

einmal herzlich alle willkommen alle die teilnehmen wollen hier zu unserem adercore Webinar zum Thema
ki-technologien datenschutzkonform ins Unternehmen integrieren mein Name ist Simon Kranik ich bin seit 5 Jahren data scientist hier bei der ADAC und möchte oder freue mich sehr in dieses Thema heute näher zu bringen
kurzer Überblick was haben wir heute vor wir starten mit einer kurzen Vorstellung der Aqua wer sind wir was machen wir und
gehen danach direkt über in die Welt der latein-technologie sein und dies heute hauptsächlich gehen soll bzw eigentlich ausschließlich gehen soll dafür wollen wir einmal die Frage beantworten was
sind denn eigentlich diese Darstellung Model welche Anwendungsmöglichkeiten und Herausforderungen sehen wir dabei und
wollen dann ganz ja ganz stark aber auf das Thema Software Service solutions und Open Source Modelle eingehen ja also das
sprechen wir dann vom Thema Chat versus Open Source Alternativen zu so etwas und das ist denke ich mal auch ein großer Punkt für viele die jetzt hier eingeschaltet haben das ganze wollen wir
Ihnen dann im Anschluss in der kurzen Demo zeigen das heißt wir haben so ein Open Source Modell selber gehostet wir
hoffen das ganze auf Asche und wir werden Ihnen zeigen dass es geht dass man das Ansprechen kann und dass man das
dann eben auch nutzen kann sobald man das ganze selber hauste da ist man eben plötzlich ein bisschen freier und kann so ein lang Model mit eigenen Daten verbinden und ja so eine eigene Knowledge Base quasi im Hintergrund anbinden auf die das Modell dann Zugriff
hat und ja im Anschluss daran wird dann noch Zeit sein für die ein oder andere Frage okay damit kurz zur Vorstellung der adercore wer sind wir um das Ganze kurz zu machen die ADAC ist ihr Begleiter auf
dem auf ihrer Cloud journey auf dem Weg zum digitalen Business ja das heißt wir entwickeln und Management
in der Cloud davon fast eine ganze Reihe von Lösungen wir bieten sowohl selbstgebaute clouds an auf die Sie Ihre ähm oder die ihre IT ähm Architektur stellt um ihre Lösungen ja zu hausten wir bieten aber auch Beratung im Bereich er glaub Migration zu den großen Diensten an wie z.B zu den großen dann wie Asche AWS etc
aufgrund unserer Historie aus dem Bereich des hausdings lassen Sie sie damit dann auch nicht alleine wir haben große großen Mitarbeiterstab der sich mit Cloud operations beschäftigt das heißt wir haben wenn Sie das wünschen bis zu 24 7 Monitoring und Management
ihrer Cloud Plattform das umschließt alles was dabei eben anfallen kann also instant change der kostet die Management
und ja ich möchte jetzt gar nicht auf alles eingehen aber ein Punkt noch zu heben und zwar das Thema Sicherheit und
Compliance wir haben eine sehr lange und stolze Historie zum Thema Sicherheit und Compliance und ja sind da eben sehr stark verwurzelt und lassen das eigentlich nie außer Acht und
damit was ist auch schon zu unserer ganz kurzen Vorstellungen wir starten direkt mit unserem Thema für heute nämlich den

Einführung in Large Language Models: Was sind LLMs und welche Open Source Alternativen gibt es zu ChatGPT?

large language models ich habe mir gedacht ich starte das ganze einfach mal mit so einer kurzen

auf dem radarfolie auf dem Radar soll bedeuten wenn man sich mit diesem Thema ein bisschen beschäftigt dann fliegen einem aktuell unglaublich viele Begriffe um die Ohren sei es sowas wie PPT oder

GP4 Dinge die man auf jeden Fall quasi mittlerweile wahrscheinlich gehört hat aber sobald man sich vielleicht mit tech

news beschäftigt kommen dann auch solche Begriffe wie Lama oder Alpaka oder wie Kuna

immer wieder unter aus dem Bereich der Open Source Modelle Begriffe wie hacking face oder lang chain das sind Technologien die sich im Bereich jetzt hier auf tun bzw Communitys die sich in den Bereich auf

tun und hervor tun als ganz zentrale Anlaufstellen und auch sowas vielleicht wie Einkaufen Vektor Datenbanken der

op-steck je nachdem welchen ähm auf welchem Level sie sich quasi mit diesen Dingen beschäftigen ganz gut sein dass

einige manche oder viele von diesen Begriffen Ihnen bereits um die Ohren geflogen sind und unser Ziel heute ist

es jetzt auch nicht wirklich jedes einzelne dieser jeden einzelnen dieser Begriffe irgendwie zu durchleuchten und

genau zu beschreiben was das denn ist sondern unser Ziel für heute ist es Ihnen einfach die Möglichkeit zu geben

dieses ganze Feld der large language Model und der Begrifflichkeiten die sich damit aktuell immer wieder auftun einmal

ein bisschen zugänglicher zu machen ja wir wollen dass Sie Möglichkeiten dass Sie die Möglichkeit haben Dinge wie gut

open GPT etc so ein bisschen in Kontext zu setzen aber auch zu verstehen okay Begriffe wie Alpaka Lama wie Kuna wenn

die fallen dann sind wir im Bereich der Open Source Modelle ja wenn ich sowas höre wie äh Transformer hugging face and

weddings das sind zwar alles Begriffe die ich so im Detail vielleicht gar nicht kennen muss aber ich weiß sie kommen auf weil wir im Bereich bewegen

und Begriffe wie vektordatenbank Samantha Versuche und so weiter das sind Technologien die quasi auf der Welle jetzt mit Schwimmen die eine sehr gute

Synergie bieten zu dem was da language

liefern also alles Dinge die man jetzt nicht im Detail beleuchten muss aber wie

gesagt wir hoffen dass Sie am Ende von diesem Webinar ein ganz kleines bisschen besser gewappnet sind diese Dinge in

freier Wildbahn einschätzen zu können um das tun zu können finde ich ist es aber ganz ganz wichtig erstmal zu

Was sind LLMs überhaupt?

