

Warum KI die falschen Fragen stellt (und was das für Führung bedeutet)

Sprachmodelle liefern schnelle Antworten – doch ihre Fragelogik weicht systematisch von menschlichem Denken ab. Eine neue Studie zeigt, warum vor allem Führungskräfte sehr sorgfältig abwägen sollten, wann sie Algorithmen in ihre Entscheidungsfindung einbeziehen.

Von [Arnaud Chevallier](#), [Frédéric Dalsace](#) und [José Parra Moyano](#) • 14.04.2026, 07.07 Uhr

Auf dem Markt für Gen-AI-Tools ist ein regelrechtes Wettrennen im Gange. Tools auf Basis großer Sprachmodelle wie ChatGPT, Claude und DeepSeek bieten ihren Kunden fast monatlich neue Funktionen. Und mit jeder neuen Ausbaustufe verschiebt sich der Anspruch: Aus reinen „Antwortmaschinen“ werden zunehmend „Gesprächsagenten“. Systeme also, die nachfragen, um am Ende nützlichere Antworten zu liefern.

Ein Beispiel ist WaLLM, ein für WhatsApp entwickelter Chatbot. Er begnügt sich nicht mit kurzen Antworten, sondern stellt Anschlussfragen, offeriert Listen mit aktuellen Trendanfragen und präsentiert sogar eine „Top Question of the Day“. Ähnlich funktionieren OpenAIs DeepResearch-Funktion und die in Singapur entwickelte KI Manus: Auch diese Systeme sind darauf ausgelegt, Nutzerinnen und Nutzer gezielt zu befragen. Wenn Anfragen mehrdeutig sind, treffen sie keine vorschnelle Annahme mehr, sondern stellen Rückfragen, um das Risiko von Frust oder Fehlern zu senken.

Diese Systeme können inzwischen Dialoge führen wie Berater, Kolleginnen oder Lernpartner. Indem sie nachfragen, sammeln sie die Informationen, die sie für passgenauere Antworten und Empfehlungen benötigen. Der Ansatz erinnert an gute Führung: Wer klügere Fragen stellt, trifft meist bessere Entscheidungen.

Trotz erster Analysen zur semantischen Struktur solcher Fragen ist bisher wenig darüber bekannt, welche Arten von Fragen diese Systeme tatsächlich stellen. Für Unternehmensleitungen stellt das ein Risiko dar. Denn die Fragen, mit denen Entscheidungen vorbereitet werden, steuern, welche Informationen gesucht werden – und prägen damit das Ergebnis. Deshalb ist es entscheidend, zu verstehen, ob sich die Fragen von Large Language Models (LLMs) von denen unterscheiden, die Führungskräfte stellen, und ob sie sich auch von Modell zu Modell unterscheiden. Beides beeinflusst Nutzerinnen und Nutzer unmittelbar – wenn dies für sie auch nicht offensichtlich ist.

inzu kommt: Weil sich agentische Funktionen rasant weiterentwickeln, geht es bei Fragen längst nicht mehr nur um das Zusammenspiel von Nutzer und Modell. Neue KI-Systeme antworten nicht nur, sondern organisieren Arbeitsschritte eigenständig: Sie zerlegen Aufgaben, planen Prozesse, prüfen Zwischenergebnisse, treffen Zwischenentscheidungen und stoßen – innerhalb ihrer Vorgaben – Handlungen an, um

ein Ziel zu erreichen. Dafür stellen sie sich – im wörtlichen Sinn – selbst Fragen. Wer versteht, wie ein Modell in solchen Prozessen fragt, kann dessen Empfehlungen besser einordnen – und gezielter bewerten, wo die KI hilft und wo sie womöglich in die Irre führt.

Dieser Artikel zeigt: Sprachmodelle fragen anders als Menschen. Und: Ein verlässliches Muster, das für alle Modelle gilt, gibt es nicht. Gerade wenn es um große Entscheidungen geht, sollten Führungskräfte deshalb den Einsatz von LLMs sehr bewusst abwägen und steuern. Wie das gelingen kann, zeigen wir im Folgenden.

