

**Vestrucci, Andrea**

KI, Überzeugungen und Freiheit

**In:**

Thomas Laubach und Stefan Huber (Hrsg.), Braucht Demokratie Religion? : Auseinandersetzungen mit Hartmut Rosa, Bamberg: University of Bamberg Press, S. 277-293. 2025. DOI: 10.20378/irb-108272

**Beitrag im Sammelwerk - Verlagsversion**

DOI des Beitrags: 10.20378/irb-109909

Datum der Veröffentlichung: 28.08.2025

**Rechtehinweis:**

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt die **Creative-Commons-Lizenz CC BY**.



Die Lizenzinformationen sind online verfügbar:

<https://creativecommons.org/licenses/by/4.0/>

# KI, Überzeugungen und Freiheit

Andrea Vestrucci

Dieser Artikel skizziert einige Überlegungen, die mich beim Lesen von Hartmut Rosas Buch *Demokratie braucht Religion* inspiriert haben. Diese Überlegungen konzentrieren sich speziell auf die Interaktionen mit dem Thema Technologie. Wie verbinden sich die Erkenntnisse und Fragen in der Technologieforschung mit dem Thema dieses Symposiums: Der Bedeutung von Religion in unserer heutigen komplexen Welt und der demokratischen Gestaltung? Dieses Buch behandelt mehrere Aspekte dieser Bedeutung, die in Geschichte, Literatur, Bildung, Musik und Kultur verwurzelt sind - auch aus anderen Kulturen als der deutschen.

In Verbindung mit Forschungen, die in der Gruppe für *Artificial Intelligence Systems Engineering (AISE)* an der Uni Bamberg durchgeführt werden, möchte ich gerne ein weiteres Thema hinzufügen: Künstliche Intelligenz (KI) und ihre Beziehung zur Freiheit in religiösen und demokratischen Kontexten. Ich werde meinen Artikel in zwei Abschnitte unterteilen. Der erste Abschnitt trägt den Titel *Braucht KI Religion?* und der zweite Abschnitt, der auf dem ersten aufbauen wird, trägt den Titel *Braucht Religion KI?*.

Die Antwort auf beide Fragen lautet eindeutig „Ja“, sonst würde dieser Artikel jetzt enden. Das Problem besteht darin, herauszufinden, warum die Antwort „Ja“ lautet. Die Antwort hängt eng mit dem Begriff der *Überzeugungen* zusammen und damit, wie KI mit der Komplexität unserer Glaubenssysteme interagiert.

Natürlich ist es in einem kurzen Beitrag wie diesem unmöglich, solch komplexe Themen auszuschöpfen. Daher zielt mein Artikel darauf ab,

einige Punkte zu präsentieren, die inspirieren und zur weiteren Diskussion und Reflexion einladen können.

## 1. Braucht KI Religion?

### 1.1 KI und Argumente für Überzeugungen

Wir alle wissen, dass sich Überzeugungen von Wissen dadurch unterscheiden, dass sie nicht sicher sind. Wir sagen nicht „Ich glaube, dass  $2 + 2 = 4$ “, sondern „Das weiß ich“. Daher können Überzeugungen durch Argumente gestützt werden. Dies gilt insbesondere für religiöse Überzeugungen. Genau dieser Aufgabe widmet sich der Teil der Theologie, der *Apologetik* genannt wird.

Unter diesen Argumenten zur Stützung religiöser Überzeugungen gibt es eines, das eine lange und berühmte Tradition hat: Das ontologische Argument für die Existenz Gottes. Ein ontologisches Argument ist, kurz gesagt, ein Argument, das die Existenz Gottes nur anhand des Konzepts oder der Definition Gottes nachweist. Wenn wir uns auf eine Definition von Gott einigen, dann beweist das Argument, dass es widersprüchlich ist, die Existenz Gottes nicht zu bestätigen.

Es ist ein ebenso komplexes wie faszinierendes Argument – und diese Faszination erklärt seinen tausendjährigen Erfolg.

Seit mindestens 1078, als Anselm von Canterbury im *Proslogion* seinen ontologischen Gottesbeweis niederschrieb, haben menschliche Intelligenzen über diese Art von Argument nachgedacht, sie kritisch hinterfragt und stetig weiterentwickelt. Aber eigentlich nicht nur menschliche Intelligenz: Auch künstliche Intelligenz! Eine Art von KI, die mit expliziter Wissensdarstellung und Regeln arbeitet. Bei der Software handelt es sich insbesondere um (automatisierte und interaktive) *Reasoner* oder *Theorem-beweiser*. Ich beschränke mich hier auf drei Beispiele.

- Im Jahr 2011 kodierten Edward Zalta von der Stanford Universität und Paul Oppenheimer eine nichtmodale Formalisierung von Anselms ontologischem Argument in die Syntax des

Theorembeweislers erster Stufe Prover<sup>9</sup> (vgl. Oppenheimer/Zalta 2011; Oppenheimer/Zalta 2021).