verstehen was large language Model denn eigentlich sind was tun die denn eigentlich und um das zu beantworten

gehen wir ein kleines in kleinen Schritt zurück zu statistischen sprachmodellen und das ganze tun wir weil ich der

festen Überzeugung bin dass jeder von ihnen so eins schon mal benutzt hat und ich spreche jetzt nicht von Chat GPT

sondern von etwas viel kleineren der Autovervollständigung auf ihrem Handy ja wenn sie auf ihrem Handy einen Satz eintippen wie z.B heute ist ein wunderbarer Tag dann wird ihr Handy einen nächsten ein nächstes Wort vorschlagen das ist seiner Meinung nach oder das auf Basis von einem statistischen Sprachmodell das im Hintergrund läuft das wahrscheinlichste ist es für den Auswahl von Worten von Wörtern geben in der Mitte ist quasi immer das wahrscheinlichste Wort und links und rechts davon ist es zumindest am iPhone sind dann andere Wörter die durchaus auch vorkommen können das heißt dass der das Ziel eines statistischen language Models ist es das nächste Wort in einem oder zu einem bestehenden Text auszugeben jetzt kann man das Ganze aber iterativ durchführen und so einen ganzen Text erzeugen lassen und das machen wir jetzt einfach mal bzw das habe ich im Vorfeld gemacht und die Ergebnisse hier einfach mal notiert und wir schauen mal was dabei rauskommt ja heute ist ein wunderbarer Tag war der Eingangstext und jetzt geht's weiter zum Spazieren gehen benötigt eine gesunde Ernährung ist sehr wichtig was denn da jetzt passiert irgendwie hat das Modell so ein bisschen den Faden verloren und das liegt daran dass statistische sprachmodelle nicht für so einen langen Kontext gedacht sind ja ich habe hier so ein bisschen versucht quasi zu markieren was das Modell zu jedem Zeitpunkt gesehen hat in Form von diesem kleinen blauen Text den grauen Text den hat das Modell schon komplett vergessen es sieht nur den blauen Text und ja wenn man sich den blauen Text anguckt dann war Ernährung ist sehr wichtig hier ein gute Fortsetzung von dem von dem Wort von dem von dem Satz aber wie gesagt der Kontext den komplett verloren und da in diese Lücke oder in dieses Problem stoßen jetzt die Länge ist das heißt erstmal genau dasselbe rein von der Idee her wie diese statistischen sie führen einen Satz weiter der eingegeben wurde probieren wir das ganze mal mit einem Satz erstelle eine kurze Beschreibung von large language ich habe es extra weit weggelassen und das ganze so ein bisschen zu Prime und gucken was passiert wenn ich das durch einen Model schicke gut das nächste Wort ist Models soweit so gut aber es macht weiter erläutere wie sie sich auf künstliche Intelligenz und maschinelles Lernen auswirken okay soweit schon mal sehr viel besserer Text ja der Text ist schön kohärent macht Sinn ist eine schöne Fortsetzung von dem was ich hier eingegeben habe und ich habe das ganze jetzt mal markiert dieser blaue Bereich der wurde nie kleiner ja large language sind eben in der Lage ein sehr großen Kontext zu verarbeiten vielleicht ist ihnen aber aufgefallen aufgefallen dass ich hier links ja und ich habe sie nicht vollständig formuliert aber ich habe eine Frage eigentlich versucht zu stellen ich habe versucht das Modell zu fragen ob es mir eine kurze Beschreibung geben kann was es getan hat ist aber einfach nur meine

Frage weiter vorzuführen und das ist ja etwas dass sie wahrscheinlich wenn sie Betriebe benutzen so nicht in dem sie so nicht begegnet sind ja wenn Sie eine Frage stellen dann beantwortet das diese Frage ja es reagiert als Gesprächspartner auf das was Sie gesagt haben um dahin zu kommen einmal ganz kurz eine Erklärung was ein Darm im Hintergrund getan hat bzw wurde auf einer gigantischen Menge an Text trainiert ja das heißt wir haben ganz ganz viele Trainingsdaten in diesem Modell mitgegeben und das hat gelernt welche Wörter typischerweise mit anderen welchen mit anderen Wörtern hinsetzen auftauchen und hat daraus quasi so ein Bild von menschlicher Sprache im Inneren sich geschaffen und ja seine Aufgabe den nächsten das nächste Wort zu predigen ist dabei aber nie vergessen und genauso ist vergeht auch der macht auch oder dieses Modell macht auch nichts anderes warum wirkt es jetzt also wie ein Gesprächspartner man nennt das ganze jetzt ein konversationelles Lautstärke und der Trick liegt nicht in einer Veränderung des Modells sondern einer Veränderung oder einer Augmentierung der Trainingsdaten ich weiß nicht wie viele von ihnen sich an den Schulunterricht an die deutsche Unterricht oder Englischunterricht zurück erinnern können aber da hat man das ein oder andere Mal diese kleinen gelben Heftchen gehabt in denen ein konversationsstil zum Beispiel Shakespeare oder sowas ein konversationsstil vorgegeben wurde wir hatten immer Sprecher eins Sprecher zwei Sprecher 1 Sprecher zwei und die unterschiedlichen Sprecher haben eben in dem Text etwas gesagt und der nächste Sprecher bezog sich dann auf den vorher gegangenen Text und ja genau diesen Stil diesen Textil hat man sich jetzt zu Nutze gemacht und hat ganz ganz ganz ganz viele Beispiele erzeugt die diesen Textil folgen und man hat dafür solche sogenannten mitgegeben wenn ich hier ein prompter und ein Assist t Token und wenn ich diese kleinen Tokens jetzt mitgeben nachdem ein Modell wie Jetski auf diesen Token trainiert wurde dann passiert folgendes ich gehe dann gleich drauf ein warum das passiert erstelle eine kurze Beschreibung Models ist das denn das Language Models sind fortschrittliche Computerprogramme die Menschen ähnliche Texte erstellen können ja was passiert hier ganz einfach das Modell für weiterhin einfach Text fort mehr macht es nicht aber es hat gelernt dass Texte einer bestimmten Form folgen und zwar dieser ein prompter fragte was eines System antwortet ein prompter fragt noch etwas eines System antwortet wieder das heißt das Language hat einfach gelernt diese Form einzuhalten und jetzt haben wir ein Programm darum geschrieben mit wir meine ich jetzt an der Stelle open bzw jeder der so ein Chatbot jetzt baut wir haben ein Programm darum beschrieben bei dem wir ganz klar sagen okay jeder Input des Users wird in diese Tokens verpackt und sobald das Sprachmodell von sich aus auf die Idee kommt dieses prompter Token auszugeben dann ist es fertig dann wird das Modell unterbrochen und seine