Wie Führungskräfte fragen

Wenn es darum geht, relevante Informationen zu sammeln und bessere Entscheidungen zu treffen, standen gute Fragen schon immer im Zentrum. Wir wissen außerdem: Die Arten von Fragen, die Führungskräfte stellen, beeinflussen, welche Entscheidungen sie treffen. In unserer Forschung haben wir fünf Fragetypen identifiziert, die Entscheidungen unterstützen. Wir nennen dieses Set den „Leaders’ Question Mix“ (LQM).

1. **Investigative Fragen** („Was ist bekannt?“) klären Ursachen und verdeutlichen mögliche Lösungen. Sie helfen Entscheidungsträgern, das Problem – oder potenzielle Optionen – besser zu verstehen. Mögliche Fragen sind: „Was sind die grundlegenden Ursachen?“ „Wie umsetzbar ist diese Option?“
2. **Spekulative Fragen** („Was wäre, wenn?“) erkunden Alternativen und Szenarien. Sie stellen Annahmen infrage und prüfen, ob es einen anderen Weg geben könnte. Fragen wie „Was ist ein anderer Blick auf das Problem?“ oder „Was ist unser Plan B?“ fördern Innovation und erweitern den Möglichkeitsraum.
3. **Produktive Fragen** („Wie geht es jetzt weiter?“) bringen den Prozess in Bewegung. Sie zielen auf Plan, Ressourcen, Timing und Entscheidungsreife: „Haben wir, was wir für den nächsten Schritt benötigen?“ „Sind wir bereit, zu entscheiden?“ So werden Ideen zur Umsetzung gebracht.
4. **Interpretierende Fragen** („Und was bedeutet das?“) ziehen Schlüsse aus der Analyse. Sie lenken den Blick auf Konsequenzen und Einordnung: „Was haben wir gelernt?“ „Wie passt das zu unseren Zielen?“
5. **Subjektive Fragen** („Was bleibt unausgesprochen?“) machen emotionale oder politische Faktoren sichtbar – im Entscheidungsteam oder im Umfeld. Fragen wie „Was bereitet dir an dieser Entscheidung am meisten Sorge?“ oder „Wer wird dagegen sein, wenn wir weitermachen?“ holen oft jene Dynamiken an die Oberfläche, die Entscheidungen in der Praxis mitprägen.

Anders als Ärztinnen, Therapeuten, Journalistinnen oder Juristen lernen Führungskräfte in der Regel nicht, systematisch Fragen zu stellen. Sie verlassen sich häufig auf ihr Urteilsvermögen – und entwickeln mit der Zeit eine Art Standardrepertoire an Fragen. Neigen sie dabei stärker zu bestimmten Fragetypen und vernachlässigen andere,

können sie zu gänzlich anderen Schlussfolgerungen kommen als ihre Kollegen oder Kolleginnen – selbst wenn sie sich alle in derselben Situation befinden. Bedenklicher noch: Wer Fragen eines bestimmten Typs dauerhaft ignoriert, entwickelt leicht blinde Flecken. Unsere These: Effektive Führungskräfte können ihren inneren Autopiloten abschalten und ihren Fragenmix an die jeweilige Entscheidung anpassen, statt reflexhaft bei dem zu bleiben, was sich vertraut oder bequem anfühlt.

Um Führungskräften dabei zu helfen, die eigenen Vorlieben zu erkennen, haben wir den LQM-Test entwickelt. Zunächst stellten wir eine Liste typischer, im Entscheidungsprozess nützlicher Fragen für jeden Fragetyp zusammen. Den Teilnehmenden legten wir dann Fragenpaare vor, wobei die Fragen jeweils aus unterschiedlichen Kategorien stammten. Auf einer Skala von -3 bis +3 sollten die Teilnehmenden ihre Präferenz angeben. Dies wiederholten wir zehnmal, sodass jeder Fragetyp mit den anderen vier abgeglichen wurde. Anschließend übersetzten wir die Rohwerte in Prozentanteile, die zusammen 100 Prozent ergeben und den bevorzugten Fragenmix einer Führungskraft abbilden. Die Werte reichen von 0 Prozent (am wenigsten bevorzugt) bis 40 Prozent (am stärksten bevorzugt).