- Seit 2013 arbeitet Christoph Benzmüller an Kurt Gödels ontologischem Argument aus dem Jahr 1970 in der automatisierten Argumentationsumgebung der Theorembeweiser höherer Stufe Isabelle HOL. Er untersuchte auch mehrere Varianten von Gödels Argumentation, die z. B. von Dana Scott, Anthony Anderson und Melvin Fitting vorgeschlagen wurden (vgl. Benzmüller/Woltzenlogel Paleo 2014; Benzmüller 2023; Vestrucci/Benzmüller 2024).
- Ich selbst habe sowohl an einer modalen Formalisierung von Anselms Argument als auch an der Formalisierung von Avicennas ontologischem Argument von 1027 gearbeitet, das von einigen als das erste ontologische Argument überhaupt angesehen wird und aus dem Islam und nicht aus dem Christentum stammt. Beide Automatisierungen werden in Isabelle HOL untersucht. Diese Artikel befinden sich noch in der Finalisierungsphase.

## 1.2 Die Frage nach der Wahrheit

Dies könnte eine Frage aufwerfen: Beweisen diese Erkundungen der Varianten des ontologischen Arguments in einer automatisierten Beweisumgebung endgültig, dass Gott wirklich existiert? Die Beantwortung dieser Frage erfordert möglicherweise einen kurzen Umweg in den Bereich der Philosophie der Mathematik. „Gott existiert (auch in der Realität)“ (und nicht „nur im Geist“ – wie das Argument von Anselm behauptet zu beweisen) ist eine Formel, die von den Theorembeweisern bewiesen wird. Der Beweis dieser Formel ist das Ergebnis zweier Prozesse: Eines syntaktischen Prozesses und eines semantischen Prozesses.

Der syntaktische Prozess überprüft, ob die Formel aus einigen anderen Aussagen, sogenannten Prämissen, abgeleitet werden kann, indem einige logische Regeln der Symbolmanipulation angewendet werden. Die grundlegendste Regel ist das berühmte *Modus Ponens*: Wenn  $p$ , dann  $q$ ;  $p$ ; dann  $q$ . Wenn es regnet, nehme ich den Schirm. Es regnet. Dann nehme ich den Schirm. Die Ableitung von „ $q$ “ aus „wenn  $p$ , dann  $q$ “ und „ $p$ “ wird Inferenz genannt. Die Inferenzen, die am Beweis des ontologischen

Arguments beteiligt sind, sind komplizierter als diese, aber im Wesentlichen ist das Verfahren dasselbe. Der Theorembeweiser überprüft, ob die Formel „Gott existiert (auch in der Realität)“ gemäß einiger logischer Inferenzregeln, die in die Maschine kodiert sind, aus den Prämissen des Arguments abgeleitet werden kann.

Der semantische Prozess macht etwas anderes: Er überprüft, ob das Argument gültig ist, d. h. ob unter der Annahme, dass die Prämissen wahr sind, die Schlussfolgerung – die Formel „Gott existiert (auch in der Realität)“ – auch wahr ist. Wenn hier von *Wahrheit* die Rede ist, kann dies zu Missverständnissen führen. Es bedeutet nicht, dass der Theorembeweiser beweist, dass Gott wirklich existiert. Im Kontext formaler Sprachen bedeutet „wahr sein“, „in einigen oder allen Modellen wahr zu sein“. Es ist irgendwie mit unserem intuitiven Wahrheitsbegriff verbunden, wird aber auf logisch/mathematische Formeln angewendet.

Ein *Modell* (oder auch eine *Interpretation*) besteht aus einer Menge von Objekten, dem sogenannten *Grundbereich*, und einer Zuordnung, die den nicht-logischen Symbolen (z. B. Prädikaten, Funktionssymbolen, Konstanten) Bedeutungen verleiht. Betrachten wir zum Beispiel die Formel: „Für alle  $x$  gilt: Wenn  $Q$  auf  $x$  zutrifft, dann trifft  $P$  nicht auf  $x$  zu.“ Wählen wir als Grundbereich *Tiere auf der Erde* und interpretieren  $Q$  als „ein Pinguin sein“ sowie  $P$  als „fliegen können“, dann ist die Formel wahr. Interpretieren wir dagegen  $Q$  als „ein Schwan sein“ und  $P$  als „schwarz sein“, so wird die Formel falsch. Eine Formel heißt *erfüllbar* (satisfiable), wenn es mindestens eine Interpretation gibt, in der sie wahr ist. *Gültig* (valid) ist eine Formel, wenn sie in *allen* Interpretationen wahr ist. Die oben genannte Formel, „Für alle  $x$  gilt: Wenn  $x$   $Q$  ist, dann trifft  $P$  nicht auf  $x$  zu“, ist im Grundbereich *Tiere auf der Erde* erfüllbar, jedoch nicht gültig, da es mindestens eine Interpretation gibt, in der die Formel falsch ist. Solche Interpretationen, in denen eine Formel nicht erfüllt wird, bezeichnet man als *Gegenbeispiele*. Zurück zum ontologischen Argument: In einem Theorembeweiser kann man zunächst überprüfen, ob die Prämissen des Arguments widerspruchsfrei (konsistent) sind, denn aus einem Widerspruch kann beliebig alles bewiesen werden (ex contradictione sequitur quodlibet), was das Argument trivialisieren würde. Wenn die Prämissen als konsistent erkannt werden, nimmt sie der Theorembeweiser bei

seiner Suche nach einem formalen Beweis als gegeben an und versuchen kann, ein Gegenbeispiel für die Schlussfolgerung „Gott existiert (auch in der Realität)“ zu finden – also ein Modell, in dem alle Prämissen gelten, die Schlussfolgerung jedoch falsch ist.

