1.2 Lehren über KI vs. Lehren mit KI

Kernbotschaften

- ***KI ist ein Navigationsgerät.*** Es rettet die Fahrerin, die die Karte bereits im Kopf hat, und hindert diejenige, die sie nicht hat, am Lernen. Die Wirkung hängt von der Person ab, nicht vom Produkt.
- ***Wir vertreten die Auffassung, dass Lehren über KI die Voraussetzung für Lehren mit KI ist.*** Man muss die Maschine kennen, um die Antwort überprüfen zu können, die sie erzeugt hat.
- **Wissen ist nicht Tun.** Lehrkräfte bewerten ihr KI-Wissen höher als ihre Fähigkeit, damit zu unterrichten, und Fortbildung allein schließt diese Lücke nicht.
- ***Lehrkräfte nutzen KI hinter den Kulissen, nicht im Unterricht.*** In Deutschland nutzt die Hälfte der Lehrkräfte KI mehrmals im Monat oder häufiger; davon lässt weniger als ein Drittel die Schülerinnen und Schüler im Unterricht damit arbeiten.
- ***Pädagogik schlägt Technologie.*** KI zu nutzen, um die eigene Arbeit zu überarbeiten und darauf aufzubauen, schlägt die bloße Nutzung zum Üben oder Nachschlagen; Gruppenarbeit schlägt Alleinarbeit; Feedback schlägt Gamification.
- ***Besseres Feedback bedeutet nicht besseres Lernen.*** In einem Experiment führten KI-Kommentare, die besser bewertet wurden als die einer Lehrkraft, zu Überarbeitungen, die nicht besser waren. Entscheidend ist, ob das Feedback *aufgenommen* wird.
- ***KI steigert den Output, nicht das selbstgesteuerte Lernen.***
- ***Klare Regeln und intensive Nutzung treten gemeinsam auf.*** In einer großen US-Befragung nutzten an Schulen mit einer KI-Richtlinie mehr Lehrkräfte KI – und sparten dadurch insgesamt mehr Zeit, nicht weniger.
- ***Bewahren Sie Ihre eigene Landkarte.*** *Entlastung wird zu Dequalifizierung, wenn KI die Arbeit übernimmt, die Urteilsvermögen aufbaut und erhält.*

> *Wenn Sie sich nur einen Satz von dieser Seite merken:* **KI ist ein Navigationsgerät. Es rettet die Fahrerin, die bereits eine Karte im Kopf hat, und hindert die Fahrerin, die keine hat, still und leise daran,**

> **jemals eine aufzubauen.**

Das Navi-Problem

Um ein Londoner Taxi zu fahren, muss man zunächst eine Wissensprüfung (den „Knowledge") bestehen: 320 festgelegte Routen über die 113 Quadratmeilen im Umkreis von sechs Meilen um Charing Cross, wofür Transport for London bis zu zwei Jahre Vorbereitungszeit einräumt, bevor die Prüfungen überhaupt beginnen.

Auch geprüfte Fahrer:innen nutzen ein Navi, und niemand denkt deswegen schlechter von ihnen. Die Wissensprüfung wurde nicht wertlos, als das Gerät aufkam. Das Navi wurde zu dem, was die Fahrt sicher macht. Eine Fahrerin mit der Stadt im Kopf nutzt das Navi für das, was sie nicht wissen kann: den Unfall auf der Brücke oder den heutigen Verkehr. Und sie bemerkt, wenn es sie durch eine Straße lenkt, die seit April aufgerissen ist.

Setzen Sie nun eine Neuling in dasselbe Taxi mit demselben Navi. Sie erreicht jede Adresse. Eine Zeit lang ist sie von der erfahrenen Fahrerin nicht zu unterscheiden. Doch indem sie den Anweisungen des Navis folgt, baut sie nichts auf. Nach einem Jahr, in dem sie auf Anweisung links abgebogen ist, kann sie die Stadt immer noch nicht ohne das Gerät durchqueren, und sie kann nicht erkennen, wann es sich irrt.

