

Transkript P1

Interviewer

Ja, also wir waren bei Level nochmal zurück zu Level 2. Was kam denn da so bei dir raus? Was hast du gemacht mit dem mit dem Teil. Was hast du so für Sachen angeguckt. Was hättest du gesagt, wie viel ist da rausgekommen?

Proband

Ja genau, also ich habe mir ein Waschbecken angesehen, weil ich dann quasi auch zum Kontrollieren mit den anderen Versionen gemacht, die Ergebnisse auch ein bisschen vergleichen zu können. Genau, ich hab mir da dann ein Bild ausgesucht er hat dann allgemein gesehen, was auf jeden Fall an möglichen Tasks wäre, also Spiegel, Waschbecken et cetera. Der hat jetzt nicht unbedingt geguckt, ob sie dreckig sind, der hat einfach nur so gesagt, mach dir mal bitte sauber und dann relativ ausführlich zum Waschbecken und zum Spiegel was geschrieben. Das ganz, also ich weiß nicht, ob das hier jetzt schon reinpasst, aber er hat halt vom Output dann relativ so viel geschrieben, dass das die Textbox, die man bekommen hat, ein bisschen gesprengt hat. Und wenn man dann halt auf die Textbox drauf klickt und da reinseht, dann kann man da auch nicht so gut runter scrollen. Dann auch dieses Texteingabefeld die ganze Zeit drüber. Das ist ein bisschen schwierig, da wegzuklicken, dass man dann auch alles sehen. Sehen können. Genau. Aber an sich hat er das gedacht, was ich jetzt erwarten würde, wenn man sich das Waschbecken anguckt, was man da jetzt auch sauber machen soll und was nicht.

Interviewer

Aber eine Sache, die du dir noch angemerkt hast, ist auf jeden Fall. Aber das hat jetzt nichts mit dem Modi zu tun, sondern allgemein. Dass er das Waschbecken putzen wollte. Auch wenn es eigentlich schon geputzt worden ist, oder? Ja, also quasi. Wenn du jetzt, wir stellen uns vor, du hast so ne so ne App. Die dir eben diese Tasks erstellen soll. Würdest du dann also würdest du dann per se wollen, dass der das selbstständig erkennen sagt? Nee, das sieht wunderbar aus, du brauchst es nicht mehr putzen und sagst du ja, ich weiß ja selber, ob ich es putzen will. Nicht, und deswegen mache ich ein Bild oder halt nicht.

Proband

Wahrscheinlich kommt es dann grundsätzlich darauf an, wer die Nutzen soll.

Interviewer

Ja, ich frage dich jetzt.

Proband

Und jetzt mich persönlich. Ich find's okay, wenn der einfach alles aufschreibt und dann kann ich aber selbst interpretieren, was davon jetzt sinnvoll ist oder nicht. Dann ist natürlich die Frage, was genau der Nutzen dann für mich wäre.

Interviewer

OK, genau. Dann dann, dann gehen wir mal weiter so der dritte Modus, der dritte Modus war ja der dann mit dem mit dem Antippen.

Proband

Mhm.

Interviewer

Wo du mir macht das Bild und dann gibt er. So Vorschläge.

Proband

Also die. Ergebnisse waren also die Vorschläge waren passend. Ich hatte das jetzt glaub ich 3 oder 4 mal ausprobiert, da waren jetzt keine total abstrusen Vorschläge dabei. Und die Ergebnisse waren auch ähnlich wie bei den vorigen. Was mir aufgefallen ist, weder der noch der andere haben sich jetzt die. Lesen oder sowas angeguckt oder die Wand, das ist dann die Frage, ob das schon drin ist, dass er die nicht erkennen soll oder ob er die eigentlich erkennen sollte, aber dass das mir aufgefallen, dass er die nie angesprochen hat.

Interviewer

Genau also die Idee dahinter ist tatsächlich, dass es ihn nicht anspricht. Das wär eher n Manko gewesen, sondern wenn du den Cleaning Task hast per se. Wenn du es drin haben möchtest, dann hätt ja potenziell den Text gehabt oder du hättest Dinge auswählen können, dass er das macht. Aber per se sollte er immer quasi das Hauptobjekt so sieht auf dem Bild putzen. Deswegen hat er auch weiß nicht ob du irgendein Test mal damit gehabt hast oder nicht, wenn du auf einmal so ganz viele kleine Klasse Sachen hast. Ob er jetzt. Und was ja davon jetzt tatsächlich, was jetzt tatsächlich die Aufgabe, weil es ist ja auch, wenn. So n Mensch muss ja auch immer n bisschen zurückdenken und vorstellen. OK, ich bin jetzt nicht n Computer, mach das sondern tatsächlich n Mensch macht es und du gibst denen nur n Bild und auf dem Bild sind einfach 17 Sachen drauf und woher soll der wissen was man da jetzt putzen soll? Was sind die Aufgaben wenn du absolut keine weitere Informationen

Interviewer

Also wenn er jetzt wie er bei dir sagt, die Sachen eigentlich relativ gut erkannt hat und gesagt hat, was man da putzt, was und so. Hier schon mal. Also hat.