Antwort ist quasi abgeschlossen und damit haben wir plötzlich dieses dieses Gefühl von einem Gesprächspartner nur auf Basis von Trainingsdaten hier aus diesen Modellen herausgekitzelt und um einmal kurz zu beweisen in Anführungszeichen dass das tatsächlich auch die genauso funktioniert habe die einfach mal gefragt ob er mir nicht eine Liste von zehn solchen Tokens geben kann die für den Wechsel von dem Gesprächspartner Gesprächspartner typischerweise in Modellen benutzt werden ja hat gesagt alles klar weiß gerne hier sind zehn Beispiele für solche Tokens die solche Sätze oder die den Gesprächswechsel markieren erstens end of text fertig das ist hier passiert hat das nächste Tokenity ausgeben wollte war das Token dass er selber benutzt um das Ende seines Satzes zu markieren und deswegen hat es einen Satz einfach beendet das Modell hat quasi seinen eigenen Ausschalter betätigt und hat damit hier den Text frühzeitig unterbrochen und aus der Liste von zehn Beispielen wurde einfach eine Liste von einem Beispiel und das nächste Beispiel das haben wir schon gar nicht mehr zu Gesicht bekommen also das so ein kleiner Blick hinter den Vorhang hinter die Kulissen was passiert denn da eigentlich ich hoffe damit sind sie auf jeden Fall schon mal ein kleines bisschen ja ein kleines bisschen schlauer dazu was denn eigentlich tun

Anwendungs-möglichkeiten eines LLMs so weit so schön aber was bedeutet das denn jetzt für uns welche Anwendungsmöglichkeiten bietet mir ein solches konversationelles Modell ich denke viele von ihnen haben sehr viele Ideen was man mit Modellen machen kann nichtsdestotrotz habe ich einfach mal ein paar zusammengefasst auf der linken Seite fangen wir an mit Text Manipulation wir können diesen Modell sagen erstelle eine Zusammenfassung dieses Dokuments und dann kommt ein Text den wir mitgeben dabei noch mal als Erinnerung hätten wir kein konversationelles Training durchgeführt auf dem Modell also hätten wir ihm nicht beigebracht diesen Gesprächspartnerwechsel durchzuführen dann würde er welches Dokument wie auch immer mitgeben würde einfach Fortsetzen der würde gar nicht verstehen dass er das seine Aufgabe ist tatsächlich das zu tun was wir Ihnen hier fragen und da kommt die ganze Magie eben ins Spiel okay also nächstes Beispiel übersetze diese E-Mail von Deutsch nach Englisch da ist jetzt natürlich eine Bedingung dran geknüpft nämlich dass das Modell sowohl Texte von auf Deutsch als auf Texte auf Englisch in sein Trainingsdaten drin hatte und am besten auch Übersetzung von dem einem zu dem anderen ja also das ist immer ganz wichtig zu berücksichtigen die Trainingsdaten sind grundlegend für die Fähigkeiten des Modells verantwortlich okay geführte Generierung ist es Beispiel oder beziehungsweise nächste Kategorie etwas das bestimmt viele von ihnen schon mal durch oder schon mal ausprobiert haben Verfasser eine E-Mail an mein Team mit einem Workshop Reminder ja das heißt wir sagen dem Modell das ist ein textgenerieren soll und packen

diesen Text aber so ein bisschen in unsere eigenen Regeln und sagen ihm okay ich möchte dass der Text ungefähr so oder diesen diesen Inhalt enthält machen mal spannender wird es dann zum Beispiel bei Code Generierung hier unten erstelle SQL

Code um die fünf Produkte mit der größten gewinnmarsche in 2022 zu ermitteln ja das ist für viele auch so

ein Punkt gewesen bei JET GPT bei dem man gemerkt hat oh das Modell ist ja oder diese diese dieses Tool ist ja das

doch tatsächlich viel nützlicher sich ursprünglich gedacht habe hierbei muss man natürlich sagen wenn ich so eine Frage stelle dann muss

ich den Modell irgendwie noch ein bisschen Kontext darüber mitgeben wie sieht meine Datenbank aus welche Spalten

habe ich welche Tabellen habe ich und ich musste Modell am besten auch irgendwie erklären und irgendwie sagen

können was diese Spalten bedeuten bzw sie müssen mindestens mal sprechende Namen haben also wenn ich so ein rein technische Datenbank habe in der alles

nur mit IDs bezeichnet ist dann hat das Modell hier natürlich auch Schwierigkeiten außer ich erkläre alles

ganz ganz intensiv und dann bin ich meistens auch schneller das ganze einfach selber umzusetzen oder jemanden

zu fragen der selber ist Geld schreiben kann aber nur als Idee sowas hier ist durchaus möglich Dritter Bereich dritte Sparte question

and ring ganz kurz das wird der Bereich sein den wir uns später auch näher noch anschauen werden also bleiben Sie dran

werden sie dann auch noch live einmal in Aktion sehen aber was kann das jetzt zum Beispiel machen sagen wir mal wir

stellen die Modell die Frage was wurde im Customer Generation Workshop besprochen ja

jetzt müssen wir den Modell natürlich irgendwie noch den Inhalt des Customer Generation Workshops mitgeben so ein bisschen wie hier links bei erstelle

eine Zusammenfassung dieses Dokuments aber wenn wir das tun dann wird das Modell verstehen das ist die sagen wir

zum Beispiel das Protokoll oder die Präsentation aus diesem Workshop sieht und wird verstehen welche Teile davon

meine Frage beantworten wird das Ganze zusammenfassen und die Frage eben versuchen bestmöglich zu beantworten andere Frage die man stellen könnte ist

wer kennt sich im Unternehmen mit vektoratenbanken aus ab einer bestimmten Unternehmensgröße ist das bestimmt etwas dass man so oder so ähnlich gerne mal

einfach eine Maschine fragen würde und die Idee dahinter ist wir geben dem Modell einfach alle möglichen Meeting

Protokolle mit ja und dann gibt's vielleicht irgendwann mal ein Meeting in dem das Wort Weg dort hatten Bank gefallen ist oder in dem über Vektor

Datenbanken gesprochen wurde inklusive einer Teilnehmerliste und das Modell versucht jetzt nach Möglichkeit das zu

kombinieren und uns den Auto zu geben okay hier gab es ein Meeting die Leute waren beteiligt fragt doch mal die die könnten sich damit auskennen

dass der Wunsch okay schöne Anwendungsmöglichkeiten kommen wir zu den Herausforderungen

Herausforderungen von LLMs: Am Ball bleiben, Datensicherheit, Integration, Modellqualität, Fallstricke im Lizenz-Dschungel, Entwicklungskosten