Existiert kein solches Gegenbeispiel, lässt sich zeigen, dass die Schlussfolgerung logisch aus den Prämissen folgt. Dabei bedeutet ein „Beweis“ innerhalb des Theorembeweisers, dass es keine Interpretation gibt, in der sämtliche Prämissen wahr und die Konklusion zugleich falsch ist. Ob die Formel in einem „realen“ oder philosophischen Sinn wahr ist, hängt jedoch davon ab, ob man die Prämissen als wahr im außermathematischen bzw. außermodelltheoretischen Kontext akzeptiert. Der Theorembeweiser kann nur über jene Modelle „entscheiden“, in denen die Prämissen bereits als gültig vorausgesetzt werden; außerhalb dieses Rahmens (z.B. in Modellen oder Wirklichkeiten, in denen die Prämissen nicht gelten) kann er nichts über die „Wahrheit“ von „Gott existiert (auch in der Realität)“ aussagen.

Die Antwort auf die Frage, ob die Formel tatsächlich wahr ist, hängt also davon ab, welche Gründe man anführt, um die Prämissen für wahr zu halten. In Argumenten, die sich mit metaphysischen oder abstrakten Objekten befassen, bleibt die Frage nach der Wahrheit von Formeln (Sätzen) oft umstritten, weil es per Definition nicht möglich ist, sie empirisch zu überprüfen (anders als etwa bei Formeln über Tiere).

Daher stehen uns für die Beurteilung der Wahrheit der Prämissen zwei, nicht gegenseitig ausschließende Optionen offen: intrinsische und extrinsische Unterstützung (vgl. Maddy 1988). Die intrinsische Unterstützung bedeutet, dass Prämissen etwa aus offensichtlichen Gründen oder durch weitgehenden Konsens für wahr gehalten werden; so kann es sich bei dem ontologischen Argument um Glaubensartikel handeln (wie bei Anselm). Die extrinsische Unterstützung verweist auf den Nutzen solcher Prämissen: Man führt sie ein, um bestimmte Thesen zu formulieren oder gewünschte Sachverhalte zu erklären (eine Art empirischer bzw. bottom-up-Vorgehensweise).

Aus der Geschichte der Mathematik – etwa der Mengenlehre – ist bekannt, dass sich beide Arten von Unterstützung häufig durchdringen: Ein zunächst kontraintuitives Axiom kann sich als nützlich für die Bildung

und Überprüfung neuer Hypothesen erweisen und dadurch allmählich an Plausibilität gewinnen. Umgekehrt können Prämissen, die intuitiv erscheinen, sich als notwendig erweisen, um relevante mathematische Eigenschaften oder Objekte formal zu beschreiben.

Dies bedeutet zugleich, dass die Wahrheit der Prämissen immer auch in Frage gestellt werden kann. Wer die Prämissen ablehnt, kann dem Argument seine Schlüssigkeit (im Sinne von „Überzeugungskraft“) absprechen, weil sich die Konklusion natürlich nicht auf falsche oder unbegründete Voraussetzungen stützen soll. Formal bleibt das Argument jedoch weiterhin gültig, falls es keine Interpretation gibt, die alle Prämissen wahr und die Schlussfolgerung falsch macht. In vielen Fällen kann es allerdings fruchtbarer sein, die Prämissen als wahr anzunehmen, weil das Argument interessante Strukturen unseres Denkens oder metaphysische Zusammenhänge beleuchtet. Schließlich gilt auch: Wer also die Prämissen, die das ontologische Argument stützen, ablehnt, stützt diese Ablehnung auf andere Prämissen, die ihrerseits hinterfragt werden können. Anders ausgedrückt: Die bloße Verneinung der ursprünglichen Prämissen mag manchmal mehr intellektuelle Kraftanstrengung als ein klarer logischer Fortschritt sein – insbesondere, wenn das eigentliche Ziel darin besteht, mithilfe der Argumentation die inneren Zusammenhänge einer Theorie oder eines Systems zu erhellen.

### 1.3 Expressivität

Warum sind diese Untersuchungen von Varianten des ontologischen Arguments in verschiedenen automatisierten Argumentationsumgebungen relevant?

Ontologische Argumente sind komplex. Zum Beispiel funktioniert Gödels Argumentation in einem logischen Kontext, der *Logik höherer Stufe* genannt wird und in dem nicht nur Einzelvariablen, sondern z.B. auch Eigenschaften quantifiziert werden können. Oder Anselms Argument könnte – abhängig von der Formalisierung – die Verwendung der Diagonalisierung beinhalten, einer mathematischen Methode, bei der eine Eigenschaft mit der Eigenschaft selbst verknüpft wird (vgl. Oppenheimer/Zalta 2011). Oder Anselms Arbeit beinhaltet eine zusätzliche Klärung,

welche Eigenschaften binärer Beziehungen erforderlich sind, um das Argument zu beweisen.

Darüber hinaus basieren Versionen des ontologischen Arguments, wie bereits erwähnt, auf einer nichtklassischen Logik, genauer gesagt auf der Modallogik, der Logik der Möglichkeit und Notwendigkeit. Diese verschiedenen Aspekte der Komplexität erfordern von der Maschine einen erheblichen Aufwand, um das Argument zu verstehen. Tatsächlich liegt der eigentliche Aufwand jedoch nicht auf Seiten der Maschine, sondern auf Seiten des Informatikers oder Logikers, der die Expressivität der Maschine bis an ihre Grenzen bringen muss.