Für unsere Studie haben wir den LQM-Test mit Hunderten Führungskräften durchgeführt. Im Durchschnitt verteilen Führungskräfte ihre Fragen relativ gleichmäßig über die fünf Bereiche: Etwa 17 bis 22 Prozent ihrer Fragen entfallen auf jede Kategorie. Dies nutzten wir als Ausgangspunkt, um zu vergleichen, ob – und wie stark – KI-Chatbots in Entscheidungssituationen anders fragen als Menschen.

Wie KI-Systeme fragen

Stellen KI-Systeme dieselben Arten von Fragen wie erfahrene Führungskräfte? Verteilen sich diese Fragen ähnlich auf die fünf Kategorien? Um das zu prüfen, wählten wir 13 weitverbreitete LLMs aus dem Umfeld von Anthropic, OpenAI, Google, DeepSeek, Mistral, Perplexity und xAI. Mit jedem LLM führten wir denselben Test durch wie mit Führungskräften; jedes System musste den Test 200-mal absolvieren. Diese Ergebnisse verglichen wir mit den Antworten von mehr als 1600 Führungskräften, die die Selbsteinschätzung jeweils einmal vornahmen. Dabei zeigten sich vier Muster:

❓ LLMs und Führungskräfte nutzen einen unterschiedlichen Mix an Fragen. Bei allen von uns getesteten LLMs wich der Fragenmix in mindestens drei der fünf Kategorien signifikant von dem der Menschen ab. Während wir unsere Fragen relativ gleichmäßig über alle fünf Typen verteilen, streuen die LLMs deutlich stärker: Einige Kategorien werden überbetont, andere fallen unter den Tisch. Bei den Führungskräften lag die Spannweite zwischen den am stärksten und den am schwächsten gewichteten Kategorien bei 4 Punkten (von 17,8 Prozent bei subjektiven bis 21,8 Prozent bei

interpretierenden Fragen). Bei den LLMs war diese Spannweite häufig deutlich größer, in einem Fall lag sie sogar bei 23,1 Punkten – nämlich bei LLM Gemini 2.5, das interpretierende Fragen deutlich stärker bevorzugte als produktive.

☒ Bei interpretierenden und produktiven Fragen zeigen sich klare, wiederkehrende Unterschiede. 10 der 13 Modelle gewichteten interpretierende Fragen stärker als Führungskräfte. Und alle 13 Modelle stellten weniger produktive Fragen als Menschen.

☒ Die einzelnen Sprachmodelle unterscheiden sich stark voneinander. Wir haben jede mögliche Paarung von LLMs daraufhin geprüft, wie sehr sich die Verteilung der Fragen unterschied. Das Ergebnis: Bei allen 78 Paaren gab es statistisch signifikante Abweichungen – selbst dort, wo es sich um Weiterentwicklungen desselben Systems handelte (etwa Grok 3 und Grok 4).

☒ Insgesamt sind LLMs in ihrem Fragenmix konsistenter als Führungskräfte – aber nicht durchgängig. Meist fällt die Streuung innerhalb eines Modells geringer aus als bei Menschen. Zwei Modelle (Gemini 2.5 Pro und Grok 4) zeigten jedoch in mehreren Fragetypen eine höhere Varianz.

Warum diese Unterschiede wichtig sind

Die Unterschiede im Frageverhalten haben praktische Folgen. Wenn LLMs beim Brainstorming, beim Lösen von Problemen und bei der Unterstützung von Entscheidungen eine größere Rolle übernehmen, kann ein unausgewogener Mix an Fragen blinde Flecken erzeugen. Denn jede Fragekategorie lenkt die Aufmerksamkeit auf einen anderen Aspekt der Entscheidung. Wird einer davon nicht berücksichtigt, kommt er im Prozess womöglich gar nicht erst auf den Tisch. Nicht selten merken Führungskräfte erst spät, dass eine entscheidende Frage fehlt. Wie in den folgenden vier Beispielen:

Beispiel 1: Ein Modell, das selten produktive Fragen stellt – etwa Gemini 2.5 Pro oder ChatGPT 5 –, kann zwar Ideen liefern, lässt Teams aber allzu leicht zentrale Punkte übersehen: das richtige Tempo, die Verteilung von Ressourcen und/oder konkrete nächste Schritte. In schnelllebigen Umgebungen kann das Entscheidungen unnötig verzögern (oder auch zu vorschnellen Entscheidungen führen). Die Folge ist ein schlecht getakteter Prozess. Diese typische Managementfalle verschärft sich, wenn die KI nicht dazu beiträgt, den Entscheidungsprozess rechtzeitig nachzujustieren.

