

## Existenzielle Risiken von Künstlicher Intelligenz

Die Entwicklung von Künstlicher Intelligenz (KI) wird gegenwärtig umfassend vorangetrieben und feiert insbesondere Erfolge bei Spielen: Nachdem Computer im Schachspiel schon lange gewonnen haben, besiegt 2016 die von Google entwickelte künstliche Intelligenz AlphaGo den mehrfachen Weltmeister Lee Sedol im komplexen Go-Spiel und 2017 gewinnt die Software Libratus gegen vier der weltbesten Pokerspieler. Poker war insofern eine Überraschung, als erfolgreichen SpielerInnen ein hohes Maß an Intuition und psychologischem Geschick zugeschrieben wird. Selbst in Fernseh-Quizsendungen, in denen es darum geht, Wörter und deren Kontext zu analysieren, können Computer Menschen besiegen<sup>43</sup>. Führende Internetkonzerne wie Amazon, Apple, Microsoft, Facebook und Google investieren Milliarden in die KI-Forschung. Es ist nicht nur hinsichtlich der Forschungsgelder abzusehen, dass künstliche Systeme immer mehr lernen, den Systemen stehen auch immer mehr Daten und Experimentierfelder zur Verfügung (mit jedem neuen Dienst, den die Plattformen einführen).

Offen sind drei zentrale Fragen: Die Frage danach, was KI-Systeme in wenigen Jahren können werden, was sie wollen, also welchen Zielhorizonten sie folgen werden, und schließlich die Frage, was dies mit den Menschen und der Gesellschaft machen wird.

Wesentlich für die komplexen und weitgehend unvorhersehbaren Entwicklungspfade ist das „Deep Learning“, das die KI-Forschung stark verändert hat. Hierbei werden Informationen in künstlichen neuronalen Netzen verarbeitet, die analog zum menschlichen Gehirn in hierarchischen Stufen angeordnet sind und damit ein "tiefergehendes Lernen" ermöglichen sollen. Die starke Unvorhersehbarkeit liegt darin, dass diese KI-Systeme aus den gigantischen Datensätzen selbstständig Konzepte und Weltmodelle entwickeln, was oft in Analogie zum Lernen von Kindern dargestellt wird. Mit dem exponentiellen Wachstum von Daten und Rechnerkapazität kann damit eine Komplexität und Geschwindigkeit an Lernfähigkeit in wenigen Jahren erreicht werden, die die individuelle und politische Kontrolle dieser Systeme als unmöglich erscheinen lässt.

Im Feuilleton steht oft die Frage im Vordergrund, was das originär Menschliche der menschlichen Intelligenz ausmacht, wenn KIs komplexe Aufgaben in hoher Geschwindigkeit lösen werden<sup>44</sup>. Gleichzeitig ist die relevante Frage auf politischer Ebene die nach der Steuerung von Prozessen im Kontext der Unberechenbarkeit zukünftiger KI-Fähigkeiten.

Diese Unvorhersehbarkeit zu konkretisieren, bedeutet zugleich, die „ontologische Expansion“ (Tuomi 2012) zur berücksichtigen: Mit KI können künstliche Weltmodelle entstehen, die ganz anders als unsere bekannten

---

<sup>43</sup>[futurezone.at/science/computer-watson-siegte-bei-jeopardy/24.563.160](http://futurezone.at/science/computer-watson-siegte-bei-jeopardy/24.563.160).

<sup>44</sup>[zeit.de/2018/14/kuenstliche-intelligenz-menschen-maschine-verhaeltnis/komplettansicht](http://zeit.de/2018/14/kuenstliche-intelligenz-menschen-maschine-verhaeltnis/komplettansicht).

Weltmodelle sind, aber auch zu anderen Handlungsstrategien von Künstlichen Intelligenzen führen können.

### Zitierte Quellen

Tuomi, I., 2012, Foresight in an unpredictable world, *Technology Analysis & Strategic Management* 24(8), 735-751.

(PSR)