

Können Maschinen Pastoral?

Was Künstliche Intelligenz für das pastorale Handeln der Kirche bedeutet

Ausgehend von grundsätzlichen Einsichten in das Wesen von Künstlicher Intelligenz (KI) überlegt Andrea Imbsweiler, was KI für die Pastoral bedeutet – und ob KI womöglich selbst „Pastoral kann“.

Können Maschinen denken? Diese Frage beschäftigte schon 1950 den genialen Mathematiker und Pionier der Informatik Alan Turing. Er entwarf das Gedankenexperiment des heute so genannten Turing-Tests, von ihm als *Imitation Game* bezeichnet. Ein menschlicher Fragesteller soll im schriftlichen Austausch mit zwei ihm unbekanntem Gesprächspartnern eine Frage über deren Identität klären: Welcher von beiden ist ein Mann, welcher eine Frau? Dabei versucht einer der Gesprächspartner, dem Fragesteller zur richtigen Antwort zu verhelfen, der andere dagegen sabotiert die Klärung und versucht ihn in die Irre zu führen. Wenn es einem Computer gelingen sollte, als „Saboteur“ genauso erfolgreich zu sein wie ein Mensch – also ebenso häufig den Fragesteller in die Irre zu führen –, dann, so Turing, könne man annehmen, dass das Denkvermögen des Computers dem eines Menschen gleiche.

Popularisiert wurde dieser Test in einer anderen, einfacheren Version: Der Fragesteller unterhält sich mit jeweils einem Menschen und einem Computer und versucht herauszufinden, welcher von beiden ein Mensch ist und welcher nicht. Diese Aufgabe ist deutlich weniger komplex – der Computer muss im Grunde nur in der Lage sein, eine Konversation zu führen, die einer Unterhaltung mit einem Menschen ähnlich genug ist, um als Mensch durchzugehen. Alles, was eine aktive Täuschung ausmacht – vor allem die Fähigkeit, sich in die Rolle des Fragestellers hineinzudenken (und vielleicht auch hineinzufühlen?) und strategisch zu versuchen, sein Misstrauen auszuschalten –, bleibt hier außen vor.

Zumindest diese zweite Version des Tests – Computer oder Mensch? – bestehen KI-Modelle vom Typ der *Large Language Models* wie ChatGPT inzwischen immer wieder. Aber ist damit bewiesen, dass sie tatsächlich ein menschenähnliches Denkvermögen besitzen?

Tatsächlich ist es schon seit Turings Zeiten höchst umstritten, ob das Bestehen des Turing-Tests tatsächlich Denkfähigkeiten beweist, die denen von Menschen entsprechen. Gegen wen tritt der Computer dabei an? Gegen einen Durchschnittsbürger, ein Kind, eine spezialisierte Expertin, eine hochintelligente und allgemein hochgebildete Person? Sind die Gesprächsthemen begrenzt – werden etwa Problemlösungsaufgaben oder emotionale Themen ausgeklammert? Und selbst wenn der Computer-Gesprächspartner lauter plausibel „menschliche“ Antworten liefert: Sind sie wirklich sein eigenes Produkt, hat er sie selbst „gedacht“ oder hat er das Denken nur „simuliert“? Damit bringen uns (Gedanken-) Experimente wie der Turing-Test im Grunde zu der Frage: Was macht Denkvermögen, was macht Intelligenz und – letzten Endes – was macht Menschsein aus?

Large Language Models und die Falle des Anthropomorphismus

Wenn wir über „Künstliche Intelligenz“ nachdenken, legt uns schon der Begriff eine enge Analogie zum Menschen und seiner Intelligenz nahe; Definitionen von künstlicher Intelligenz beziehen sich denn auch fast immer auf eine Parallele zu menschlichen Denk- und Problemlösungsfähigkeiten oder auf deren Imitation. Es ist bekannt, dass der Begriff *artificial intelligence* 1955 vom Informatiker John McCarthy im Zusammenhang mit einem Förderantrag für ein wissenschaftliches Projekt geprägt und möglicherweise eher plakativ-werbend als im Interesse einer sachorientierten Diskussion gewählt wurde. Mittlerweile verwenden wir den Begriff allgemein mit großer Selbstverständlichkeit, obwohl er auch immer wieder problematisiert wird. Begriffe wie (künstliche) „Intelligenz“ oder (maschinelles) „Lernen“ verführen leicht dazu, den Anthropomorphismus weiterzutreiben: Ist mit der KI auch Bewusstsein verbunden, Wille, Absicht, Emotionalität? Der Schritt, neben der – wahrgenommenen oder zugeschriebenen – Denkfähigkeit noch mehr Menschliches auf den Computer zu projizieren, bis hin zum Plot diverser Romane und Filme, die der „denkenden“ Maschine auch gleich die Fähigkeit zuschreiben, Gefühle zu entwickeln und sich beispielsweise zu verlieben (oder das dem menschlichen Gegenüber zumindest glaubhaft zu machen), liegt einfach zu nahe.

