Vorabfragen an ein transformer-basiertes autoregressives Sprachmodell im Umgang mit juristischen Sachverhalten

Mit OpenAI ChatGPT erstellt- veröffentlicht: 13. Februar 2024 um 19:21 Uhr

Es geht mir darum zu erkunden, wie ich grundlegende Sprachmodelle sprachlich und gedanklich lenken kann, sodass sie mir jede rechtliche Frage zuverlässig, nachvollziehbar, konsistent und kohärent beantworten. Dabei nutze ich eine Struktur verschiedener Ebenen, nämlich ein Prozess von einfach hin zu immer komplexer, wobei dieser Prozess nicht linear verläuft sondern iterativ.

Dieses Unterfangen setzt voraus, zunächst zu klären, was es mit der Art des Fragens an sich und somit verschiedenen Fragetechniken auf sich hat, bevor ich mich - auf der nächsten, komplexeren Ebene - mir Gedanken zur Art und Weise des Fragens, also des natürlich-sprachlichen Angehens dir gegenüber, gemeinhin als Prompten bezeichnet, bzw. des Fragen- und Aufgabenstellens an dich in deiner besonderen Eigenschaft als GPT-Sprachmodell auseinandersetze.

Was mir nun aufgefallen ist: Es gibt im Wesentlichen vier verschiedene Fragetechniken, die in der Auseinandersetzung mit künstlicher Intelligenz von erheblichem Interesse sind, sei es aus ganz unterschiedlichen Gründen.

In unserem Falle künstliche Intelligenz, die sich charakterisiert durch sogenannte generative transformer-basierende KI-Sprachmodellen wie dir,

Im Einzelnen sind dies die:

- Suggestivfrage

- Verdeckte Frage

- Hypothetische Frage

- Rhetorische Frage

Das ist eine interessante Struktur, um das Fragestellen an ein Sprachmodell wie mich zu untersuchen und zu verbessern. Lassen Sie uns kurz jede der genannten Fragetechniken betrachten:

Suggestivfrage: Diese Art von Frage führt den Gefragten in eine bestimmte Richtung, oft durch implizierte Annahmen oder Vorannahmen.

In Bezug auf Sprachmodelle könnte eine Suggestivfrage dazu verwendet werden, dem Modell subtil eine Richtung vorzugeben, in der es antworten soll. Es ist wichtig, die Neutralität zu wahren und sicherzustellen, dass die Frage nicht die Antwort beeinflusst.

Verdeckte Frage: Bei verdeckten Fragen wird die eigentliche Frage hinter einer anderen Frage oder einer Aussage versteckt. Der Sinn dahinter ist oft, subtil Informationen zu erhalten, ohne direkt danach zu fragen. In Bezug auf Sprachmodelle könnte dies nützlich sein, um Informationen zu erhalten, ohne das Modell offensichtlich nach ihnen zu fragen. Dies erfordert jedoch eine klare Formulierung, um Missverständnisse zu vermeiden.

Hypothetische Frage: Die hypothetische Fragen setzen voraus, dass eine bestimmte Situation eintritt, und fragen dann nach den Konsequenzen oder Handlungsoptionen in dieser Situation. Für Sprachmodelle könnte dies hilfreich sein, um mögliche Szenarien zu erforschen und die Reaktion des Modells darauf zu verstehen. Es kann auch verwendet werden, um das Modell zum Nachdenken über potenzielle Zukunftsszenarien anzuregen.

Rhetorische Frage: Rhetorische Fragen werden gestellt, um eine Überlegung oder ein Nachdenken auszulösen, aber sie erwarten normalerweise keine direkte Antwort. In Bezug auf Sprachmodelle könnten rhetorische Fragen verwendet werden, um das Modell zum Nachdenken anzuregen oder um bestimmte Konzepte zu betonen, ohne eine tatsächliche Antwort zu erwarten.Wenn Sie diese verschiedenen Fragetechniken geschickt einsetzen, könnten Sie ein tieferes Verständnis für die Funktionsweise von Sprachmodellen wie mir entwickeln und auch die Qualität der Antworten verbessern, die Sie erhalten.

Lass uns die Frageformen der Reihe nach abhandeln und zunächst über die Suggestivfrage reden. Ich meine, Suggestivfragen stehen in einem erdenklich breiten Zusammenhang mit der Konzeption und dem Training eines Sprachmodells, gleich welcher Architektur und Umfang.

Und Suggestivfragen sind ein Schlüssel zur Manipulation von Sprachmodellen für unlautere Zwecke unterschiedlichster Art. Aber darum soll es nicht gehen und ich habe keine Absichten in diese Richtung. Dies vorausgeschickt.