Proband

Jetzt nie gesagt. Hey Putz mal bitte den Schornstein oder sowas. So ne Sachen sind tatsächlich nicht passiert.

Interviewer

Genau. Also du hättest jetzt auch gesagt, dass von der Information, die du herbekommen hast und auch von dem System an sicher so wie du es jetzt benutzt hast, war es jetzt eigentlich schon relativ robust und ist jetzt nicht irgendwie ausgeschweift oder hat halluziniert oder irgendwelche komischen Sachen von sich gegeben.

Proband

Also er hat jetzt keine komischen Sachen ausgegeben, aber manchmal hat er sehr viel geschrieben und manchmal relativ wenig. Da konnte man nicht unbedingt erkennen, woran es lag.

Interviewer

Ich wollt Grad sagen und da hast du auch nicht gemerkt, woran es lag und es lag auch nicht irgendwie zwischen den verschiedenen Leveln, dass er da verschiedene verschieden große Informationen, Blöcke rausgeworfen, das konnte man an sich eigentlich immer wieder bei ganz unterschiedlichen.

Proband

Von den Leveln beobachten es schien relativ willkürlich. Ich hab zumindest kein muster erkannt.

Interviewer

Wunderbar. Ja, dann gehen wir doch kurz zum letzten Modus, da wo man nur ein Bild macht. Ich vermute was davon. Wie haben wir ja gerade angesprochen, die Informationen sind da eigentlich überall relativ gleich geblieben, auch die Anzahl und wie viele gemacht hat, wobei du ja meintest, beim Dritten hat er manchmal ein bisschen mehr beschrieben. Okay. Gut, dann noch mal kurz zu zu, nur damit es auch noch vermerkt ist. Das erste Level, klar, da muss man einfach nur selber schreiben. Hier geht es mir jetzt darum, wenn. Aufgefordert worden wärst. Du machst jetzt bitte ne Aufgabe für dieses Ding auf deinem Handy. Wieviel hättest du denn per se geschrieben? Du hast jetzt hier die Vergleiche, wieviel das Ding dir darunter geballert hat, wieviel hättest du denn geschrieben wenn es. Wenn man nicht, wenn man nicht mal, wenn man nicht gefordert hat, sondern es geht hier jetzt wirklich nur um deinen Eigennutzen, erstmal, wieviel hättest du geschrieben und hättest gesagt, ja es passt schon, aber auch musst du auch Bedenken, die Aufgabe könnte, dann müsste es nicht du übernehmen, die könnte auch irgendein anderer übernehmen mit dem Aspekt im Hintergrund.

Proband

Ich am Smartphone ne, dann hätte ich wahrscheinlich einfach nur geschrieben.
Waschbecken saubern. Fragezeichen.

Interviewer

Und das hättest du in die Description überhaupt irgendwas geschrieben.

Proband

Wahrscheinlich nicht.

Proband

Ich bin dabei, da wäre der Titel Description genug, vermute ich mal.

Interviewer

Genau so, weil so geht es zum Beispiel auch mir. Ich hab also ich hab die App, ich hab quasi die n bisschen bessere App da ja verwendet und da war es auch so. Man hat eigentlich also eigentlich nur die Headline gemacht und die Description ja hat man sich ja gedacht, Junge also auf dem Smartphone da jetzt nen halben Text reingeben hab ich auch keinen Bock.

Interviewer

So gut, dann kommen wir mal kurz zu erfordernd also. Gut, auch fordernd bei beim ersten Modi. Gut, es sind 2 Textfelder die da, da weiß man eigentlich was man zu tun hat, wenn man da. Reinschreibt. Also ich hatte.

Proband

Eigentlich. Auch bei den anderen Sachen keinerlei Probleme konnte. Wenn ich rausgekriegt habe ich gesehen wie es gemeint ist. Was jetzt offensichtlich aufgefallen ist, dass ich nicht verstanden habe, wie das erste Level funktioniert.