Nummer eins und für mich aktuell ganz ganz spannend am Ball bleiben der Zug der Technologien und der Entwicklung von large language Models sowohl im Bereich der Politik um die großen Modelle Jacky verboten in Italien dsgvo concerns in Richtung open da bewegt sich ganz ganz viel da tut sich ganz ganz viel und es ist bei einer extrem schwierig allein in dem Bereich irgendwie am Wald zu bleiben was denn eigentlich aktuell ja was was da los ist was die Modelle in Zukunft können dürfen was vielleicht in Zukunft da passieren wird allein das ist schon schwierig zum anderen ist aber auch der Open sourcebereich unglaublich unglaublich schnell ja also hier kommt stündlich kommen neue Modelle auf den Markt täglich kommen gute Modelle auf den Markt und wöchentlich kommen bahnbrechende Modelle auf dem Markt die eben im Open Source Bereich hier wie gesagt Wellen schlagen ohne Ende wenn man natürlich sich in diesem ganzen Kontext nicht bewegt dann ist das ein riesiger Berg an Modellen bei denen ist leider nicht vielleicht irgendwie einfach nur nach most recent zu filtern sondern man muss natürlich ein bisschen in der Community drin bleiben und verstehen welche Modelle werden denn aktuell wofür angewandt okay also am Ball bleiben Herausforderung Nummer eins Herausforderung Nummer 2 Datensicherheit ja wir hatten jetzt eben dieses Thema von meetingprotokolle hochladen aber möchte ich das wirklich mit Open verbinden also möchte ich wirklich meine Meeting Protokolle eine fremde Firma die was auch immer damit macht allein schon neue Modelle damit trainiert ist vielleicht schon was was ich gar nicht möchte also meine Daten sollen nach Möglichkeit sicher bleiben und das können sie irgendwie nicht wenn ich solche Software Service Lösung verwende das gleiche fährt von hinten einmal aufgezeichnet Integration schön und gut dann lasse ich eben das mit den privaten Daten aber dann fehlt irgendwie die Integration zu seinem eigenen Business es fehlt irgendwie dass das Modell mehr machen kann als einfach nur eine Frage zu beantworten wie wer war der erste Präsident der Vereinigten Staaten ja ich möchte ja spezifische Fragen zu meinem Business hier stellen können oder zu meiner Firmenpolitik oder zu Verträgen mit Kunden oder was auch immer und ich will ohne bedenken dass das Modell integriert ist und diese Fragen beantworten kann okay das klingt ist alles erstmal sehr magisch aber Thema modellqualität ist die nächste Herausforderung gerade im Bereich Open Source sind wir oder die Open Source Community ist noch nicht an dem Punkt dass man jetzt einfach sagen kann das ist eins zu eins Ersatz für jetpity Jackie BT die großen Modelle die sind immer noch führend was die modellqualität angeht und ich denke viel von ihnen wird auch bekannt sein dass GPT auch alles andere als fehlerfrei ist ja also da gibt's auch schon viele Punkte wo man sagt da halluziniert das Modell gerne mal das spuckt das fälschlicherweise oder falsche Informationen aus im Bereich von Bing Chat zum Beispiel von Microsoft wird das Modell auch gerne mal beleidigend obwohl das gar nicht tun soll also die sind alle auch noch nicht perfekt und die Open Sour Community die hinten noch ein

kleines bisschen hinterher da kommen wir aber gleich noch drauf oder gehen noch gleich ein bisschen mehr drauf ein
beobachtet werden und muss eben auch gut eingeschätzt werden noch zwei Punkte der
nächste lizenzdschungel ja gerade für eine sourcebereich wichtig ein Open Source Modell
das hat zum einen die Lizenz desjenigen der es trainiert hat also derjenige der den Programmcode geschrieben hat aber
auch die Lizenz der Datensätze die benutzt wurden um das Modell zu trainieren ist relevant ja und das sind
gerne mal ist gerne mal nicht nur ein Datensatz sondern viele auf einmal und jeder einzelne Datensatz bringt seine
eigene Lizenz mit oder zumindest sollte er das wenn das nicht tut ist es generell schon mal eine rote rote Flagge
aber jetzt kannst durchaus passieren dass quasi einen ein Modell zwar Open Source und mit Apache zwei Leistungs
markiert ist so dass man es kommerziell nutzen darf aber auf Daten trainiert wurde die diese Lizenz eben nicht erfüllen und damit darf man auch das
Modell nicht füllen man darf den Quellcode zwei kommerziell nutzen aber das Modell selber nicht kommerziell einsetzen also das ist ein bisschen
Dschungel der nicht so einfach zu navigieren ist und da muss man eben auch ein bisschen wie das Thema Ball bleiben
guten Überblick bekommen darüber westeten Datensätze sind aktuell beliebt welche Lizenzen haben die welche Modelle wurden auf was trainiert etc letztes
Beispiel denn oder letzte Herausforderung Entwicklungskosten muss ich mir jetzt Gigabyte Weise Textdaten irgendwie aus
dem Rippen schneiden muss die vorverarbeiten in der konversationellen Form muss dann eine gpu-maschine mit irgendwie
4 8 wie vielen GPUs auch immer dahin stellen die dann erstmal für eine Woche das ganze Ding trainieren darf oder wie
funktioniert das also sind die Entwicklungskosten wirklich so eine große Hürde das Wirken oder nicht da
kann ich schon mal beruhigen nein sind sie nicht ja man muss trotzdem eine GPU Maschine sich hinstellen aber bei den Modellen die aktuell am Markt sind
Modelle da reicht eine kleine die ist nicht viel viel teurer als eine produktiv Maschine die eben production
sogar Cloud auf CPU basiert übernimmt da gehen wir später auch noch gerne bei den Fragen dann näher drauf ein okay das soll es einmal mit dem ganzen
Herausforderungen gewesen sein gibt
einige aber dafür sind wir ja da Thema Software ist es Service versus Open Source ich habe hier eine kleine
SaaS und Open Source: Welche Modelle können ChatGPT ersetzen?
Grafik mitgebracht die wurde von den Erstellern von vicuna einem Open Source Modell erzeugt und das ganze ist ein
Benchmark von verschiedenen Modellen gegen Chat GPT deswegen ist hier mit 100% markiert das soll nicht bedeuten
dass die Spitze der veranstange ist gut für diesen Test jetzt schon aber die 100% können hier durchaus gesprengt
werden aber erstmal ganz kurz was worum geht's jetzt hier also bitte rechts ein Modell das bei diesem Test eben mit 100%