Beispiel 2: Eine KI, die nur selten subjektive Fragen stellt – wie etwa Sonar oder ChatGPT 5 –, kann den menschlichen Faktor leicht ausblenden. Dann fehlt womöglich genau das Nachhaken, das in Organisationen oft entscheidend ist: „Wie geht es den Stakeholdern mit dieser Veränderung?“ „Haben wir die Perspektive derjenigen eingeholt, die am stärksten betroffen sein werden?“ Derlei Versäumnisse können in problematischen Empfehlungen münden. Ein System kann etwa mit Verve eine Kostensenkungsinitiative

vorschlagen, ohne die Stimmung der Mitarbeitenden oder unausgesprochene Einwände deutlich zu machen. Führungskräfte spüren häufig intuitiv, wie es um Stimmung und Ausrichtung im Team steht. Verlässt sich eine Managerin zu stark auf die Empfehlung der KI, übersieht sie womöglich wichtige Signale – Signale, die sie früher davon abgehalten hätten, eine Entscheidung weiterzuverfolgen, die auf dem Papier überzeugt, aber wegen emotionaler oder politischer Faktoren zum Scheitern verurteilt ist.

Beispiel 3: Auch die Überbetonung investigativer und interpretierender Fragen, wie viele LLMs sie zeigen, kann zum Problem werden. Natürlich ist es wertvoll, Daten gründlich zu prüfen und einzuordnen. Doch wenn ein Modell zu lange in diesem Modus bleibt, kann eine Diskussion leicht in eine Analyse abdriften – oder bekannte Fakten wieder und wieder durchkauen. Erfahrene Moderatorinnen und Moderatoren von Diskussionen spüren, wann der Punkt erreicht ist, an dem es vom Verstehen ins Handeln gehen muss. LLMs fehlt dieses Gespür oft: Sie stellen weiter interpretierende Fragen („Was bedeutet das?“), obwohl das Team im Prozess schneller vorankäme, wenn es jetzt Entscheidungen vorbereiten und nächste Schritte festziehen würde.

Beispiel 4: Die Unterschiede zwischen den LLMs bedeuten: Schon die Wahl des KI-Modells kann die Perspektive eines Teams verschieben. Nutzt eine Organisation etwa eines, das selten spekulative Fragen stellt („Könnten wir das auch anders machen?“), werden innovative Alternativen womöglich nicht berücksichtigt, weil sie weder von Menschen noch von der Maschine ins Spiel gebracht werden. Umgekehrt kann ein Modell, das zu viele klärende, also investigative Fragen stellt, Nutzerinnen und Nutzer ausbremsen und frustrieren: Manche Nutzer und Nutzerinnen berichten, dass die Systeme hängen bleiben und immer wieder Bestätigung einholen, bevor sie weitermachen.

Damit das Fragen Wirksamkeit entfaltet, braucht es einen vielfältigen, ausgewogenen Mix – etwas, um das auch Führungskräfte ringen und das KI erst noch lernen muss. Kurz gesagt: Kippt ein System stark in Richtung eines bestimmten Fragetyps, kann es Recherchen und Austausch unbemerkt in eine einzige Richtung lenken – und so den Entscheidungsprozess negativ beeinflussen.

Was Chefinnen und Chefs tun können

Wenn LLM-Systeme anfangen, Fragen zu stellen, werden sie zu verführerischen Partnern. Viele Führungskräfte beantworten lieber ein paar Nachfragen, als lange am perfekten Prompt zu feilen. Man lehnt sich zurück, weil es sich anfühlt, als kümmern sich das System schon um den Rest – es wird ja nachfragen, wenn etwas unklar ist. Und plötzlich wirkt es, als reiche es, nur noch abzunicken, was das Modell vorschlägt.