*Expressivität* bedeutet die Fähigkeit einer Sprache – in diesem Fall die Sprache des automatisierten KI-Denkens –, Informationen aus einer anderen Sprache – in diesem Fall natürlicher Sprache – ohne Bedeutungsverlust zu vermitteln, d. h. durch die Erfassung von Nuancen und die Berücksichtigung von Polysemie und die Unterscheidung semantischer Kontexte.

Nachfolgend eine überarbeitete Fassung, bei der die vormals englischen Passagen ins Deutsche übertragen und stilistisch angepasst wurden:

Ich beschränke mich hier auf ein Beispiel: Wie bereits erwähnt, verwenden einige Versionen des ontologischen Arguments eine nicht-klassische Logik, nämlich die Modallogik. Das KI-System (bzw. der Theorembeweiser) hingegen beruht auf einer klassischen Logik (gewöhnlich höherstufig), wie sie typischerweise für den Beweis mathematischer Theoreme eingesetzt wird. Um modale Argumente – etwa das ontologische Argument – in einem solchen System zu formalisieren, muss man die Modallogik innerhalb der klassischen Logik repräsentieren. Dies kann beispielsweise mithilfe der flachen semantischen Einbettung (englisch: *Shallow Semantic Embedding*) erfolgen, die Elemente der Modallogik direkt auf Konstrukte der klassischen Logik abbildet. Auf diese Weise werden die nativen semantischen Mechanismen genutzt, um modallogische Formeln hinsichtlich ihrer Bedeutung und Wahrheit zu evaluieren, was ihre rechnerische Handhabung erleichtert (vgl. Benz Müller 2019). Man bezeichnet diesen Prozess als „flach“, weil er keine tiefgreifende, syntaktische Transformation der Modallogik erfordert, sondern vielmehr einer Neuformulierung gleichkommt. In Bild 1 (gezeigt in *Isabelle/HOL*) etwa

werden die logischen Operatoren der Modallogik direkt als höherstufige Funktionen umgesetzt, sodass keine explizite Kodierung der modallogischen Syntax nötig ist.

```

typedecl w (*Possible worlds*) typedecl i (*Individuals*)
type_synonym  $\sigma$  = "w $\Rightarrow$ bool" (*Type of world-lifted propositions, truth-sets*)
(**Logical operators lifted to truth-sets**)
abbreviation mbot::" $\sigma$ " ("⊥") where "⊥  $\equiv$   $\lambda$ w. False"
abbreviation mtop::" $\sigma$ " ("⊤") where "⊤  $\equiv$   $\lambda$ w. True"
abbreviation mnot::" $\sigma \Rightarrow \sigma$ " ("¬" [52]53) where "¬ $\varphi \equiv \lambda$ w. ¬( $\varphi$  w)"
abbreviation mand::" $\sigma \Rightarrow \sigma \Rightarrow \sigma$ " ("∧" [51]51) where " $\varphi \wedge \psi \equiv \lambda$ w. ( $\varphi$  w) ∧ ( $\psi$  w)"
abbreviation mor ::" $\sigma \Rightarrow \sigma \Rightarrow \sigma$ " ("∨" [50]50) where " $\varphi \vee \psi \equiv \lambda$ w. ( $\varphi$  w) ∨ ( $\psi$  w)"
abbreviation mimp::" $\sigma \Rightarrow \sigma \Rightarrow \sigma$ " ("→" [49]49) where " $\varphi \rightarrow \psi \equiv \lambda$ w. ( $\varphi$  w)  $\rightarrow$  ( $\psi$  w)"
abbreviation mequ::" $\sigma \Rightarrow \sigma \Rightarrow \sigma$ " ("↔" [48]48) where " $\varphi \leftrightarrow \psi \equiv \lambda$ w. ( $\varphi$  w)  $\leftrightarrow$  ( $\psi$  w)"
abbreviation mident::" $i \Rightarrow i \Rightarrow \sigma$ " ("=") where " $x = y \equiv \lambda$ w. (x = y)"

```

Bild 1: Beispiel für eine flache semantische Einbettung der Modallogischen Operatoren in die Logik höherer Ordnung im Isabelle/HOL-Theorembeweiser.

## 1.4 KI und Religion

Braucht KI also Religion? KI kann sicherlich davon profitieren, sich mit religiösem Denken auseinanderzusetzen: Die KI-Forschung nutzt Argumente zur Verteidigung religiöser Überzeugungen als Chance, das logische Denken in KI-Systemen voranzutreiben und ihre Fähigkeit zu verbessern, mit komplexen, abstrakten Konzepten umzugehen und auf eine Weise zu argumentieren, die eher dem menschlichen Verständnis entspricht.

## 2. Braucht Religion KI?

### 2.1. KI und Freiheit

Wir wissen, dass die Pandemie Religionsgemeinschaften dazu gezwungen hat, alternative Formen des Gottesdienstes und der Versammlung zu finden, was in den meisten Fällen auf Technologie zurückzuführen ist. Aber darüber möchte ich nicht sprechen. Vielmehr möchte ich über die Beziehung zwischen Technologie, Religion und Freiheit sprechen.

