

Einstiegs Geschichte für beide Mystery-Gruppen

Jacob Coxon ist 27 Jahre alt, gebürtiger Brite, und hat in den letzten drei Jahren an vorderster Front mitgeholfen, einige der leistungsfähigsten KI-Systeme der Welt zu trainieren – erst bei OpenAI, dann bei Anthropic. Am Abend des 8. September 2026 sitzt er vor seinem Laptop und schreibt einen Thread auf X: „Ich habe heute bei Anthropic gekündigt. Keines der beiden Unternehmen handelt verantwortungsvoll. Sie rasen geradewegs auf selbstverbessernde Superintelligenz zu und spielen mit unserem Leben.“

Über Nacht explodiert sein Post: 90 Millionen Aufrufe in 24 Stunden. Ein Kollege, der bei Anthropic ein Sicherheitsteam leitet, meldet sich zustimmend zu Wort: Er persönlich halte es für zu über 10 Prozent wahrscheinlich, dass KI innerhalb der nächsten zehn Jahre die gesamte Menschheit auslöschen könnte. Journalist:innen weltweit greifen die Geschichte auf. Gleichzeitig finden andere heraus: Coxon kündigte kurz bevor seine Firmenanteile bei Anthropic wertvoll geworden wären – er hatte also finanziell nichts mehr zu verlieren. War es Mut oder eine berechnete Geste?

Leitfrage: „Warnen die Menschen, die künstliche Intelligenz selbst bauen, zu Recht vor ihrem eigenen Werk – und was genau befürchten sie eigentlich?“

Ablauf des Gruppenpuzzles

- Die Klasse bzw. der Kurs wird in Gruppe A „Denker“ (Philosophie/Ethik-Fokus) und Gruppe B „Macher“ (Politik-Fokus) aufgeteilt (bei größeren Kursen mehrere Gruppen pro Typ, je 3–4 Personen).
- Beide Gruppentypen erhalten 4 identische Ankerkarten (S1–S4, gelb) mit dem Coxon-Fall – das gemeinsame Fundament.
- Zusätzlich erhält jede Gruppe 10 Karten zu ihrem Teilthema, plus 3 optionale Vertiefungskarten für leistungstärkere Gruppen oder Leistungskurse.
- Bearbeiten Sie in Ihrer Gruppe zunächst nur die Karten Ihres Teilthemas. Legen Sie das Set der neun Icon-Kärtchen (Kapitel 5) daneben und ordnen Sie jeder Sachkarte das Icon-Kärtchen zu, das inhaltlich am besten passt – diskutieren Sie uneindeutige Fälle in der Gruppe.
- Ordnen Sie die Karten anschließend so an, dass ein Wirkungsgefüge entsteht, kennzeichnen Sie Zusammenhänge zwischen den Karten (verstärkend, abschwächend, unklar) und halten Sie Ihr Ergebnis mitsamt den zugeordneten Icon-Kärtchen auf einem Flipchart-Plakat fest.
- Beantworten Sie anschließend schriftlich Ihre Teilfrage (Gruppe A: „Muss eine KI ‚böse‘ sein, um die Menschheit zu gefährden?“ · Gruppe B: „Wenn KI nur Software ist – warum sorgen sich Sicherheitsleute um Stromnetze, Fabriken und Waffen?“).
- In der Tandem-Phase trifft eine A-Gruppe auf eine B-Gruppe; beide verbinden ihre Plakate über die Ankerkarten und beantworten gemeinsam die große Leitfrage (vgl. Tandem-Arbeitsblatt, Kapitel 10).

Materialhinweis: Farbcodierung

- Gelb = gemeinsame Ankerkarten (in beiden Sets identisch)
- Blau = Karten nur für Gruppe A (Denker)
- Grün = Karten nur für Gruppe B (Macher)
- Grau/neutral = Ethik-Positionskarten, für beide Gruppen gleich

Es empfiehlt sich, die Kartensets zusätzlich auf entsprechend farbigem Papier auszudrucken, um das Sortieren und spätere Zusammenführen zu erleichtern.