Doch genau darin liegt das Problem: LLMs schlagen nicht nur ein mögliches „Menü“ an Fragen vor – sie bestimmen mit, welche Optionen überhaupt sichtbar werden.

Führungskräfte sollten daher nicht voraussetzen, dass eine „fragengetriebene“ KI automatisch alle relevanten Punkte abdeckt oder dieselben Fragen stellt wie ein Mensch. Damit KI-Fragen zu besseren strategischen Überlegungen führen – und nicht zu Fehlsteuerung oder endloser Analyse –, benötigt die KI weiterhin menschliche Führung. Im Folgenden skizzieren wir mehrere Schritte, mit denen sich LLM-gestütztes Fragen gezielt steuern und mit guter Entscheidungsfindung verbinden lassen:

Für Ihre ganz konkrete Entscheidung gilt: Lassen Sie sich nicht von Bequemlichkeit leiten. Fragen Sie sich immer: Was benötigen wir wirklich, um diese Entscheidung zu treffen? Überlegen Sie, ob Sie besser fahren, wenn Sie ein KI-System einbeziehen, oder ob der Prozess bewusst den Menschen überlassen werden sollte. Falls Sie KI nutzen: Entscheiden Sie, welche Art Unterstützung Sie brauchen – ein System, das gezielt nachfragt, oder eines, das in erster Linie Vorschläge liefert.

Wählen Sie das System, mit dem Sie arbeiten, bewusst aus. Unsere Forschung zeigt: Jedes LLM stellt andere Fragen – das gilt sogar für unterschiedliche Versionen desselben Modells. Achten Sie deshalb darauf, *wie* ein System fragt, und nehmen Sie eines, das zur Entscheidung passt, die Sie treffen müssen. Es kann sich lohnen, mehr als ein System heranzuziehen: So erhalten Sie verschiedene Blickwinkel und kommen eher zu den Fragen, die wirklich zählen. Besonders bei unternehmensinternen Systemen ist das wichtig – sonst wächst die Gefahr, dass sich Gruppendenken verfestigt oder bestehende Schief lagen und Vorurteile unbemerkt fortgeschrieben werden.

Behalten Sie die Kontrolle. Das heißt: Prüfen Sie die Vorschläge der KI konsequent. Solche Systeme wirken oft sehr sicher – auch wenn die Grundlage dafür fehlt. Sie können uns allzu leicht davon überzeugen, dass ein Ergebnis hochwertig ist, obwohl es das nicht ist. Zudem neigt die Technologie zu schmeichelnden, zustimmenden Formulierungen, die eher das bestätigen, was Nutzer ohnehin hören wollen. Es mag angenehmer sein, kompetent klingende Empfehlungen einfach zu übernehmen – besonders wenn sie in freundliche Sätze verpackt sind. Aber Führung bedeutet, zu tun, was notwendig ist, nicht was sich gerade bequem anfühlt.

Testen Sie die Vorschläge, die Sie bekommen. Fragen Sie sich nicht: „Kann ich das glauben?“ – um dann nach bestätigenden Belegen zu suchen. Fragen Sie stattdessen: „Muss ich das glauben?“ – und gehen Sie aktiv auf die Suche nach Gegenargumenten und Gegenbelegen. Zur Kontrolle gehört auch, die Zusammenarbeit regelmäßig nüchtern zu bewerten: Worin ist die KI stärker als ich? Worin bin ich stärker?

Stellen Sie auch den Entscheidungsprozess selbst auf die Probe. Führungskräfte haben selten genügend Zeit. Deshalb versuchen sie häufig, am Anfang eines Prozesses möglichst viel zu analysieren – und verschieben Entscheidungen auf den spätesten Zeitpunkt –, in der Hoffnung, die Unsicherheit dadurch so weit wie möglich zu reduzieren.