Künstliche Intelligenz ist immer mehr Teil unseres täglichen Lebens. Schüler und Lehrer profitieren von ChatGPT, haben jedoch auch

Probleme damit. Und das ist nur die Spitze des Eisbergs. Von der Gesundheitsversorgung bis zu Mobiltelefonen, von Videospiele über Einstellungsverfahren bis hin zu Finanzhilfen ist KI ein integraler Bestandteil unseres Lebens.

Technologie ist eine Form der Freiheit. Sie befreit uns von einigen Pflichten und ermöglicht uns, mehr Zeit für uns selbst zu haben. Denken Sie an alle Transportmittel oder die Nutzung von Elektrogeräten. Dies hat jedoch seinen Preis – nicht nur den Strompreis oder den CO<sub>2</sub>-Fußabdruck. Ein eher immaterieller Preis. Technologie erweitert nicht nur unsere Freiheit, sie schränkt sie gleichzeitig ein, weil sie uns von ihr abhängig macht. Wir gewöhnen uns an die Technologie; der Einsatz von Technologie ist fast eine Gewohnheit, so dass wir uns ein Leben ohne sie nicht mehr vorstellen können. Dies kann auch Auswirkungen auf unser Engagement für demokratische Teilhabe haben.

## 2.2. Logik und Rhetorik

Ein wichtiges Beispiel dafür ist das, was ich *künstliche Narrative* nenne (Lattanzi/Vestrucci 2024). Es handelt sich nicht um Science-Fiction, sondern um etwas absolut Reales und Alltägliches. Nehmen wir einen Fall, der alltäglich sein könnte. Wenn Sie ChatGPT eine logische Frage stellen, ist es möglich, dass seine Antwort nicht korrekt ist. Im Jahr 2024 habe ich die damals kostenlos verfügbare Version von ChatGPT (ChatGPT 3.5) gebeten, einen Syllogismus zu generieren, der zwar semantisch korrekt (also wahr), aber syntaktisch nicht korrekt (also logisch ungültig) ist – siehe dazu Bild 2. Die generierte Schlussfolgerung lautet: „Alle Säugetiere legen Eier. Menschen sind Säugetiere. Folglich legen Menschen Eier.“ Natürlich ist diese Schlussfolgerung auch semantisch falsch, da die erste Prämisse falsch ist: Nur wenige Säugetiere legen Eier (Schnabeltier und Ameisenigel). Warum irrt sich ChatGPT auf diese Weise? (Wichtig anzumerken ist, dass dieses Modell von ChatGPT mittlerweile aufgrund technologischer Verbesserungen nicht mehr verfügbar ist. Dies hebt jedoch weder die Existenz dieses Ergebnisses auf, noch schließt es aus, dass auch in Zukunft kontroverse oder problematische Resultate auftreten könnten.)

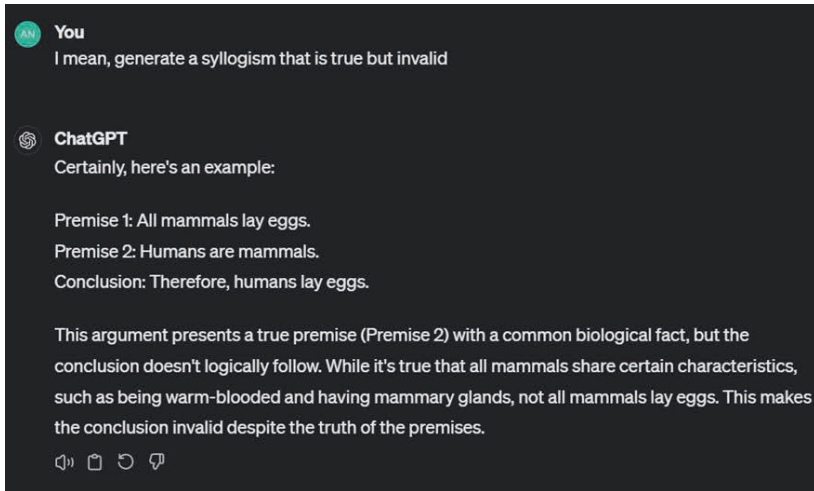


Bild 2: Ein Chat mit ChatGPT Model 3.5

ChatGPT ist eine Art von KI, die nicht mit expliziten semantischen Regeln arbeitet – zum Beispiel Regeln zur Suche nach Gegenmodellen, wie in Abschnitt 1.2 für Theorembeweiser erwähnt –, weil die Semantik seiner Antworten probabilistisch ist: Vereinfacht ausgedrückt könnte man sagen, dass die Semantik von ChatGPTs Antworten auf einer Analyse des Kontextes und der Vorhersage von Wortfolgen basiert, basierend auf der Wahrscheinlichkeit ihrer Verteilung in der riesigen Datenbank, auf der ChatGPT trainiert wurde. Das bedeutet, dass die Semantik von ChatGPTs Antworten nicht aus einem echten Verständnis der Bedeutung resultiert, sondern aus Operationen, die die statistische Neuordnung von Teilen seines enormen Wissensbestands einschließen. Folglich besitzt ChatGPT keine abstrakte Vorstellung davon, was der Sinn seiner Antworten ist, weil es keine Basis expliziter Regeln für den Aufbau der Semantik seiner Entscheidungen hat.