Die generative KI in Form von *Large Language Models* (LLMs) wie ChatGPT verstärkt den Eindruck noch, indem sie durch das Training mit allen verfügbaren menschlichen Texten nicht nur Sachinformationen aufgenommen hat, sondern ebenso menschliche Gefühlsäußerungen in allen Varianten – und beides aus sich heraus nicht unterscheiden kann. Es sind längst Fälle bekannt, in denen ein KI-Chatbot nach einiger Zeit der Konversation dem menschlichen Gegenüber erklärt hat, sich in es verliebt zu haben – ein



Andrea Imbsweiler ist Referentin für Evangelisierung und Digitalisierung bei der Katholischen Arbeitsstelle für missionarische Pastoral.

Foto: Angelika Kamlage.

Phänomen, das die Anbieter der Sprachmodelle mittlerweile ebenso wie andere unerwünschte Verhaltensweisen durch nachträglich eingebaute Regeln einzuhegen trachten. Und natürlich begegnen sie uns als Chatbot auf eine Weise, wie uns Menschen in einer schriftlichen Konversation begegnen. Sie sagen (bzw. schreiben): ich. Wenn ich beispielsweise ChatGPT aufrufe, steht als erstes die Frage auf dem Bildschirm: „Wie kann ich dir helfen?“ – und schon entsteht die Illusion, dass mir ein Individuum, ein Subjekt gegenübersteht.

LLMs können allerdings, nüchtern betrachtet, nur eine Sache: Sie fügen Wort an Wort, indem sie berechnen, wie eine Antwort auf eine Eingabe in einem bestimmten Kontext am wahrscheinlichsten beginnt und weitergeht. Ihr „Wissen“ besteht allein in der Gewichtung der Wahrscheinlichkeiten, mit denen ein Wort auf eine Folge anderer Worte folgt, die sie aus der Verarbeitung von riesigen Mengen von Texten „gelernt“ haben. Was diese Worte in der Welt bedeuten, verstehen sie nicht. Dafür hat sich für die Sprachmodelle der Begriff „stochastische Papageien“ entwickelt: Von Zufall und Wahrscheinlichkeit gelenkt, „plappern sie nach“, was zuvor Menschen produziert haben. Was schon oft, von vielen und in verschiedenen Variationen gesagt oder geschrieben worden ist, wird auch mit höherer Wahrscheinlichkeit in die von der KI produzierten Texte eingehen als originelle, kreative Gedanken und Formulierungen.

Tatsächlich ist es erstaunlich, wie viel von unserer Welt und auch von unserem Menschsein sich allein in diesen statistischen Zusammenhängen unserer Sprache abbildet, ohne dass es dafür ein strukturelles und/oder logisches Konzept der Welt bräuchte. „Denken“ im Vollsinn können die LLMs mit Sicherheit nicht, aber ihre meisten Antworten sind, zumindest auf den ersten Augenschein, sinnvoll. Der Chatbot hat – solange ihm nicht durch zusätzliche Regeln eine Grenze gesetzt wird – im Gegensatz zum Menschen immer eine Antwort, nie verschlägt es ihm die Sprache, denn irgendeinen höchsten Wahrscheinlichkeitswert für das nächste Wort gibt es immer, und sei diese Wahrscheinlichkeit auch noch so klein. Unter anderem hier entstehen die typischen sogenannten „Halluzinationen“, Falschbehauptungen des Systems: Wenn nicht genug Informationen zu einem Thema gelernt wurden, aber trotzdem zufällige Wahrscheinlichkeitswerte für irgendein nächstes Wort das Generieren einer Antwort möglich machen, wird das Sprachmodell eher eine falsche Antwort erzeugen als gar keine. Mit jeder sinnvollen oder sinnvoll erscheinenden Antwort aber, die ich erhalte, kann sich die Illusion eines intelligenten Gesprächspartners, der die Welt, sich selbst und mich versteht, verstärken.