Eine Alternative besteht darin, den Zeitplan anders zu strukturieren und zwischen Entscheidung und Umsetzung einen festen Termin für ein Pre-Mortem einzuplanen – also eine kurze Übung, in der man so tut, als sei die Entscheidung bereits gescheitert, um mögliche Ursachen früh sichtbar zu machen. Dabei wird die getroffene Wahl gedanklich in die Zukunft verlängert: „Angenommen, wir entscheiden uns für Option X. Sechs Monate später hat sich das als Fehler erwiesen – welche Fragen hätten wir vorher stellen müssen?“ Diese Perspektive aus der Zukunft hilft, blinde Flecken aufzudecken, die im laufenden Prozess leicht übersehen werden – von Menschen ebenso wie von LLM-Assistenten.

Testen Sie die Vorschläge, die Sie bekommen. Fragen Sie sich nicht: „Kann ich das glauben?“ – um dann nach bestätigenden Belegen zu suchen. Fragen Sie stattdessen: „Muss ich das glauben?“ – und gehen Sie aktiv auf die Suche nach Gegenargumenten und Gegenbelegen. Zur Kontrolle gehört auch, die Zusammenarbeit regelmäßig nüchtern zu bewerten: Worin ist die KI stärker als ich? Worin bin ich stärker?

Stellen Sie auch den Entscheidungsprozess selbst auf die Probe. Führungskräfte haben selten genügend Zeit. Deshalb versuchen sie häufig, am Anfang eines Prozesses möglichst viel zu analysieren – und verschieben Entscheidungen auf den spätesten Zeitpunkt –, in der Hoffnung, die Unsicherheit dadurch so weit wie möglich zu reduzieren.

Eine Alternative besteht darin, den Zeitplan anders zu strukturieren und zwischen Entscheidung und Umsetzung einen festen Termin für ein Pre-Mortem einzuplanen – also eine kurze Übung, in der man so tut, als sei die Entscheidung bereits gescheitert, um mögliche Ursachen früh sichtbar zu machen. Dabei wird die getroffene Wahl gedanklich in die Zukunft verlängert: „Angenommen, wir entscheiden uns für Option X. Sechs Monate später hat sich das als Fehler erwiesen – welche Fragen hätten wir vorher stellen müssen?“ Diese Perspektive aus der Zukunft hilft, blinde Flecken aufzudecken, die im laufenden Prozess leicht übersehen werden – von Menschen ebenso wie von LLM-Assistenten.

Testen Sie die Vorschläge, die Sie bekommen. Fragen Sie sich nicht: „Kann ich das glauben?“ – um dann nach bestätigenden Belegen zu suchen. Fragen Sie stattdessen: „Muss ich das glauben?“ – und gehen Sie aktiv auf die Suche nach Gegenargumenten und Gegenbelegen. Zur Kontrolle gehört auch, die Zusammenarbeit regelmäßig nüchtern zu bewerten: Worin ist die KI stärker als ich? Worin bin ich stärker?

Fazit

In einer Arbeitswelt, in der KI-Systeme zunehmend Rückfragen stellen, kann die Qualität der Fragen, die wir von LLMs erhalten, ebenso wichtig sein wie die Qualität ihrer Antworten. Vorausschauende Führungskräfte nehmen das zum Anlass, die eigene Entscheidungsarbeit zu schärfen. Wenn wir nachvollziehen und steuern, wie diese Modelle fragen, – und die Führung behalten, wenn wir sie in Entscheidungsprozesse

einbinden – können wir von ihrer Perspektive profitieren und zugleich ihre blinden Flecken ausgleichen. Daraus kann eine neue Form der Zusammenarbeit entstehen: menschliches Urteilsvermögen, ergänzt durch KI-gestützte Einsichten, um die Fragen zu stellen (und zu beantworten), die bessere Entscheidungen möglich machen.

Ein oft Peter Drucker zugeschriebenes Bonmot lautet: „Die schwerwiegendsten Fehler sind nicht die, die wir machen, weil wir die falschen Antworten geben, sondern weil wir die falschen Fragen stellen.“ Im Zeitalter von KI bleibt genau das der Punkt: Die Verantwortung für die richtigen Fragen liegt beim Menschen – auch wenn ein System sie formuliert.