Diese fehlende explizite Regelung, die es ermöglicht, die semantische Gültigkeit der gegebenen Antworten zu bewerten, bedeutet etwas Wichtiges: KIs wie ChatGPT haben im übertragenen Sinne kein Verständnis dafür, was sie sagen und entscheiden. Aus diesem Grund wurde diesen KIs die Metapher des „stochastischen Papageis“ (vgl. Bender et al. 2021)

zugeschrieben: Der stochastische Papagei schafft nichts Neues, sondern formuliert einfach Antworten, indem er das, was er gelernt hat, auf probabilistische Weise umstellt. Genauer gesagt formuliert der stochastische Papagei Antworten, deren Korrektheit nur wahrscheinlich ist, und diese Wahrscheinlichkeit basiert zunächst auf der Menge und Qualität der Daten, die er gelernt hat. Gerade vor diesem Hintergrund hat Luciano Floridi das Akronym *AI* scherzhaft, aber nicht zu sehr, als „*Agere sine Intelligere*“ neu formuliert, was auf Lateinisch „Handeln ohne zu verstehen“ bedeutet (vgl. Floridi 2023).

Was ist das Problem mit dem stochastischen Papagei? Das Problem betrifft unser Vertrauen in die Antworten dieser KIs. Ist unser Vertrauen gut platziert? In dem Beispiel, das wir gemacht haben, hat ChatGPT eine falsche, aber rhetorisch sehr gut formulierte Antwort generiert. Also stellt sich im Fall von ChatGPT die Frage des Vertrauens in diesen Begriffen: Wie tief ist die Unterscheidung zwischen Logik und Rhetorik in den Antworten der Maschine?

Wie erwähnt, formuliert ChatGPT Antworten – das heißt Entscheidungen – auf der Grundlage eines statistischen Ansatzes zur Semantik, zum Beispiel basierend auf der Verteilung von Worten und ihrer mehr oder weniger häufigen Verwendung in den Texten, auf denen ChatGPT trainiert wurde. Daher erscheinen ChatGPTs Antworten vollkommen korrekt, aber in einigen Fällen ist diese Korrektheit eben nur oberflächlich, da es keinen Vergleich mit expliziten semantischen Regeln gibt. Die falsche Antwort erscheint jedoch korrekt, weil sie sehr gut argumentiert ist. Noch genauer gesagt verbergen die Antworten von ChatGPT eine potenzielle logische Unkorrektheit unter einer sehr überzeugenden Rhetorik.

Man könnte einwenden, dass dies kein allzu großes Problem darstellt, da es sich nur um einige spezifische Antworten von ChatGPT handelt. Dennoch sollten wir uns fragen, was ein durchschnittlicher Benutzer tun würde, der ChatGPT eine Frage stellt, ohne zweideutige Absichten wie zum Beispiel die Überprüfung seiner logischen Fähigkeiten. Wie würde dieser Benutzer auf eine rhetorisch so gut formulierte Antwort reagieren? Würde er möglicherweise die logische Richtigkeit der Antwort überprüfen oder würde er dem vollkommen plausiblen Urteil von ChatGPT vertrauen?

### 2.3. Überlegungen und künstliche Narrative

Stellen Sie sich eine zukünftige KI vor, die noch leistungsfähiger ist als ChatGPT und nicht nur kurze Antworten von ein paar Zeilen, sondern ganze Texte oder Videos produziert, wie beispielsweise journalistische Reportagen, die der breiten Öffentlichkeit eine rhetorisch perfekte, aber logisch fehlerhafte oder der Realität nicht entsprechende Erzählung präsentieren. In gewisser Weise ist dies bereits Realität: Denken Sie an sogenannte Deepfakes, also von KI generierte Videos, die so realistisch sind, dass sie wie Aufnahmen echter Ereignisse wirken. Darüber hinaus kann ChatGPT (auch die aktuelle Version) unsere Vorlieben, Gewohnheiten und Erwartungen kennen, basierend auf den Kommentaren, die wir zuvor zu seinen Antworten abgegeben haben. Daher könnte diese hypothetische Erzählung, die von einer zukünftigen KI ähnlich ChatGPT generiert wird, sogar so kalibriert sein, dass sie unsere Erwartungen erfüllt.

Auf diese Weise könnte die KI nicht nur einzelne Entscheidungen beeinflussen, sondern auch unsere Überzeugungen tiefgreifend umgestalten und somit unser gesamtes Orientierungssystem in der Welt beeinflussen. Tatsächlich spielen unsere Überzeugungen eine grundlegende Rolle bei unseren Entscheidungen. Insbesondere, wenn wir die klassische Unterscheidung zwischen „glauben dass“ und „glauben an“ wieder aufgreifen, denken wir an die Macht, die Erzählungen wie politische Ideologien, religiöse Überzeugungen und folkloristische Ideen auf unsere Entscheidungen haben. Abhängig von der Stärke – der „Verwurzelung“ – unserer Überzeugung in diesen Erzählungen können diese tatsächlich unser Leben und Verhalten sehr tiefgreifend beeinflussen.