abgeschnitten hat weil es der Maßstab war das große open-ei Flaggschiff aktuell immer noch mit 170 Milliarden Parametern Modellparameter und technisch zu werden das sind die ganzen vielen kleinen Stellschrauben die das Modell hat um bei dem Training zu lernen was es eben lernen muss also je mehr Parameter desto zumindest in der Theorie desto besser kann das Modell Dinge lernen muss nicht zwangsläufig die Daten sind eben mindestens genauso wichtig aber dass die Idee so 170 Milliarden Parameter bei Tee daneben Google bald 540 Milliarden Parameter also noch mal um einen Faktor von 3 bis 4 mehr und erreicht leider schon gar nicht mehr die den Benchmark von Jacky oder den Menschen mag werden das heißt hier sehen wir schon Parameter sind nicht alles und ähm ja wenn wir links daneben gucken und wir schauen uns mal die Namen der Modelle an Lama 13b und 513b ähm das heißt 13 Milliarden also es ist ein Markierer dafür wie viele Parameter hier benutzt wurden 13 Milliarden gegenüber 540 Milliarden und 170 Milliarden bei den beiden rechten Modellen also viel viel weniger weniger als 10% im Vergleich schon zu Chat geschweige denn Wahrheit und trotzdem sind diese Modelle in der Lage zumindest wie Kuna hier an der Stelle eine sehr sehr ordentliche Performance zumindest laut diesen Benchmark zu leben das ist aber noch viel spannender macht ist der Zeitlauf oder der zeitliche Ablauf in dem das ganze hier passiert ist Lama wurde Ende Februar von Meta ehemals Facebook auf dem Markt gebracht keine zwei Wochen später hat die Universität von Stanford das Alpaka Modell veröffentlicht und hier steht jetzt eine Woche eigentlich sind es eher ein bis zwei Wochen eher zwei Wochen etwa zwei Wochen später wurde dann auf Alpaka basierend dass wir Tuner Modell auf dem Markt gebracht und wir sind plötzlich bei einem Benchmark Wert der für die allermeisten Fälle ausreicht um tatsächlich einfach zu ersetzen viele viele Dinge kannst noch nicht so gut wie TT aber das soll jetzt gar nicht Thema sein entscheidend ist hier das ganze ist innerhalb von einem Monat passiert also Lama wurde veröffentlicht kurz danach kam Alpaka kurz danach kam wie cooler und es ist eben durch die Decke gegangen das ganze war Anfang April seitdem hat sich schon wieder so viel getan das kann ich in der Grafik mit drin ist das ist unglaublich also der Markt der Open Source Modelle oder die Idee Community-Bereich bewegt sich wahnsinnig schnell aber das hat natürlich nicht nur wir erkannt und auch nicht nur die Open Sour Community das ganze hat auch Google erkannt das von ein paar Tagen gab es eingelegt das Internet Dokument von Google indem sie ganz klar sagen wir haben im aei-bereich kein alleinstellungsmerk mehr mehr wir kriegen es nicht hin hier am Ball zu bleiben we have no Mode das Open also der Meinung von Google nach geht das gleiche eben auch für sowas Gemeinschaft aber da haben wir das

haben wir eben gerade gesehen
das also eine Aussage von Google zu diesem Trend und schließen tun sie das ganze
mit wie ein notpositioned to win
this armsrace and noizer is open ai worauf möchten wir damit hinaus auch wenn die
Open Source Modelle
aktuell vielleicht noch nicht den Benchmark knacken bleiben Sie am Ball gucken Sie
sich diesen Bereich gerne mit uns zusammen an das tut sich was und das
tut sich was sehr sehr interessant ist ja damit habe ich jetzt lange um den
heißen Brei geredet von wegen da tut sich ganz viel kommen wir zu Demo ja weil das
ist ja dann der Teil der das
auch ein bisschen zeigt die wird relativ kurz sein aber sie zeigt ebenso ein bisschen
in einem ganz ganz einfachen
Demo eines Knowledge Retrieval und Answering System
Konzept was denn möglich wäre wir wollen Ihnen ein knowledge Retriever question
Anzeigen System zeigen basierend auf
eigenen Dokumenten so wie es zu einer Magen gekündigt habe was bedeutet
das ein kurzer Überblick denn der User
stellt eine Frage diese Frage soll ein LLM beantworten ja bevor das LLM das tut
geht diese Frage allerdings an ein sogenanntes Knowle System dieses System
ist einfach ein Sammelbegriff und beinhaltet Dokumente und für den Datenschutz
ganz ganz wichtig unserer
Meinung nach eine Access Control ist auf diesen Dokumenten das haben wir hier
einfach mal mit eingebaut und ganz klar zu machen schon von Sekunde 0 schon
beim
Aufbauen der der noch Space ist es wichtig darauf zu achten diese Dokumente eben
nur Usern zugänglich zu machen
die die Dokumente auch lesen dürfen und eben dann down the line auch dem LLM
nur
das zur Verfügung zu stellen was der User der das Ende befragt sehen dürfte ja
also das ne einmal hier das neue System Dokumente mit einer Control ist und darauf
gehe ich jetzt gar nicht näher
ein wie das technisch funktioniert sonst würden wir den Namen ein bisschen
sprechen eine semantische Suchfunktion auf diesen Dokumenten
okay also die Frage geht rein in das nordische viel System da sind die Informationen
drin aber eben ganz ganz
viele Informationen gepaart mit einer semantischen Suche führt das zu einer
Sammlung von relevanten Dokumenten bzw
relevante Dokument Ausschnitte und diese beiden Dinge werden jetzt an das LLM
geschickt und zwar einfach in Form von der User Curry also das heißt die Frage vom
von dem Menschen wurde wird einfach
ergänzt das können sie also bei JET GPT in der Oberfläche auch machen indem Sie
einfach Ihre Frage stellen und dann Kontext mitgeben mehr passiert hier
nicht also ich möchte einfach noch mal erwähnen dass keine Magie kein externer
Dienst oder sonst was im Spiel sondern wir sagen einfach nur das ist die Frage
des Nutzers hier ist erweiterter Inhalt bitte beantwortet die Frage ja und dieser
erweiterte Inhalt wo der herkommt
dass es eben das spannende nämlich aus unserem eigenen Dokument verstanden
gut okay damit soll das jetzt die Frage
beantworten soweit so gut was machen wir oder was nutzen wir als Technologie
dafür bevor