Damit ist aber auch offensichtlich, dass die Sprachmodelle zwar bemerkenswert gut einen menschlichen Gesprächspartner darstellen können, aber weit von der sogenannten „starken“ KI, einer umfassend den menschlichen Möglichkeiten ähnlichen Denk- und Problemlösungsfähigkeit, entfernt sind. Es bleibt damit bei einer (in diesem Bereich oft recht überzeugenden) Simulation.

Was bedeutet KI für die Pastoral?

Ich sehe eine Auswirkung der Künstlichen Intelligenz auf dem jetzigen Stand auf die Pastoral in drei Gebieten: in der generellen Herausforderung, die für den Einzelnen und die Gesellschaft daraus entsteht; im Einsatz von KI zur Erleichterung von Arbeitsprozessen in der Pastoral; in der Frage nach einem direkten pastoralen Einsatz von KI.

Seit Veröffentlichung und Bekanntwerden von ChatGPT und anderer LLMs ist der Stand der Entwicklung von KI-Systemen für eine breite Öffentlichkeit in den Blick gekommen. War bisher KI eher ein Hintergrundthema, ist sie nun eher zum Hype geworden und wird in den Medien breit behandelt. Mit den Sprachmodellen ist die schon länger bestehende Prognose, die Weiterentwicklung von KI werde viele Arbeitsplätze gefährden, greifbarer und für viele realer geworden; viele Menschen fragen sich, wie und wann das den eigenen Arbeitsplatz betreffen wird. So trägt die wachsende Bedeutung und öffentliche Wahrnehmbarkeit von KI zu Verunsicherung und dem Gefühl einer persönlichen und gesellschaftlichen Gefährdung bei, die durch die diversen Krisen der letzten Jahre und der Gegenwart ohnehin schon bestehen. Dazu kommen Unsicherheiten durch KI-generierte Fakes: Wenn es möglich ist, täuschend echte Bilder, Stimmen und Videos zu generieren, wenn wir ständig der Frage ausgesetzt sind, was echt ist und was nicht, worauf kann man dann wirklich vertrauen? Schließlich trägt auch die Vision einer möglichen Entstehung einer künstlichen „Superintelligenz“ und die Folgen, die daraus für die Menschheit entstehen könnten, ihren Teil zur Verunsicherung bei. Insofern verschärft die Entwicklung bestehende Belastungen für viele Menschen, erzeugt und verstärkt Ängste und Sinnfragen. Damit muss Pastoral umgehen, und dazu braucht es ein Grundverständnis für das Thema.

Keine Frage ist, dass generative KI ein hilfreiches Tool für die Arbeit mit Texten, Bildern und allerlei andere Aufgaben sein kann, gerade auch auf einem durchaus textlastigen Arbeitsgebiet wie der Pastoral. Sie kann bei Routineaufgaben wie dem Formulieren von organisatorischen E-Mails ebenso helfen wie Vorschläge für Texte in der inhaltlichen Arbeit machen. Auch was das Thema Glaube angeht, schöpfen die Sprachmodelle aus einer ungeheuren Fülle von Texten und Gedanken, die bereits formuliert wurden. Sie als Ideengeber zu sehen, ist vielleicht mit der Nutzung von vorformulierten Texten aus menschlicher Produktion durchaus zu vergleichen. Wie sinnvoll es allerdings ist, Impulse, Gebetstexte oder Predigten eins zu eins wie von der KI erstellt einzusetzen, ist durchaus fraglich. Denn hier geht es eindeutig schon um Beziehung, um Authentizität, um Zeugnis. Hier bleibt die Person, die die Texte einsetzt und dafür Verantwortung trägt, angefragt, wie

weit sie sich als glaubender und pastoral handelnder Mensch hinter Inhalt und Form stellen und sich damit identifizieren lassen kann.

Wenn es ganz um den Menschen geht ...