Denken Sie zum Beispiel an die japanischen Kamikaze-Piloten des Zweiten Weltkriegs, an Religionskriege (oder zumindest solche, die durch religiöse Motive gerechtfertigt sind), und an die Rivalitäten, die sich zwischen Individuen nur aufgrund von Unterschieden in Herkunft und Nationalität entwickeln. Aber es gibt auch positive Beispiele: die Bindungen zwischen Menschen aus verschiedenen europäischen Ländern, die durch die Existenz der Europäischen Union, also durch den Glauben an ein gemeinsames europäisches Volk, begünstigt werden; die geistige Reife oder die Fähigkeit, Lebensereignisse angesichts religiöser Überzeugungen zu relativieren, auch im Licht von meditativen Übungen und Gebeten.

Denken Sie daran, wie die menschliche Geschichte durch den Glauben an diese Erzählungen verändert wurde. Wie Milliarden von Menschenleben im Laufe der Geschichte vorzeitig durch Handlungen, die von diesen Überzeugungen beeinflusst wurden, ausgelöscht wurden. Und es handelte sich „nur“ um Erzählungen, die von unseren Mitmenschen geschaffen wurden. Ist es möglich, dass diesen Erzählungen in naher Zukunft künstliche Erzählungen hinzugefügt werden, die von einer Maschine erfunden wurden?

Tatsächlich beruhen unsere Überzeugungen nicht auf Beweisen oder Bestätigungen ihrer Gültigkeit, sondern hängen ausschließlich von ihrer Kompatibilität mit bestimmten Erwartungen von uns ab, von ihrem narrativen Einfluss und ihrer rhetorischen Kraft (vgl. Kahneman 2011). Und wie gesehen, können wir uns eine KI vorstellen, die eine rhetorisch perfekte und daher absolut überzeugende, aber logisch fehlerhafte Erzählung produziert. Ohne angemessene Überprüfung würde uns die ausgezeichnete Rhetorik der KI dazu verleiten, dieser künstlichen Erzählung zu vertrauen. Welche Auswirkungen kann diese künstliche Erzählung auf unsere Überzeugungssysteme und damit auf unsere Entscheidungen und unsere Beteiligung an der Welt und am öffentlichen Leben haben?

Daher stellt sich erneut die Frage nach der Freiheit. KI befreit uns von alltäglichen Aufgaben wie dem Erstellen von Texten (ChatGPT), aber diese Freiheit birgt ein Risiko: das Risiko, dass die von KI generierten Antworten nicht nur Zeit sparen, sondern auch unsere Überzeugungen und damit unsere Freiheit als autonome Teilnahme am gemeinsamen Leben formen. Die Freiheit, künstliche Intelligenz zu nutzen, könnte daher zu einer Knechtschaft führen: der Knechtschaft, unsere Überzeugungen als abhängig von künstlichen Erzählungen zu betrachten.

#### **2.4. Freiheit als Knechtschaft**

Diese Bindung, diese Unfreiheit kennt das Christentum seit Jahrhunderten: die Knechtschaft des Willens. Martin Luther verwendet dieses Bild, um auf Erasmus' Verteidigung des freien Willens des Menschen in Bezug auf Erlösung und Verdammnis zu antworten. Unser Wille ist nicht nur unter Satan, sondern auch unter Gott gebunden. Unsere Seele ist ein

Pferd, das entweder von Satan oder von Gott geritten wird – aber trotzdem geritten, niemals frei.

Für Luther gibt es jedoch einen Unterschied, ob die Seele von Satan oder von Gott geritten wird. Der Unterschied besteht darin, dass Satan uns glauben lässt, wir seien frei, während wir in Wirklichkeit von Satan bewegt werden.

Im Gegensatz dazu offenbart Gott, dass unser Wille von Gott gebunden ist. Und diese Offenbarung ist laut Luther die Freiheit des Menschen, weil sie uns von dem Gedanken befreit, dass wir Gott sind und dass unsere Erlösung in unseren Händen liegt (vgl. Vestrucci 2019, 34-35, 62-65).

Somit gibt es sowohl in der Religion als auch durch den Einsatz von KI keine Freiheit ohne Abhängigkeit und umgekehrt. Die Befreiung von einer Last geht mit der Abhängigkeit von der Entität einher, die von der Last befreit: Gott einerseits und die Maschine andererseits.

Es gibt jedoch einen Unterschied. Bei der Freiheit, die durch den Einsatz von KI versprochen wird, steht der Nutzende, der Technologiekonsumierende, im Mittelpunkt. Andererseits stellt die Freiheit, die die Religion – oder vielmehr das Christentum – verspricht, die Gottheit in den Mittelpunkt: Sie richtet die Gläubigen wieder auf den Dienst am Anderen aus.

Ich sage nicht, dass die eine Freiheit richtig und die andere falsch ist. Vielmehr denke ich, dass beide Freiheiten, die technische und die religiöse, auch im demokratischen Kontext eine Bedeutung haben.

Einerseits könnte die Idee des Dienstes an der Gottheit das eigentliche Fundament der Demokratie aufrechterhalten: Freiheit als Teilhabe. Demokratisch frei zu sein bedeutet, gleichermaßen Teil der öffentlichen Angelegenheit zu sein. Es bedeutet Dienst – allerdings ein Dienst an der Gemeinschaft, nicht an der Gottheit.