Interviewer

Ja gut, aber das könnt auch an der Beschreibung von mir liegen.

Proband

Ja, also das wäre mein Feedback. Beim ersten hab ich es anscheinend nicht verstanden, ich hab es aber tatsächlich auch nicht drauf geguckt. Ich hatte jetzt einfach assumed, dass man sich da selbst nen prompt baut, basierend auf den Vorinformationen die ich hatte. Das heißt, dass es vielleicht dann auch nicht unbedingt relevant jetzt für diese Auswertung, weil hier mit voreingenommen reingegangen bin. Deswegen habe ich es mir auch gar nicht angeguckt. Genau, aber bei den anderen Sachen war eigentlich direkt klar, was gemacht werden soll. Da hatte ich keine Probleme.

Interviewer

Okay.

Proband

Und dieses Take Picture und Select Picture war mir auch echt klar, dass mit Select gemeint ist, dass man dann 1 auswählt aus dem aus dem Handyspeicher und Tag das. Und dann halt direkt ein Foto macht über die App.

Interviewer

Ok, wunderbar. Und das war auch so. Ich mein, es sind jetzt nicht so viele Buttons, aber du hast auch quasi sobald du die Sachen benutzt hast und deren Sending gemacht hat. Du warst jetzt auch immer im klaren OK, was habe ich jetzt gerade getan und was passiert jetzt und es war jetzt nicht, dass du was gedrückt hast und dann kam da kam da die Sache oh Gott, wo bin ich jetzt gelandet?

Sprecher

Nur.

Interviewer

Das heißt, du hast dich eigentlich bei allen Sachen relativ wohl gefühlt beim Benutzen und du hast auch bei allen Dingen gesagt, ja, ich weiß 100%, was man hier von mir möchte und so wie das jetzt hier angeordnet ist, da weiß ich auch was was wie man das hier gemeint hat. OK, die einzige Unsicherheit die es gab war, weil man kann ja quasi den Output speichern bzw wenn man das Create new gemacht hat dann speichern kann man ja am Ende das ganze Ding da abspeichern.

Proband

Da weiß ich jetzt nicht unbedingt, hey, wie groß ist das, wo wird das überhaupt gespeichert? Vielleicht wäre da ne Info ganz cool, dass da in Klammern steht, wie groß das ist oder sowas, weil da könnte ja quasi alles sein. Bevor ich mir jetzt alles sicher ist, das jetzt aber nur so n kleines Textfeld da was da drin selbst abgelegt wird oder wie schaut das aus?

Interviewer

Ach so, okay. Ja das wird simple als Text gespeichert.

Interviewer

Wunderbar, dann machen wir aber einfach mal weiter. Genau. Also wenn du das Ding benutzt hast. Wie? Also hat sich das geändert über den Verlauf, wie sicher du warst, was da jetzt rauskommt oder allgemein, wenn du jetzt das System in einem anderen Kontext in dem Higher Level Kontext jetzt irgendwo das ist irgendwie verzahnt in so einem ganz großen System. Wie sehr würdest du sagen, wärst du sicher, dass da ein gescheiter Text

rauskommt, ohne noch mal Extras zu schauen? In der Art: „Ok, ich les mir lieber noch mal durch was da steht“.

Proband

Also grundsätzlich, denke ich nicht, dass LLM ohne menschliche Supervision funktioniert. Ich denke, dass man da immer gucken sollte, was da rauskommt. Aber das ist eher, das ist auch wieder voreingenommen, das ist auch unabhängig jetzt von deiner App, weil ich würde niemals ein Ergebnis aus so einem Modell nehmen und einfach sagen, ja OK, passt schon.

Interviewer

OK, aber jetzt? Wir stellen uns mal vor, das wär jetzt kein LLM gewesen, das wüsstest du jetzt nicht. Du hast nur die Texte gelesen, die da jetzt rausgekommen sind und du wüsstest nicht, wie das passiert ist. Du kennst die Technologie dahinter nicht, sondern also wirklich. Wir nehmen das mal jetzt raus. Sondern wirklich nur das anhand von den Texten, die jetzt rausgekommen sind.

Sprecher

Uff.

Interviewer

Wenn, selbst wenn es jetzt nicht so viele war.

Proband

Würde ich immer noch genauer nachsehen.

Interviewer

Wie okay also das. Wie könnte sich das für dich ändern, dass du. Nicht mehr tust. Oder sagst du einfach nee das ich also momentan siehst du da keinen keinen Weg

Proband

Eher zweiteres.