wir wechseln einmal ganz schnell das wir haben ein Asia Lake House aufgebaut basierend auf data bricks das bedeutet
Welche Technologien werden hierfür eingesetzt?
wir haben ein Data Lake auf den wir Dokumente einfach drauf werfen können dann haben wir über data bricks diese
Dokumente in die indiziert und ein bisschen vorverarbeitet und eben verwaltet und vor allem über data bricks
auch die die Zugriffsrechte organisiert ja also nächstes Control ist angelegt alles jetzt gleich wenn sie sehen in dem
ganz kleinen Rahmen so dass man das super nachvollziehen kann aber das ist gerade die grundtechnologie gehostet auf Asche und als Sprachmodell
also das Model verwenden wir Dolly witzigerweise und tatsächlich überhaupt nicht notwendigerweise eine Paarung
stolly auch von Data Brix das sind Open Source Modell das databaks auf dem Markt gebracht hat und deswegen interessant weil der brexit sich hier mit kurdos nachdem auf die Fahne geschrieben hat dass ein wirklich vollständig kommerziell nutzbares Open Source Modell auf dem Markt zu bringen das sind mittlerweile nicht mehr das einzige als
Dolly aber auf den Markt kam das war Mitte April war es eigentlich das einzige kommerziellen nutzbare open
source-modell mit diesen fähigkeitsstand das also einmal erwähnt und wir benutzen das Dolly 3 Milliarden Parameter Modell es gibt auch genau wie bei wikuna mit 13 Milliarden Parametern wir benutzen 3
Milliarden Parameter einfach weil das ganze noch mal günstiger ist im Hosting noch mal billiger zu betreiben weniger
weniger Masse an an Zahlen die durchgerechnet werden müssen das heißt es ist ein bisschen schneller und unsere
Meinung nach von der Qualität her trotzdem schon ziemlich beeindruckend aber auch natürlich hier nicht das Ende der veranstange alles klar damit
wechseln wir einmal in unsere data Brix Instanz einmal erstmal auf die Startseite ja so
dass hier ist dann das was sie sehen würden wenn sie sich bei datapest dann einloggen wie gesagt das heißt Sie
können das ganze abriegeln und in ihre eigenen Cloud Umgebung eben alles organisieren und
ja mit datablicks können wir eben jetzt wie gesagt unser dataway Haus basierend auf einem data Lake aufbauen das haben
wir gemacht jetzt wechseln wir einmal ganz kurz in einen sogenannten Notebook das ist einfach eine data science typischerweise pfeifengesteuerte
Umgebung in der wir jetzt mit unseren Daten mit unserem aber auch generell mit heißen eben interagieren können und
Dinge tun können und ja wir wollen ihn einmal ganz kurz vorstellen dass wir hier im Hintergrund Daten haben das ist
jetzt nicht der spannende Teil deswegen frage ich das Modell einmal ab das muss ganz kurz erstmal warm laufen welche Dokumente bin ich denn aktuell
laut to see also welche werden werden wir denn zur Verfügung gestellt und das sehen wir einmal den hier sind zwei
paper einmal der gpt4 technische Report dann ein Paper Sparks of artificial general intelligence und extra den
Dokumente excellink befragen das Modell gleich mal was extra ist kurz vorab ist

eine weiteres Geschäftsfeld der dacore ähm dass sich sehr stark für sehr stark aber nicht ausschließlich im Bereich von E-Mobilität platziert da kommen wir später dazu das heißt ich habe jetzt durch drei Dokumente irgendwie hinterlegt auf die Zugriff habe wie funktioniert das Ganze ganz ganz simpel wir haben Dokumente gut ich kriege jetzt nur die drei auch ausgegeben auf die Zugriff habe aber auf der anderen Seite haben wir eine Access Controller ist ich bin der witzigerweise benannten Gruppe Lorenz Reader das ist aus Teststrecken so genannt das ist immer unsere Testgruppe und ja wir haben hier drei IDs das sind dieser Dokumente die hier oben sehen und eben nicht Rechte mehr ist das nicht ja und durch data Brix und durch die Verwaltung oder Governance Möglichkeiten von X habe ich jetzt eben die Möglichkeit zu sagen diese Liste soll quasi die Sperre zu den Dokumenten für jeden einzelnen User sein das heißt bei jeder Abfrage wird geprüft welche ob der User in einer Gruppe ist die dieses Dokument sehen darf so Dokument sage ich jetzt die ganze Zeit was heißt das denn ja wir wechseln erstmal ganz kurz um zu zeigen was das hier für Dokumente sind auf diese beiden paper ich gehe mal ganz kurz hier hin zum Spice of artificial Generation paper warum haben wir das genutzt ganz einfach es ist ein Paper mit 150 Seiten Text sehr sehr viel Text hier wurde GPT 4 also das neue das kommende oder das neue Flaggschiff von Open ai wurde hier auf die Probe gestellt unglaublich viel Inhalt unglaublich viel Text sehr spannend aber eben sehr umfangreich also hier ist eine semantische Suche oder generelle suche hier drauf durchaus von Vorteil und genauso sieht es mit dem technischen Report aus 100 Seiten technische Details über GPT 4 also das neue Sprachmodell wie gesagt von oben ja also einfach interessant um diese semantische Suche einmal vorzuführen okay jetzt kann ich keine Sprachmodell auch nicht Chat gpt und oder sonstigen großen sprachmodellen so ein riesiges Dokument einfach mitgeben ohne dass es einknickt deswegen macht man noch folgendes man verwandelt jedes Dokument in eine ganze Reihe von sogenannten Chunks das ist noch der letzte Schritt der datenvorverarbeitung das heißt ich habe hier ein Dokument oder einen Ausschnitt aus dem Dokument typifiziert technische Report das ist Ausschnitt Nummer 50 dann 51 52 also wir teilen das Dokument einfach noch mal auf in gut verarbeitbare Stücke die eben trotzdem noch genug Inhalt ja aufweisen können und das diese Abschnitte diese Blöcke das ist das was das Modell letzten Endes sich dann ziehen kann und zum Beantworten seiner Frage oder der Frage des Nutzers nutzen kann okay ganz ganz kurzer Kommentar zum Model Hosting wir haben das Ganze zwar noch verpackt in Service und so weiter der auf einer VM läuft aber ganz ganz prinzipiell wenn Sie das ganze irgendwie bei sich mit einem ganz kleinen Modell vielleicht sogar ohne GPU einmal testen wollen dann brauchen Sie nicht viel mehr als das hier wie gesagt um das Ganze

stabil laufen zu lassen brauchen sie schon noch ein bisschen mehr aber die Grundidee ist wirklich das hier das Modell wird gehostet auf einer face das ist eine Webseite die ich vorhin schon mal erwähnt habe eine Community da haben wir hier das datale Story v23 billion