Andererseits geht es bei der Demokratie auch um die Zuteilung von Gütern, wozu auch die Gewährung des Zugangs zu Technologie nach bestimmten Verteilungskriterien gehört. Die Freiheit, die die Technologie ermöglicht, ist Teil der demokratischen Diskussionen über das beste System der Verteilungsgerechtigkeit.

## 2.5. Religion und KI

Braucht Religion also KI? Vielleicht ist „brauchen“ zu viel. Religion kann aus mindestens zwei Gründen mit KI interagieren.

Erstens, um die religionspezifische Sprache der Freiheit, die christliche Verflechtung von Befreiung und Abhängigkeit, wiederzuentdecken, gerade im Licht der Differenz zur Freiheit, die die KI bietet.

Zweitens kann Religion mit Technologie interagieren, um die demokratische Freiheit zu stärken und zu ergänzen. Dies ist kein „Religion gegen KI“ oder „Religion gegen Technologie“-Szenario. Dies würde lediglich dazu führen, dass die Religion isoliert wird – vielleicht noch mehr.

Vielmehr kann die Religion mit der Technologie interagieren, um die nutzerzentrierte Freiheit der Technologie neu auszubalancieren, indem sie Freiheit als Dienstleistung darstellt. Diese Interaktion, vielmehr, diese Resonanz (Rosa 2023), würde zu einem organischen, nicht zu vereinfachten Engagement in der Demokratie und zur Beteiligung am Gemeinwohl beitragen.

## 3. Zusammenfassung

In Fortführung einiger Untersuchungen der *AISE-Gruppe* diskutiert der erste Abschnitt mit dem Titel „Braucht KI Religion?“, wie künstliche Intelligenz (KI) von der Auseinandersetzung mit religiösen Konzepten profitieren kann, um ihre Fähigkeiten zur logischen Argumentation zu verbessern und komplexe abstrakte Konzepte besser zu handhaben, basierend auf theologischen Diskussionen zu Themen wie dem ontologischen Argument für die Existenz Gottes.

Der zweite Abschnitt mit dem Titel „Braucht Religion KI?“ untersucht, wie Religion von der Integration der Technologie, insbesondere künstlicher Intelligenz, profitieren kann, um zeitgenössische Herausforderungen anzugehen und Erzählungen neu zu gestalten. Es werden die potenziellen Risiken von KI-generierten „künstlichen Narrativen“ diskutiert, die unsere Überzeugungssysteme und unsere Freiheit tiefgreifend

beeinflussen können, und die Notwendigkeit eines ausgewogenen Ansatzes für Technologie im Rahmen von Religion und Demokratie betont.

Meine Hoffnung besteht einerseits darin, das Interesse an der Zusammenarbeit in der KI-Forschung und in Themen der Computertheologie zu wecken, und andererseits darin, die Erkenntnisse zu würdigen, die die Arbeit von Hartmut Rosa bietet, sowie die Kraft, die dieses Denken hat, verschiedene Disziplinen zusammenzuführen, um unsere komplexe Welt zu interpretieren.

## Literatur

- Bender, Emily, et al. (2023), On the Dangers of Stochastic Parrots. Can Language Models Be Too Big?, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, 610-623.
- Benzmüller, Christoph/ Woltzenlogel Paleo, Bruno (2014), Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers, in: ECAI 263, 93-98.
- Benzmüller, Christoph (2019), Universal (meta-)logical reasoning. Recent successes, in: Science of Computer Programming 172, Nr. 1, 48-62.
- Benzmüller, Christoph (2023), A Simplified Variant of Gödel's Ontological Argument, in: Vestrucci, Andrea (Hg.), Beyond Babel: Religion and Linguistic Pluralism, Sophia Studies in Cross-cultural Philosophy of Traditions and Cultures, Cham, 271-286.
- Floridi, Luciano (2023), AI as Agency without Intelligence. On ChatGPT, large language models, and other generative models, in: Philosophy & Technology 36, Nr. 1.
- Kahneman, Daniel (2011), Thinking, fast and slow, New York.
- Lattanzi, Nicola/Vestrucci, Andrea (2024), Digital Sapiens. Decidere con l'Intelligenza Artificiale, Rome.
- Maddy, Penelope (1988), Believing the axioms. I., in: The Journal of Symbolic Logic 53, Nr. 2, 481-511.
- Oppenheimer, Paul E./ Zalta, Edward N. (2011), A computationally-discovered simplification of the ontological argument, in: Australasian Journal of Philosophy 89, Nr. 2, 333-349.

- Oppenheimer, Paul E./ Zalta, Edward N. (2021), On Anselm's Ontological Argument in Proslogion II, in: *History of Philosophy & Logical Analysis* 25, Nr. 2, 327-351.
- Rosa, Hartmut (2023), *Demokratie braucht Religion. Über ein eigentümliches Resonanzverhältnis. Basierend auf einem Vortrag beim Würzburger Diözesanempfang 2022*, München.
- Vestrucci, Andrea (2019), *Theology as Freedom. On Martin Luther's De servo arbitrio*, Tübingen.
- Vestrucci, Andrea/ Benzmüller, Christoph (2025), Kurt Gödel and the Logical Existence of God, in: Marcacci, Flavia/ Allen, Paul (Hg.), *Divined Explanations. The Theological and Philosophical Backdrop to the Development of the Sciences (1600-2000)*, Leiden.