Proband

So alles, was so in Richtung Output und Ergebnisse geht, würde ich niemals so stehen lassen, unabhängig davon, wie es entstanden ist.

Interviewer

OK, du also du selbst. Wenn es quasi noch n anderer Mensch gemacht hätte, hättest du gesagt, Hey nee, ich les mir das trotzdem noch mal durch, weil ich kann dem OKOK.

Proband

Ist natürlich immer auch die Frage, wie relevant der Task ist. Für die Saubermacherei bin ich vielleicht eher. Gewählt darüber hinwegzusehen. Das ist dann auch wieder so eine Delegationsfrage.

Interviewer

Genau. Ja, genau. Dann anders gefragt. Was für Aufgaben also wie banal? In Anführungszeichen müssten denn die Aufgaben sein, dass du, dass du dann sagst, ja okay, da kann man das machen und da würde ich das jetzt auch nicht mehr nachlesen und da wäre das auch ein Tool, wo ich sage, ja, ich bin eigentlich relativ das relativ zufrieden mit dem was da kommt. Wenn du aus den 100 Outputs 2 schlecht sind, dann juckt mich das nicht arg viel, aber die 98, die er mir Zeit gespart hat, da das, das war schon gut.

Proband

Das stimmt. Ich glaub in dem Kontext, da mit Saubermachen und sowas passt da schon, solange der mir jetzt nicht empfiehlt irgendwie Chlor zu benutzen um jetzt nen Spiegel zu reinigen der ne Beschichtung hat oder sowas.

Interviewer

Okay. Ich meine okay ja nee, aber. Wird auch effektiv. Nicht sichergestellt, dass er das nicht machen kann. Aber ja, OK. Aber das kam jetzt bei. Auch jetzt nicht. Genau. Ja, ich mein, das ist effektiv. Die Frage, die ich dir schon mal gestellt hab. Da hast du gesagt, bist du selber nicht ganz sicher, wie könnte man denn also dieses. Wie könnte man denn erhöhen, dass du mehr Vertrauen in hättest? Also jetzt wirklich explizit LLM in dieses System. Was könnte man denn machen? Also weil so wie ich das jetzt gehört hab das in das System selbst. Wie, du hast schon gesagt, du hast nicht wirklich viel vertrauendes System. Das ist jetzt aber auch unabhängig von der App. Meine App hat jetzt aber auch nichts daran geändert, wie deine Ansichtsweise bezüglich der ganzen Sache ist. Das hat sich das. Das genau das hat nichts davon. Verändert. Was würdest du aber sagen? Was müsste man denn tun oder könnte man überhaupt in deinen Augen irgendwas tun damit?

Proband

Meinst du jetzt allgemein oder speziell für diese App.

Interviewer

Beides, wird speziell für die App erstmal.

Proband

So speziell für die App. Das hat das Problem dadurch, dass man ja. Mit über Chat GPT arbeitet hat man ja keinen wirklichen Einfluss darauf, was da jetzt reingeht an Trainingsdatensätzen und was nicht. Das heißt, da kann man dann eh immer nur mit

Voraussetzungen rumfummeln, auf die man keinen großen Einfluss hat, dadurch, dass es ja auch nicht wirklich open Source ist, kannst du auch nicht ganz reinblicken was das vielleicht fehlen könnte, das heißt da hat man dann immer so im Hinterkopf noch Zweifel, ob da vielleicht Sachen jetzt nicht drin sind, die dann doch relevant werden und man hat ja nicht die Möglichkeit das herauszufinden. Mhm, das hat n grundsätzliches Problem mit dem ganzen Ding.

Interviewer

Okay Ja.

Proband

Gerade, wenn man dann auch sieht, dass die jetzt anfangen. Eben auch LLMS mit LLMS zu trainieren, nimmt die Konferenz, was das angeht, eher ab.

Sprecher

Okay.

Interviewer

OK und auch vor vor allem das das also die Confidence was da rauskommt und auch sagst du vom Vertrauen hast, dass das System dir. Immer gute Sachen Rausspuckt.

Proband

Dann vor allem, dass ich halt auch nicht reinblicken kann, was da eigentlich passiert hinter den Kulissen.

Interviewer

OK, das ist also, das ist auch, wird auch vermutlich auch für wird sich nicht ändern, solange du nicht weißt, was da tatsächlich unter der Haube ist, kannst du dem Ganzen auch nicht wirklich vertrauen.

Proband

Richtig gute LLMS müssten halt eigentlich auch Open Source sein. Also das.

