positiv beeinflusst haben.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Nutzen für die Zielgruppe

Das Projekt "Spot The Bot" richtet sich nun an eine breitere Zielgruppe: Menschen, die sich ihrer Wissenslücken bewusst sind, gezielt nach Informationen suchen und daher besonders anfällig für Fehl- und Desinformationen sind. Der Nutzen für diese Zielgruppe ist vielfältig. Durch das Spiel erlangen die Teilnehmenden ein tieferes Verständnis dafür, wie KI-generierte Inhalte erstellt werden und welche Merkmale sie von menschlich erstellten Inhalten unterscheiden. Dieses Wissen befähigt sie, kritischer mit den Informationen umzugehen, denen sie online begegnen.

Die spielerische Herangehensweise fördert zudem das Engagement und die Motivation, sich mit dem Thema auseinanderzusetzen, was zu einer dauerhaften Kompetenz im Umgang mit digitalen Informationen führen kann.

Effekte der Open-Source-Stellung

Die dem Projekt zugrundeliegende Open-Source-Lizenz ermöglicht es, dass das Projekt beliebig geforkt und in verschiedensten Communities weiterentwickelt werden kann. Dies fördert Innovation und Kollaboration im digitalen Raum, indem es EntwicklerInnen weltweit ermöglicht wird, auf bestehender Arbeit aufzubauen, sie zu verbessern oder anzupassen, um spezifische Bedürfnisse zu erfüllen.

Die Offenheit des Projekts schafft außerdem die Grundlage zur Bildung einer *eigenen* Community, die gemeinsam an der Anpassung und Verbesserung von Methoden zur menschlichen Erkennung KI-generierter Inhalte arbeitet. Darüber hinaus erhöht die Open-Source-Natur die Transparenz und Vertrauenswürdigkeit des Projekts, was besonders in einem Bereich wie der Bekämpfung von Fehlinformationen von großer Bedeutung ist.

Weiterentwicklung der Lösung

Hinsichtlich der Weiterentwicklung von "Spot The Bot" wurden bereits erste Schritte unternommen, um Kontakte zu Online-Communities zu knüpfen. Je nach Rückmeldung und verfügbaren Finanzierungsmöglichkeiten ist eine erste attraktive Weiterentwicklung die Anpassung des Spiels an die Bedürfnisse einer spezifisch ausgewählten Community.

Dies könnte beispielsweise eine Version für schulische Bildungseinrichtungen sein, die Lehrkräfte dabei unterstützt, SchülerInnen im Umgang mit digitalen Medien zu schulen, oder eine Anpassung für journalistische Zwecke, um die Fähigkeit zur Unterscheidung von echten und gefälschten Nachrichten zu schärfen.

Persönliche und fachliche Weiterentwicklung

Die Arbeit an "Spot The Bot" hat nicht nur meine Erfahrung im Open-Source-Ökosystem vertieft, sondern auch wertvolle Kontakte zu Experten in diesem Bereich geschaffen. Durch das Projekt habe ich die Art und Weise, wie Open-Source-Projekte präsentiert und gemeinschaftlich weiterentwickelt werden können, besser verstanden. Diese Erfahrungen haben mein Fachwissen erweitert und meine Fähigkeiten in der Projektentwicklung und im Community-Management verbessert. Die Auseinandersetzung mit der Dynamik des Open-Source-Bereichs hat mir neue Perspektiven auf die Potenziale und Herausforderungen digitaler Innovationen eröffnet. Diese Erfahrungen haben meine Fähigkeiten in der Softwareentwicklung gestärkt und mein Verständnis für die Dynamik und Potenziale offener Entwicklungsprojekte vertieft. Durch die Auseinandersetzung mit Feedback habe ich zudem gelernt, wie wichtig es ist, offen für Anregungen zu sein und wie wertvoll die Kollaboration mit anderen Entwicklerinnen und Nutzerinnen sein kann.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

In den frühen Phasen des Projekts "Spot The Bot" wurden verschiedene Ansätze und Funktionen in Erwägung gezogen, die darauf abzielten, Interaktivität und Engagement der NutzerInnen zu erhöhen. Ein spezifischer Bereich betraf die Entwicklung eigener Community-Funktionen. Dazu gehörten ein umfassendes Freundschafts- und Einladungssystem, Nachrichtenfunktionen, persönliche Ranglisten und andere soziale Interaktionsmöglichkeiten innerhalb des Spiels.