Modell das von dem wir eben gesprochen haben das machen und man lernt das runter indem man einfach sagt okay der was der tropenizer ist da gehen wir jetzt gar nicht drauf ein das brauchen sie aber auf jeden Fall auch ähm und das Modell wird eben mit einer Funktion geladen und dann wird das Modell runtergeladen ja das heißt obwohl das fucking face für Sie gehostet wird das Modell ist das nur ein Repository und wenn Sie das Modell dann tatsächlich laufen lassen läuft es auf der VM auf der Sie diesen Code hier ausführen und wie gesagt wir haben das ganze jetzt noch mal verpackt in einen Endpunkt so dass das Ganze auf einer gpuvm im Hintergrund auf läuft

aber ja genug der Worte kommen wir doch einfach jetzt mal zu unserem Beispiel was dann passiert wenn wir das mit der jetzt tatsächlich etwas fragen ich starte mal mit der Suche deaktiviert deshalb haben wir als kleinen Fleck mitgegeben nur einmal zu zeigen dass sowas möglich ist ja wir fragen jetzt

Drei Beispiele einer Anwendung für die semantische Suche im Scope trainierter Daten das Modell zum Beispiel was first president of the US ja also das was wir vorhin schon mal durchgesprochen hatten so dieses nicht besonders hilfreiche aber so als Benchmark kleines Beispiel ich habe hier diese Frage gestellt wir haben den Output ein bisschen formatiert und gesehen die Antwort war schon Washington so weit so gut und dafür gibt es kein einziges Referenzdokument liegt ganz einfach daran dass ich die Suche deaktiviert habe okay dann lassen wir die Suche mal aus und Fragen noch etwas

und zwar jetzt etwas das uns tatsächlich intern interessiert und ist exceling wie gesagt die Suche ist aus ja das heißt es wird jetzt Unfug bei rauskommen aber den möchten wir Ihnen auch einmal nicht vorenthalten die Ausgabe ist extra Link ist eine wireless technology company base in mountain you California so das mag jetzt stimmen oder nicht ist aber definitiv für das was ich wissen wollte nicht korrekt ja weil ich weiß wir haben internen Dokumente die beschreiben was extra Link in unserem Kontext ist also aktiviere jetzt ne das ist einfach eine Funktion die wir geschrieben haben sie können das ganze natürlich auch standardmäßig aktiv lassen aber wenn ich die Suche jetzt aktiviere und ich stelle die gleiche Frage noch mal ist Entschuldigung

ja dann retrieve der einmal kurz Dokumente deswegen dauert ein kleines bisschen länger und wir sind diese Antwort extra Link ist eine Low Code Integration Plattform from aderco Hosting aus business is the create API's integrierte data in Clouds applications er konnte unsere Frage super beantworten und er gibt uns sogar die Referenz des Dokumentes mit dir benutzt hat um diese Frage zu beantworten ja also das kommt aus den extra Link Box Nummer 4 und hier ist der Ausschnitt in der benutzt hatte zu beantworten der Frage also wir sehen hier klar logiteation Plattform hier unten steht dann dass das Ganze von ADAC

Hausding ist und hier steht dieser Bereich über das eigene APIs erstellen und so weiter super Sache ja also das ganze Modell hat offensichtlich hier an der Stelle funktioniert und hat den Kontext den wir haben wollten verwendet aber noch ein Beispiel wir stellen die Frage how long dass die excelling Audit Log keep Login trees jetzt wird also schon ein bisschen spezifischer wir wollen etwas technisches Wissen dass wir vielleicht ansonsten irgendwo aus den Dokus fühlen müssen ja das Schöne ist die semantische Suche macht sich ans Werk findet ein Eintrag über das neue auto-log Feature in dem tatsächlich auch drin steht genau die Login schließt for six weeks und so weiter und die Antwort des Modells war die orderblock gibt's super ja es hat diesen Text genommen hat also Frage verstanden hatte Text verstanden und letztlich kommt hier äh ja der Outfit raus der irgendwo hier in diesem Text drin war kleiner Kommentar ähm ich mache das ganze hier auf Englisch deswegen weil Dolly hauptsächlich auf englischen Texten trainiert wurde ich kann es auch Deutsch befragen und vielleicht war irgendwo auch meine deutsche ein bisschen deutscher Text mit drin aber momentan ist in der Community die Vorläufermodelle sind immer alle erstmal besser auf Englisch und die Community macht sich dann als Werk das ganze quasi auch ganz schnell in alle anderen Sprachen zu übersetzen so kleinen Blick auf die Uhr muss ich sagen wir machen jetzt noch zwei kleine schnelle Beispiele und wechseln dann tatsächlich aus der Demo auch schon wieder raus hier ist noch mal eine schöne Frage die die ich ganz interessant finde und zwar keine Export wie extra den Audit Log die Antwort sagt you can download die Audit Log by clicking on the download button das steht hier in diesem Abschnitt da könnte die Display Lock Candy Safet und so weiter was ich aber wirklich toll finde ist hier steht the lock nicht excelling Audit Logs sondern the block Candy sscsv File nicht Export bei Klick in billig ja das heißt hier ist die semantische Suche tragen über eine Keywords suche hätten wir diesen Abschnitt hier nicht gefunden war die Modell aber sofort klar Moment ja safe Assassin's wie File und Export und Download sind ja eigentlich semantisch in dem Kontext hier vielleicht bedeutend also gebe ich eben entsprechend dieses Dokument aus und weiß dann auch dass der User diesen Abschnitt hier gemeint hat zu beantworten seiner Frage gebe ich also diesen Text aus der zwar nicht Export mehr enthält aber im Prinzip eben genau das ist was wir wissen wollen okay und dafür das ganze hier als Funktion geschrieben haben möchte ich Ihnen einmal nicht vorenthalten dass nicht immer hier so ein interaktives miteinander oder nicht ein interaktiver Umgang mit dem Modell notwendig ist wir können das ganze natürlich auch programmatisch machen ich kann die instruction schon im Vorfeld mitgeben ich kann auch die Suche den suchtring separat eingeben so dass ich andere