Interviewer

Genau das dann. Dann komm ich auch gleich zum letzten Punkt, das heißt um deine User Acceptance zu erhöhen um von dem ganzen System, also wie würdest du allgemein beschreiben, wie hoch wär jetzt deine Acceptance bezüglich dem System, dass du sagst, das könnte man, dass man das in. In der realen Welt verwenden kann/ sollte.

Proband

Sie kann mir gut vorstellen, dass gerade für Putzen und so weiter das völlig in Ordnung ist. Für das, was es tut. Man könnte halt wahrscheinlich davon ausgehen, dass dann, je nachdem wo auf der Welt man guckt, dass es problematisch werden könnte, weil die Trainingsdatensätze jetzt natürlich eher aus westlicher Perspektive sind. Das heißt, wir benutzen, der will dann auch eben Reinigungsmittel und Methoden sagen, die man vielleicht jetzt. Keine Ahnung in Singapur nicht benutzt. Das könnte ich mir gut denken, dass das ein Problem sein könnte. Aber es ist halt auch wieder so ein total allgemein gesehen.

Interviewer

OK. Wo hättest du jetzt per se gesagt? Hast du dich am angenehmsten? Also woher warst du am sichersten, dass du sagst, da kommt was Gescheites raus? Welches Level hättest du gesagt, war da am besten klar beim ersten Confidence and Used ist einfach nur dein eigenes Geschriebenes, das können wir halt wieder mal wieder rauslassen. Aber wenn ihr jetzt Level 2, 3 und 4 anguckst.

Proband

Ich glaube, mein Favorit, wenn man jetzt zusammenzählt, wie angenehm was zu nutzen und wie sehr, denke ich, dass er was halbwegs vernünftiges rauskommt, wenn ich das, wo man dann ein Bild reingehauen hat und dann diese Tags angeklickt hat (Level 3).

Interviewer

OK, weil auch. Also jetzt von mir aus. Ich hatte auch letztens wollte ich das System vorstellen und dann ist mir tatsächlich auch sofort aufgefallen. Ich wollte n Beispiel machen und sagen, ja hier guck mal, das könnt man ja putzen und hab dann das System das zweite genommen wo du. Das schreiben musst und dann hab hab hab ich erstmal so n paar Sekunden drüber nachgedacht. OK warte mal was schreib ich da jetzt rein? Was macht denn da Sinn und es ist genau auch so n. So n Ding auf so ne Design Barrier über die. So n Paper gesprochen hat eben, dass man per se erstmal gar nicht weiß, was man überhaupt von dem System will. Man sagt da ja gut, jetzt ist Bild, das Bild ist relativ einfach, weil das ist da wo du davor stehst, das kriegst du hin, ne. Stehst du davor und denkst dir ja gut, was muss ich denn da überhaupt machen? Was macht n Sinn da zu machen?

Proband

Weil eigentlich soll das System einem das Denken abnehmen. Und jetzt muss man wirklich anfangen zu denken.

Interviewer

Genau jetzt musst du aber trotzdem bisschen denken, ja.

Proband

Weil bei dem anderen hat wird einem ja das Denken quasi abgenommen, weil dann sieht man irgendwelche Wörter und denkt sich ja, das könnte gut passen.

Interviewer

Genau. Und bezüglich dem Full Auto System, wie war das jetzt da? Hast du da eher gedacht, OK, ja, der macht da. Ich, ich fühle mich per se nicht so ganz, ganz zufrieden, weil ich mache ein Bild und Hey, da kommt jetzt vielleicht in dem Fall ein guter Text. Aber ich habe. Ja, aber warst du zufrieden damit, dass du quasi so wenig Einflussnahme hast in das ganze Geschehen? Oder hast du gesagt das da erzählt mir einfach n bisschen n bisschen was?

Proband

Ja genau. Also jetzt kann man wahrscheinlich so, wenn man es ganz überspitzt formuliert, als Kontrollverlust bezeichnen. Ja, seit deinem zweiten musst du zu viel Kontrolle machen, dann musst du denken, bei einer Aufgabe, wo du eigentlich nicht denken willst. Beim vierten musst du gar nicht mehr denken oder dann zweifelst du. OK, was habe ich jetzt gemacht, ist das überhaupt sinnvoll und dann muss man ja auch wieder denken, wenn man sich überlegt, ob das überhaupt so hinhauen könnte. Beim Dritten hat man dann die Illusion einer eigenen Entscheidung. Man macht was rein, man kann noch ein bisschen was anklicken und dann kann man sich denken, ja gut, ich hatte meinen Einfluss hier in dieser Pipeline, ich bin zufrieden damit.