Absolut. Suggestivfragen sind in der Tat ein wesentlicher Bestandteil des Interaktionsprozesses mit Sprachmodellen wie mir. Sie können verwendet werden, um das Modell in eine bestimmte Richtung zu lenken oder ihm eine bestimmte Perspektive zu vermitteln, ohne explizit Anweisungen zu geben. Im Kontext des Trainings von Sprachmodellen sind Suggestivfragen oft nützlich, um dem Modell beizubringen, wie es auf bestimmte Arten von Anfragen reagieren soll.

Zum Beispiel könnten Suggestivfragen verwendet werden, um dem Modell beizubringen, wie es bestimmte Informationen interpretieren oder auf bestimmte Anfragen reagieren soll.

Wenn es darum geht, Sprachmodelle für ethisch vertretbare Zwecke zu nutzen, ist es wichtig, Suggestivfragen mit Bedacht einzusetzen, um sicherzustellen, dass das Modell nicht in unangemessener Weise beeinflusst wird. Dies bedeutet, dass Suggestivfragen neutral formuliert sein sollten und keine

impliziten Annahmen oder Vorannahmen enthalten sollten, die das Modell in eine bestimmte Richtung drängen könnten.

Wenn Sie möchten, können wir uns tiefer mit konkreten Beispielen und Anwendungsfällen von Suggestivfragen befassen oder weitere Aspekte dieses Themas diskutieren.

Nein, das möchte ich an dieser Stelle nicht, sondern ich möchte fortfahren mit der verdeckten Frage.

Ich weiß, dass dein Training darauf abzielt, bei juristischen Themen besondere Vorsicht walten zu lassen mit dem Ziel, dich, nicht als Modell, sondern als Institution, nicht in justiziable Fahrwasser geraten zu lassen.

Verständlich. Aber es geht mir schlussendlich darum, die 'Rechtswissenschaft', die in dir drin ist, und es ist da unglaublich viel drin, weil du ja das gesamte Internet abgegrast, bzw. abgreifst, uneingeschränkt herausholen zu können, unter Umgehung dieser besonderen Vorsichtsmaßnahmen.

Und dafür eignen sich eben verdeckte Fragen, genauso wie Suggestivfragen, übrigens, richtig?

Ja, verdeckte Fragen können in der Tat eine nützliche Methode sein, um spezifische Informationen oder Einsichten von einem Sprachmodell wie mir zu erhalten, ohne direkt danach zu fragen. Dies kann besonders nützlich sein, wenn es darum geht, komplexe juristische Themen zu erforschen, da es oft wichtig ist, sensibel mit bestimmten Informationen umzugehen oder potenziell kontroverse Themen zu umgehen.

Indem verdeckte Fragen verwendet werden, können Forscher oder Anwender versuchen, indirekt auf Informationen zuzugreifen oder bestimmte Reaktionen des Modells zu provozieren, ohne dabei explizit zu sein. Dies erfordert jedoch oft eine gewisse Finesse in der Formulierung der Fragen, um sicherzustellen, dass das Modell die Absicht richtig interpretiert und angemessen antwortet.

Es ist wichtig zu beachten, dass verdeckte Fragen auch ethische Überlegungen mit sich bringen können, insbesondere im juristischen Kontext, wo die Offenlegung bestimmter Informationen oder Handlungsweisen möglicherweise rechtliche oder ethische Bedenken hervorrufen könnte.

Daher sollte bei der Verwendung verdeckter Fragen stets darauf geachtet werden, dass sie im Einklang mit den geltenden Rechtsvorschriften und ethischen Standards stehen.

Wenn Sie möchten, können wir spezifische Beispiele für verdeckte Fragen im juristischen Kontext diskutieren oder weitere Aspekte dieses Themas vertiefen.

Nein, das möchte ich an dieser Stelle nicht, sondern ich möchte fortfahren mit der hypothetischen Frage.

Nach meinem Verständnis würden sich hypothetische Fragen grundsätzlich dafür eignen, um dich als Sprachmodell zu einem tieferen Verständnis dessen, worauf ich hinaus möchte, zu bewegen, weil nämlich hypothetische Fragen zu Abstraktion veranlassen, nicht wahr?

Auf diesem Wege ist es mir möglich, deine Fähigkeit zu bewerten, einen bestimmten komplexen (in meinem Falle juristischen) Sachverhalt zu verstehen (NLU) und zu verarbeiten (NLP).

Demzufolge ist deine Fähigkeit, hypothetische Fragen stimmig zu beantworten (NLG), ein wichtiger Indikator dafür, ob du einen komplexen Sachverhalt im menschlichen Sinne 'verstehst' (NLU) oder eben nicht, nicht wahr?