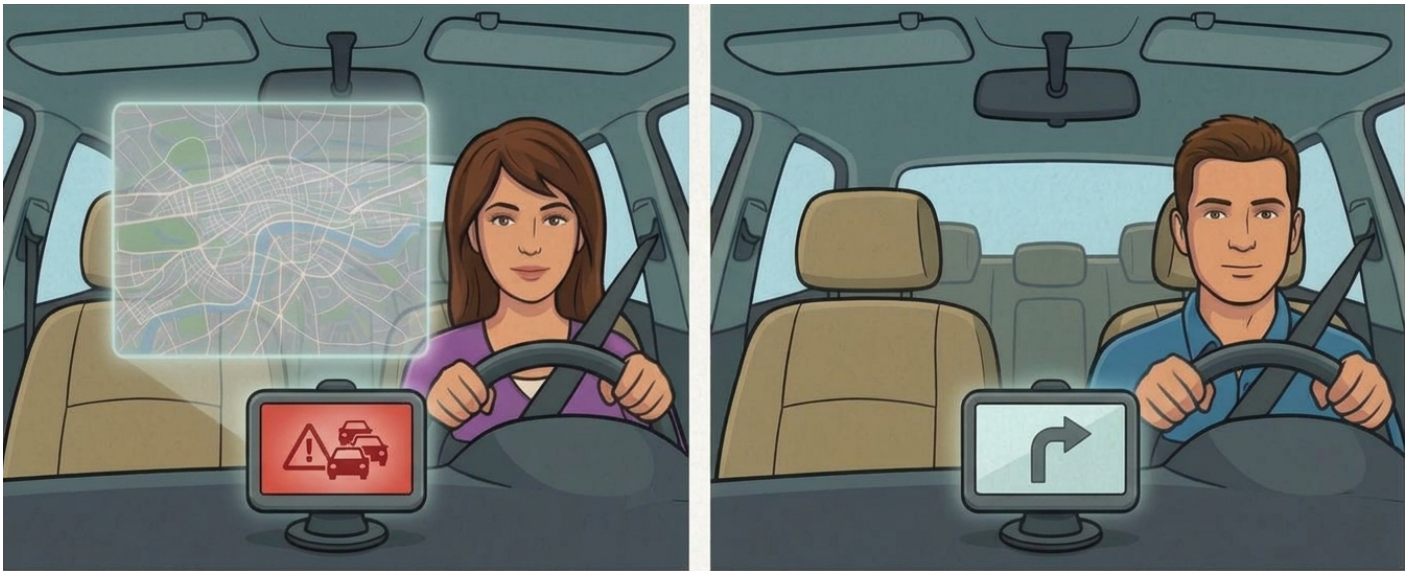


Abb. 1: „Das Ergebnis hängt vom Fahrer ab“ (KI-generiertes Bild, Google Gemini)

Auf der linken Seite beginnt die Fahrt mit einer Karte im Kopf. Das Navi ergänzt, was der Fahrer nicht wissen kann. Das nennt man „nützliches Auslagern“. Rechts hat der Fahrer keine Karte im Kopf. Das Gerät ist seine einzige Orientierung. Das nennt man „schädliches Auslagern“.

Dasselbe Navi, zwei Fahrerinnen, gegensätzliche Ergebnisse. **Der Unterschied lag nie im Gerät.** Es geht um mehr als ein Bild. Studien mit Autofahrern zeigen, dass Menschen, die sich stark auf das Navi verlassen, eine schlechtere mentale Karte ihrer Umgebung entwickeln. Eine Studie über drei Jahre zeigte, dass eine intensivere Nutzung mit einem stärkeren Rückgang des räumlichen Gedächtnisses einherging. Und die Psychologie weiß seit zwei Jahrzehnten, dass dieselbe Unterstützung, die Anfängern hilft, Personen im Weg stehen kann, die das Thema bereits beherrschen. Dasselbe gilt auch umgekehrt. Die Wirkung eines KI-Werkzeugs hängt weit stärker davon ab, was die nutzende Person bereits weiß, als vom Werkzeug selbst.

Forschende nennen dies ein Schema, das organisierte Vorwissen einer Person. Ist es vorhanden, entlastet die Abgabe von Arbeit an eine Maschine. Forschende nennen dies nützliches kognitives Auslagern. Muss Wissen jedoch erst aufgebaut werden, kann derselbe Vorgang genau das verhindern. Das nennt man schädliches kognitives Auslagern.