Dinge suche als das was instruction drin vorkommt und ich kann auch eingeben wie viele Resultate denn tatsächlich ausgegeben werden sollten also basierend auf der Art und Weise wie sie das umgesetzt haben steht Ihnen natürlich hier jetzt alles offen sie können ganz ganz ganz viel einfach machen wie sie das wollen sie haben einfach nur Sprachmodell in ihren Händen und können natürlich entwickeln drum rum wie sie das wollen und wenn ich das jetzt ausführe dann passiert was ganz spannend ist die anstruction selber war einfach nur samuraise der Text in dem Wissen dass das die Suchfunktion gleich Texte ausgeben wird die Suchfunktion selber hat nach dem Text suchen sollen welche Features excellent provided ja das bedeutet wir haben jetzt hier in den Referenzen eine ganze Reihe von Dokumenten aus den Release Notes ist das wo eben so ein paar Features vorkommen und das ist schon relativ viel Text jetzt und das Modell fast das zusammen provideos solution for building APIs Papa Funktions das ist hier dieser zweite Block ja Post Login Musical Standard das kommt hier unten vor und warte doch viel zu weit changes das haben wir eben schon gesehen das kommt hier unten vor das Modell hat also allein auf dem Hinweis Sunrise Text verstanden es ist ein Text über jede Menge Features also fasse ich auch mal zusammen was für Features ist okay um den Rahmen nicht zu sprengen mit Blick auf die Uhr wechseln wir dann jetzt leider einmal zurück in unsere Präsentation das heißt das war es jetzt schon mit der Demo können wir noch zu einem wichtigen Punkt und sein Angebot Angebotsvorstellung zur Umsetzung eines Proof of Concepts für ein selbst gehostetes LLM was wollen wir denn heute mitgeben natürlich neben der Tatsache dass sie hier hoffentlich ganz viel darüber gelernt haben was sprachmodelle sind und wie sie funktionieren und das sie sich im Open sourcebereich natürlich auch umsetzen lassen bieten wir Ihnen an sie damit nicht alleine zu lassen ja wir bieten also die Umsetzung eines individuellen Konzepts mit einem gehosteten darstellungsmodel in der Umgebung sowie einer Decke Dokument Governance solution mit dem SRD Haus das was ich also gerade eben erklärt hatte was wir umgesetzt haben und das Language selber das ja das bleibt ihnen offen wir haben jetzt aktuell Dolly gezeigt aber es hat sich schon wieder unglaublich viel auf dem Markt getan da geht's natürlich dann darum das richtige Sprachmodell für ihre Zwecke mit Umzug mit einzubeziehen z.B auch ein echtes konversationelles Modell das so Frage und Antwort spielte mitspielen kann so dass sie einen Chatbot für Kundenservice ist qash Service erhalten können ja das ist unser Angebot sprechen sehr gerne an wir beraten Sie natürlich wahnsinnig gerne und ähm ja wenn sie nach dem Webinar diesbezüglich Fragen haben einfach an solutions.com damit möchte ich mich einmal ganz herzlich bei Ihnen bedanken für Ihre Aufmerksamkeit bis hierhin und ja mein Team ist jetzt im Hintergrund sortieren Ihrer Fragen und da geben wir noch mal eine Sekunde bis

die Fragen eintrudeln na wir sind jetzt in der Frage Antwort Runde das heißt das ganze wird es ein bisschen freier okay die erste Frage ja die erste Frage haben wir schon wie lange würde die Umsetzung eines solchen proof of concepts dauern so das ist natürlich eine Frage die hängt ganz stark davon ab wie viel sie denn tatsächlich umgesetzt haben wollen wir können aber ganz klar sagen und System ist es eben gesehen haben inklusive vielleicht noch einer Web Oberfläche so dass sie nicht immer über die data Wutz Instanz mit dem Modell interagieren wollen die wird sobald etwa 10 bis 15 Personen Tagen landen das heißt wir werden das Modell mit ihnen hosten wir werden dafür sorgen dass die GPU Ressourcen auf Eltern zur Verfügung gestellt bekommen werden diese im Cluster also dann für Sie bereitstellen also das Modell informieren für Sie bereitstellen werden daraufhin mit ihnen einen dataik anbinden bzw anlegen Haus darauf aufziehen mit Hilfe von Ender bricks und quasi diese Analytics und gab es Doktor mit governance sozusagen umsetzen das ganze dann verbinden über eine semantische Suche und ja anstelle von diesem data Ricks Output das Ganze noch einmal in webuy Gießen und das wäre dann unser Proof of Concept wie gesagt 10 bis 15 PT ungefähr dafür aber natürlich offen für alle Ergänzungen die danach noch kommen genau ich hoffe das hat die Frage soweit beantwortet ich bin gern bereit für mehr fragen da kommt doch schon die nächste müssen wir an solche Sprachmodell noch auf eigene Daten trainieren damit es in der gezeigten Form antworten kann wie lange dauert so ein Training nein müssen sie nicht die gezeigte Form ist eben das was wir ganz besonders hervorheben wollten das ist das Dolly 3 Milliarden Parameter Modell jetzt ist es natürlich durchaus möglich dass sie entweder ganz wenig Text nur in der Firma haben das ist zum Training eignen würde und dann können Sie das Modell direkt so verwenden und es kann diesen Text so wie es in der Zeit in Form war analysieren wenn Sie viele Daten haben dann helfen wir aber beraten natürlich aber auch gerne dabei das Modell nach zu trainieren und auf bestimmten Tasks die sie die ihnen wie zum Beispiel dem Erstellen von Code und Datenbanken anzusprechen fein zu tun so nennt man dieses nachtrainieren und ja wie gesagt da da sind wir gerne auch Ihr Partner und da hatten sie da gerne aber notwendig per se ist das ganze nicht wir können das Modell also direkt heute noch einfach verwenden okay und damit ein Blick auf die Uhr tatsächlich fällt mir gerade auf wir sind ja schon bei 14:45 Uhr das heißt ich muss das Ganze leider hier beenden ich hatte natürlich sehr gerne noch ganz viel Zeit mit ihnen verbracht aber ich freue mich dafür darauf dass Sie und ich freue mich darüber dass sie so zahlreich entschieden sind dass sie so Weges Interesse gezeigt haben ich hoffe wie gesagt wir konnten Ihnen ein bisschen was heute vermitteln und noch einmal das Erinnerung falls Sie Interesse daran haben oder falls es auch noch weitere Fragen zu dem Thema haben gerne an solutions bemühen sobald wie möglich auf sie zurückzukommen damit

wünsche ich Ihnen einen schönen Tag verabschiede mich und ja noch einmal vielen
Dank dass Sie dabei waren