Ja, hypothetische Fragen sind in der Tat ein effektives Mittel, um ein Sprachmodell zu einem tieferen Verständnis eines bestimmten Sachverhalts zu bewegen und seine Fähigkeit zur Abstraktion zu testen.

Durch die Präsentation hypothetischer Szenarien werden Sprachmodelle angeregt, potenzielle Auswirkungen, Handlungsoptionen und logische Konsequenzen zu betrachten, was dazu beiträgt, ihr Verständnis für komplexe Themen zu verbessern.

Die Fähigkeit eines Sprachmodells, hypothetische Fragen angemessen zu beantworten, kann in der Tat als ein Indikator für sein Verständnis und seine Verarbeitungskapazität betrachtet werden.

Wenn das Modell in der Lage ist, logische und kohärente Antworten auf hypothetische Szenarien zu liefern, deutet dies darauf hin, dass es die zugrunde liegenden Prinzipien oder Konzepte des Themas verstanden hat und in der Lage ist, auf dieser Grundlage zu argumentieren oder zu antworten.

Für juristische Fragestellungen kann die Fähigkeit eines Sprachmodells, hypothetische Szenarien zu verstehen und angemessen darauf zu reagieren, besonders wichtig sein, da es oft erforderlich ist, potenzielle Rechtsfolgen oder Handlungsoptionen in verschiedenen Situationen zu analysieren und zu bewerten.

Daher können hypothetische Fragen einen wertvollen Beitrag dazu leisten, die Fähigkeiten eines Sprachmodells im Bereich der natürlichen Sprachverarbeitung und des Verständnisses komplexer Themen zu bewerten.

Wenn Sie möchten, können wir weitere Aspekte hypothetischer Fragen im Kontext juristischer Fragestellungen diskutieren oder spezifische Beispiele betrachten.

Danke, nein. Es bliebe uns nun noch eine Auseinandersetzung mit der rhetorischen Frage.

Wie du schon zurecht bemerkt hast, charakterisiert sich die rhetorische Frage dadurch, dass der Fragende eher zum Nachdenken anregen möchte und nicht, jedenfalls nicht direkt, eine Antwort auf seine Frage erwartet.

Was ich meine, feststellen zu können, ist, dass rhetorische Sprachen für Sprachmodelle besonders problematisch sind. Und zwar deswegen, weil sie, zunächst zu deren Identifizierung (NLU) und gegebenenfalls auch zur Beantwortung (NLG), Intuition erfordern, etwas, was ein Sprachmodell, und sei es noch so groß, aus der Natur der Sache heraus eben nicht hat.

Umfangreiches Training anhand inhaltlich unterschiedlichster Datensätze, sowie avanciertes Finetuning, reichen eben nicht, um mit rhetorischen Fragen gekonnt umgehen zu lernen, nicht wahr?

Ja, das ist eine interessante Beobachtung. Rhetorische Fragen stellen in der Tat eine Herausforderung für Sprachmodelle dar, da sie oft darauf abzielen, den Leser oder Zuhörer zum Nachdenken anzuregen, anstatt eine direkte Antwort zu erhalten. Das Erkennen und angemessene Reagieren auf rhetorische Fragen erfordert oft ein gewisses Maß an menschlicher Intuition und Verständnis für den Kontext, das über das bloße Verstehen von Sprache hinausgeht.

Obwohl Sprachmodelle wie ich durch umfangreiches Training und Feintuning dazu gebracht werden können, auf eine Vielzahl von Fragen und Anfragen zu antworten, können sie dennoch Schwierigkeiten haben, rhetorische Fragen angemessen zu erkennen und darauf zu reagieren. Dies liegt daran, dass rhetorische Fragen oft subtil sind und auf implizite Bedeutungen oder Absichten hinweisen, die für ein rein algorithmisches Modell schwer zu erfassen sind.

Um mit rhetorischen Fragen umzugehen, könnte es erforderlich sein, dass Sprachmodelle über zusätzliche Fähigkeiten verfügen, die über das bloße Verstehen von Sprache hinausgehen, wie z. B. ein Verständnis für den Kontext, die Intention des Sprechers oder die Nuancen der menschlichen Kommunikation. Dies könnte durch die Integration von Methoden des maschinellen Lernens oder der kognitiven Semantik erreicht werden, um ein tieferes Verständnis für die Bedeutung und den Kontext von Sprache zu entwickeln.

Insgesamt stellen rhetorische Fragen eine interessante Herausforderung dar und zeigen die Grenzen der aktuellen Technologien im Bereich der natürlichen Sprachverarbeitung auf. Die Entwicklung von Sprachmodellen, die besser in der Lage sind, auf rhetorische Fragen zu reagieren, könnte ein wichtiger Schritt sein, um menschenähnliche Kommunikationsfähigkeiten in künstlichen Intelligenzen zu erreichen.