



Computer- linguistik

Wenn sich plötzlich alles ändert – Fluch und Segen des KI-Hypes

Annemarie Friedrich

Prof. Dr. Annemarie Friedrich ist Professorin für Computerlinguistik an der Universität Augsburg und Vorsitzende der Deutschen Gesellschaft für Sprachtechnologie und Computerlinguistik. Sie publiziert sowohl zur Modellierung der Struktur und Semantik von Sprache als auch zu sprachtechnologischen Anwendungen.

Als Computerlinguistin beschäftige ich mit der maschinellen Verarbeitung menschlicher Sprache mittels Methoden der künstlichen Intelligenz (KI). Noch vor kurzer Zeit musste ich regelmäßig harte Überzeugungsarbeit leisten, was die Relevanz meiner Arbeit betraf. Ich musste aufzeigen, welchen Benefit Sprachtechnologie bietet.

In den letzten Jahren erfuhr das Fachgebiet durch technologische Durchbrüche einen massiven Zulauf in Forschung und Entwicklung. Die Sprachtechnologie steht plötzlich im Zentrum der Öffentlichkeit und sieht sich mit teilweise überzogenen Erwartungen konfrontiert: Eine KI, die in der Lage ist, wie ein Mensch zu formulieren, muss doch eigentlich genauso intelligent sein?

Heute hat sich meine Argumentationskette umgekehrt: Um zu überzeugen, dass auch weiterhin Forschung in meinem Fachgebiet nötig ist, muss ich zunächst begründen, dass die KI dieses oder jenes Problem *nicht* lösen kann.

Die Verarbeitung natürlicher Sprache mit Computerprogrammen bedient zwei verschiedene Fragestellungen:

- Erstens kann das Erstellen von Anwendungen das Ziel sein, die Sprach- oder Textdaten verarbeiten, um Nutzer:innen Vorteile zu bieten.
- Zweitens kann die Modellierung der Sprache selbst das Ziel sein, um Gesetzmäßigkeiten festzustellen, Sonderfälle zu isolieren, die Entwicklung von Sprache über die Zeit zu beobachten – allgemein gesagt: um *über* Sprache zu lernen.

An der Frage, welche Rolle die Linguistik in der maschinellen Sprachverarbeitung spielt oder spielen sollte, scheiden sich bereits seit mindestens einem Jahrzehnt die Geister. Die jüngsten Durchbrüche scheinen dem Lager rechtzugeben, das für End-to-End-Lösungen plädiert. Aktuelle große Sprachmodelle werden

direkt auf Eingabe- und Ausgabertexten trainiert, ohne dass sie eine explizite linguistische Analyse durchführen.

Was bedeutet nun der KI-Hype für Wissenschaftler:innen der Computerlinguistik, die ihr Fachgebiet auch aus Interesse an der Sprache als Forschungsgegenstand gewählt haben? Haben Arbeiten zur Modellierung von Sprache ihre Relevanz verloren? Lohnt es sich noch, in diesem Bereich zu forschen? Diese Fragen können zur ganz persönlichen Sinnkrise führen. Die starke Dynamik der Disziplin erfordert eine flexible Anpassung der eigenen Forschungsstrategie und ein Überdenken unserer eigenen Vorstellung von Sprache, Wissen und Intelligenz.

Die in der Computerlinguistik entwickelte Methodik hat jedoch keineswegs an Relevanz verloren. Es besteht kein Zweifel, dass aktuelle Ansätze entscheidende Durchbrüche im Bereich der Sprachproduktion erzielt haben und dass für manche Anwendungen auch ohne sprachliche Analyse gute Ergebnisse erzielt werden können.

In der Diskussion wird nur oft vergessen, dass die Auswertung und Bewertung, die für eine Weiterentwicklung essenziell ist, durchaus mit bewährter und oft sprachwissenschaftlich motivierter Methodik erfolgt. Aktuelle Sprachmodelle weisen noch gravierende Mängel in der systematischen Erfassung und Abbildung der Bedeutung von Text auf. Diese zu untersuchen und Lösungen oder zumindest Verbesserungen zu finden, ist nach wie vor Gegenstand unserer Forschung:

Wie gehen die Modelle mit Verneinung oder Unsicherheit in der extrahierten Information um? Ist die Antwort richtig, wenn man Subjektivität und Mehrdeutigkeit beachtet? Außerdem ist eine Zunahme an interdisziplinären Arbeiten zu beobachten – teilautomatisierte sprachliche Analysen haben große Relevanz beispielsweise für die Kommunikations-, Politik- oder Sozialwissenschaften.

Die Sprachwissenschaft bietet eine reiche Tradition an Theorien, um die Struktur und Bedeutung von Sprache zu durchdringen. Ich wünsche mir, dass auch zukünftige Generationen von Computerlinguisten den Sinn hinter einer soliden sprachwissenschaftlichen Ausbildung erkennen. Ich wünsche mir, dass sie in der Lage sind, ihren Forschungsgegenstand

»
Sprache ist viel mehr als Tokens und Zahlen. Sie ist das wunderbare Zusammenspiel von Text, Bedeutung und Kontext.
 «

nicht nur als Tokens und Zahlen, sondern als das wunderbare Zusammenspiel von Text, Bedeutung und Kontext zu begreifen, welches durch die Evolution menschlicher Sprache entstanden ist und das sich auch heute noch täglich weiterentwickelt. Nur so können wir daran arbeiten, dass KI-Modelle in Zukunft ein echtes Verständnis natürlicher Sprache entwickeln.

