

# Mathematikaufgabe 137

---

[Home](#) | [Startseite](#) | [Impressum](#) | [Kontakt](#) | [Gästebuch](#)

**Aufgabe:** Beweisen Sie, daß es ohne Vertrauen in Künstliche Intelligenz niemals Autonome Systeme geben wird.

**Beweis:** Die einfachste Form einer autonomen Entscheidung ist die sogenannte Wenn-Dann-Beziehung (lat. Subjunktion):

$$\text{if } A \text{ then } B \Leftrightarrow (A \Rightarrow B).$$

Wenn die Aussage  $A$  wahr ist, muß die notwendige Bedingung  $B$  zwingend erfüllt sein.<sup>1</sup> Eine Entscheidung ist nicht autonom, wenn sie der menschlichen Zustimmung  $H$  bedarf, denn dann gibt es im Minimum zwei notwendige Bedingungen  $B$  und  $H$ , die erfüllt sein müssen, d.h.

$$A \Rightarrow B \quad \text{und} \quad A \Rightarrow H,$$

damit gilt:  $A \Rightarrow B \wedge H$ . Wenn nun der Mensch seine Zustimmung verweigert,<sup>2</sup> entscheidet er, daß der Befehl  $B$  nicht ausgeführt wird.

Gibt es nun mindestens zwei hinreichende Bedingungen  $A$  und  $\bar{H}$  für die Aussage  $B$ , d.h. gelten die Subjunktionen

$$A \Rightarrow B \quad \text{und} \quad \bar{H} \Rightarrow B,$$

so reicht es, daß mindestens eine hinreichende Bedingung erfüllt ist, damit gilt:  $A \vee \bar{H} \Rightarrow B$ . Wenn der Mensch die Aufgabe  $\bar{H}$  indes nicht erfüllen kann und außerdem Zweifel daran bestehen, ob  $A$  wahr ist,<sup>3</sup> dann kann der Befehl  $B$  niemals zur Ausführung kommen.

Nehmen wir nun als konkretes Beispiel den Fall eine Drohne, wobei die Aussage  $A$  die Identifizierung eines Ziels erfordert und die Aussage  $B$  das Drücken des Feuerknopfs. Die Aussage  $A \Rightarrow B \wedge H$  läßt sich dann wie folgt in Worte fassen: Der Mensch hat bei erfolgter Identifizierung des Ziels durch Künstliche Intelligenz die Möglichkeit, die Auslösung des Schusses zu verhindern.

Die Aussage  $A \vee \bar{H} \Rightarrow B$  bedeutet andererseits: Es reicht, wenn entweder die Künstliche Intelligenz oder der Mensch das Ziel eindeutig identifizieren, damit der Schuß ausgelöst werden kann.

Wenn es allerdings weder dem Menschen gelingt, das Ziel zweifelsfrei zu identifizieren,<sup>4</sup> noch ausreichend Vertrauen in die Künstliche Intelligenz gesetzt wird, daß ihr das gelungen ist, dann kann der Schuß niemals freigegeben werden.

Gesichtserkennungsalgorithmen können einen Menschen mit hoher Wahrscheinlichkeit identifizieren. Wenn nun der Mensch blind bestätigt, was die Maschine herausgefunden hat, dann

---

<sup>1</sup>  $B$  darf nicht falsch sein, weil aus einer wahren Aussage keine falsche Aussage folgen kann.

<sup>2</sup> In diesem Fall ist  $H$  falsch.

<sup>3</sup> Im Zweifelsfall ist  $A$  falsch.

<sup>4</sup> Was häufig der Fall ist

## Mathematikaufgabe 137

---

braucht man keine Künstliche Intelligenz, denn der Mensch übernimmt sowieso die volle Verantwortung. Überläßt man hingegen die Identifikation den Algorithmen, weil der Mensch diese Fähigkeiten nicht besitzt, dann dürfte er strenggenommen den Feuerknopf nicht betätigen, auch wenn er am längeren Hebel sitzt und letztgültige Entscheidungsinstanz ist. Feuern darf er nämlich nur, wenn er mindestens so gut identifizieren kann wie die Maschine. Aus diesem Dilemma kommt man ohne Zutrauen in die Leistungsfähigkeit der Künstlichen Intelligenz nicht heraus. Im einen Fall wälzt man das Töten lediglich von der Maschine auf den Menschen ab, im anderen muß man es ohnehin blindlings der Maschine überlassen. Dann stellt sich allerdings die Frage, warum man überhaupt eine Technik entwickelt, die nicht besser sein darf als der Mensch,

was zu beweisen war.