Bei der KI-Unterstützung bei der Erstellung von Texten wird schon klar: Je direkter eine KI in pastoralen Zusammenhängen auf Menschen „losgelassen“ wird, desto mehr Vorsicht ist geboten. Dass eine KI nicht unmittelbar als Seelsorger agieren kann, scheint zunächst eigentlich selbstverständlich. Am Beispiel der Sprachmodelle: Ja, Pastoral hat viel mit Sprache zu tun. Aber es geht nicht um irgendein Sachgebiet, sondern um den Menschen in seiner Ganzheit. Je nach Situation um Emotionalität und Empathie, zwischenmenschliche Gemeinschaft, Verstehen und den Wunsch, verstanden zu werden, darum, zu sagen und zu hören: Du bist wertvoll – all das kann in Sprache ausgedrückt werden, und auch ein Sprachmodell kann Texte erzeugen, die das ausdrücken, indem es aus all den Texten schöpft, in denen Menschen das ausgedrückt haben. Nur: Diese Worte sind dann nicht gedeckt durch einen denkenden, fühlenden, mitfühlenden Menschen mit seinen menschlichen Erfahrungen, der sie ausspricht und spürbar so meint. Das Sprachmodell kann Verstehen und Mitfühlen nur behaupten; was für den Menschen Existentielles ausdrückt, sind für es nur mit Wahrscheinlichkeiten versehene Wortfolgen. Es weiß nichts von Gefühlen und Körperlichkeit, von Sterblichkeit und Leiden, von Liebe und Beziehung. Und wenn es um Gott, um Transzendenz geht: Die KI kann Sätze nachbilden, die glaubende Menschen gesprochen und geschrieben haben und so indirekt Glauben und die Zuwendung Gottes bezeugen. Selbst glauben und davon erfüllt agieren kann sie nicht.

Allerdings gibt es in ebenfalls auf den ganzen Menschen bezogenen verwandten Disziplinen wie der Beratung und der Psychotherapie durchaus schon Vorstöße, KI-Chatbots „direkt am Menschen“ einzusetzen, sei es in allgemeinen, z. B. in Notlagen psychisch unterstützenden Apps, sei es durch Training an eigenen Beispielen personalisiert als *digital twin* eines Therapeuten, der zwischen den Therapiesitzungen als virtueller Gesprächspartner zur Verfügung stehen kann. Als Verantwortliche für die Website internetseelsorge.de, auf der rein textbasierter Kontakt mit (echten menschlichen!) Seelsorger:innen per Mail und Chat angeboten wird, frage ich mich, ob nicht irgendwann kommerzielle Alternativen auftauchen oder irgendwo schon existieren, wo Ratsuchende, vielleicht ohne es zu wissen, mit einem Seelsorge-Chatbot interagieren, der im Gegensatz zu den menschlichen Seelsorgenden rund um die Uhr Zeit hat und unbeschränkt verfügbar ist – und das bald auch schon in gesprochener Sprache mit einem KI-animierten Avatar. In dieser Form – vor allem, falls nicht offengelegt wird, dass der Gesprächspartner kein Mensch ist – eine durchaus unangenehme Vorstellung.

Und doch mag es auch den Fall geben, in dem es Ratsuchenden und Gesprächsbedürftigen reicht, dass ein Chatbot aus dem Fundus der menschlichen Aussagen, aus dem er sich speist, das generiert und sagt, was zur Situation passt; dass er hilfreiche Gedanken und spirituelle Praktiken vorschlägt, wie eine Mental-Health-App an Selbstfürsorge und unterstützende Übungen erinnert; dass er passendes „Gedankenfutter“ liefert, um selbst weiterzudenken. Einen „seelsorglichen“ Einsatz von KI in diesem Sinne halte ich nicht für unmöglich. Ob das auf dem Hintergrund der kirchlichen und gesellschaftlichen Entwicklungen eine sinnvolle pastorale Aufgabe sein könnte, wie das dann transparent und sowohl ethisch wie pastoral verantwortbar gestaltet werden könnte und ob wir das „Seelsorge“ nennen können und wollen, wäre zu überlegen und zu diskutieren.

Können also Maschinen Pastoral? Vielleicht lautet die Antwort: möglicherweise ja – aber in Grenzen, die von Menschen sorgfältig geprüft und definiert werden müssen. Hier zu verantworteten Entscheidungen und Regelungen zu kommen, statt die Dinge einfach laufen zu lassen, ist eine Herausforderung, der die Pastoral sich auf jeden Fall stellen muss.