Diese initialen Überlegungen zielten darauf ab, eine tiefere soziale Dimension und eine stärkere Vernetzung zwischen den SpielerInnen zu schaffen. Diese Funktionen sollten nicht nur die Spielerbindung steigern, sondern auch Austausch und Diskussion über KI-generierte Inhalte fördern.

Im Verlauf der Konzeptentwicklung stellte sich jedoch heraus, dass die Nachfrage nach solchen umfangreichen Funktionen durch potentiell interessierte Communities nicht sichergestellt ist. Insbesondere ohne klare Evidenz, dass diese Features die Kernziele des Projekts unterstützen würden, nämlich die Bildung und Sensibilisierung im Umgang mit KI-generierten Falschinformationen, erschienen diese Ansätze als zu riskant und wurden nicht weiter verfolgt.

Stattdessen wurde der Fokus darauf gelegt, eine direkte Auseinandersetzung mit KI-generierten Inhalten zu fördern. Durch diese Fokussierung konnten Ressourcen effizienter eingesetzt und sichergestellt werden, dass die Entwicklung des Projekts eng mit seinen Bildungszielen und dem Bedarf der Zielgruppe verknüpft bleibt.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Beispiel und Testen des Spiels

Besucher können direkt auf spotthebot.app eine interaktive Demo des Spiels erleben. Diese Webseite bietet einen realitätsnahen Einblick in die Spielmechanik, das Ranglistensystem und die Benutzerführung, die speziell darauf ausgelegt ist, die Erkennung von KI-generierten Fehlinformationen zu trainieren.

GitHub-Repository

Für technisch interessierte Nutzer und Entwickler ist das vollständige Projekt einschließlich Quellcode, Dokumentation und Entwicklungsleitfäden auf GitHub verfügbar. Dies ermöglicht Einblicke in die technische Umsetzung, die verwendeten Algorithmen und die Architektur des Projekts. Darüber hinaus bietet es die Möglichkeit zur Kollaboration, zur Diskussion von Verbesserungen und zur Weiterentwicklung des Projekts.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Ursprünglich lag der Nutzergruppenfokus auf Multiplikatoren wie Journalisten, Content-Moderatoren, Social-Media-Managern und Public-Relations-Experten. Im Laufe der Entwicklung wurde das Spiel mehr auf durch Des- und Misinformation besonders gefährdete Gruppen ausgerichtet. Diese Entscheidung führte zu signifikanten Anpassungen im Design und Layout der Webapp.

Um dieser neuen Zielgruppe gerecht zu werden, musste das Frontend so gestaltet werden, dass es leicht an verschiedene Communities mit individuellem Branding / CI anpassbar ist. Daher mussten Arbeitsressourcen umverteilt werden, um eine flexible Benutzeroberfläche zu entwickeln, die es ermöglicht, die Präsentation relativ schnell und unkompliziert anzupassen, ohne Kernelemente des Systems dabei verstehen oder verändern zu müssen.

Außerdem hat die Planung von Datenerfassung und -haltung mehr Zeit in Anspruch genommen als geplant, um sicherzustellen, dass die erfassten Daten nicht eingesetzt werden können, um KIs für die Erzeugung menschlich wirkender Texte zu trainieren.

Diese Anpassungen hatten Auswirkungen auf den Arbeits- aber keine auf den Kostenplan. Der Mehrbedarf an Ressourcen für die Entwicklung einer flexiblen und leicht anpassbaren Benutzeroberfläche sowie die Neugestaltung der Spielinhalte konnte durch unerwartete Effizienzsteigerungen in anderen Bereichen ausgeglichen werden, sodass die Gesamtstundenzahl und das Budget eingehalten werden konnten.