Eine Studie mit rund tausend weiterführenden Schüler:innen zeigt genau das in einem Mathematikunterricht. Eine Gruppe übte mit einem uneingeschränkten Chatbot, eine andere ohne. Während der Nutzung löste die Chatbot-Gruppe **48 % mehr Übungsaufgaben!** Dann kam eine Prüfung ohne Chatbot. Nun schnitten dieselben Schüler:innen, die zuvor so gut Übungsaufgaben gelöst hatten, **17 % schlechter** ab als die Gruppe, die nie einen KI-Chatbot genutzt hatte. Eine dritte Gruppe nutzte eine KI-Version, die so gebaut war, dass sie Antworten zurückhielt und stattdessen Rückfragen stellte. Diese Gruppe verlor nichts.

Diese dritte Gruppe ist die eigentliche Gestaltungsfrage. Ein Navi, das das Ziel nennt und Ihnen hilft, die Route selbst zu finden, bringt Ihnen die Stadt bei. Ein Navi, das sagt *„In 200 Metern links abbiegen*", tut das nicht, so gut seine Daten auch sein mögen. Forschende nennen diesen Unterschied Leitplanken (Guardrails), und **die Gestaltung entscheidet, ob Sie Ihren Schüler:innen helfen, die Karte der Stadt zu lernen, oder ob sie einfach nur das Ziel finden.**

Bitte denken Sie daran: **Sie haben das Wissen. Ihre Schüler:innen haben es, in dem Fach, das Sie unterrichten, noch nicht.** Diese Asymmetrie ist das ganze Kapitel.

„Im Moment sind die Werkzeuge, zu denen Schülerinnen und Schüler greifen, darauf ausgelegt, Probleme zu lösen, nicht zu lehren – und das ist derzeit die zentrale Schwierigkeit in meiner Arbeit.“

Zwei Stockwerke, nicht zwei Themen

In der Fachliteratur gelten diese beiden Bereiche als getrennte Felder. Man kann sie sich als zwei Stockwerke eines Gebäudes vorstellen, wobei sich die meisten Lehrkräfte im Erdgeschoss befinden und keine Treppe ins erste Stockwerk führt.

In einer Studie mit 90 Lehrkräften an bayerischen beruflichen Schulen, die sich für eine KI-Fortbildung angemeldet hatten, schätzten die Lehrkräfte ihr KI-Wissen im Schnitt mit 3,4 von 5 Punkten ein, ihre Fähigkeit, KI im Unterricht einzusetzen, dagegen deutlich niedriger mit 2,6. Wissen und Handeln hatten sich voneinander gelöst. Eine Studie in fünf europäischen Ländern untersuchte daraufhin, ob Fortbildung diese Lücke schließen kann. 736 Lehrkräfte in Frankreich, Irland, Italien, Luxemburg und Slowenien nahmen teil. Die Lehrkräfte wurden in zwei Gruppen aufgeteilt. Eine Gruppe erhielt während der Studie einen Online-Kurs und Begleitung, die andere erhielt denselben Online-Kurs nach Abschluss der Studie. Die Ergebnisse zeigten, dass das Wissen und die Fähigkeit, die Leistungsfähigkeit eines Werkzeugs einzuschätzen, zunahmen. In den meisten Ländern änderte sich der regelmäßige Einsatz von KI im Unterricht jedoch nicht: Die Lehrkräfte probierten Werkzeuge aus und kehrten dann zu dem zurück, was sie vorher getan hatten.

Das ist die Wissen-Handlungs-Lücke. Kurse erzeugen Wissen, aber die Veränderung der Praxis braucht Begleitung, Ausprobieren und Kolleg:innen zum gemeinsamen Nachdenken. **Transfer muss gestaltet werden, nicht erhofft.**

„Der einzige Nachteil dieses Webinars ist, dass wir es nicht direkt ausprobieren können.“

Was Lehrkräfte tatsächlich tun

Der robusteste Befund über alle Länder hinweg ist auf leise Weise überraschend: **Lehrkräfte nutzen KI intensiv für sich selbst und kaum mit ihren Schülerinnen und Schülern**. In Deutschland nutzen 52 % der Lehrkräfte KI-Werkzeuge mehrmals im Monat oder häufiger. Doch von denjenigen, die KI nutzen, lassen nur 29 % ihre Schülerinnen und Schüler im Unterricht damit arbeiten. Die öffentliche Debatte über „KI im Klassenzimmer“ dreht sich also größtenteils um etwas, das in der Fläche noch gar nicht stattfindet.

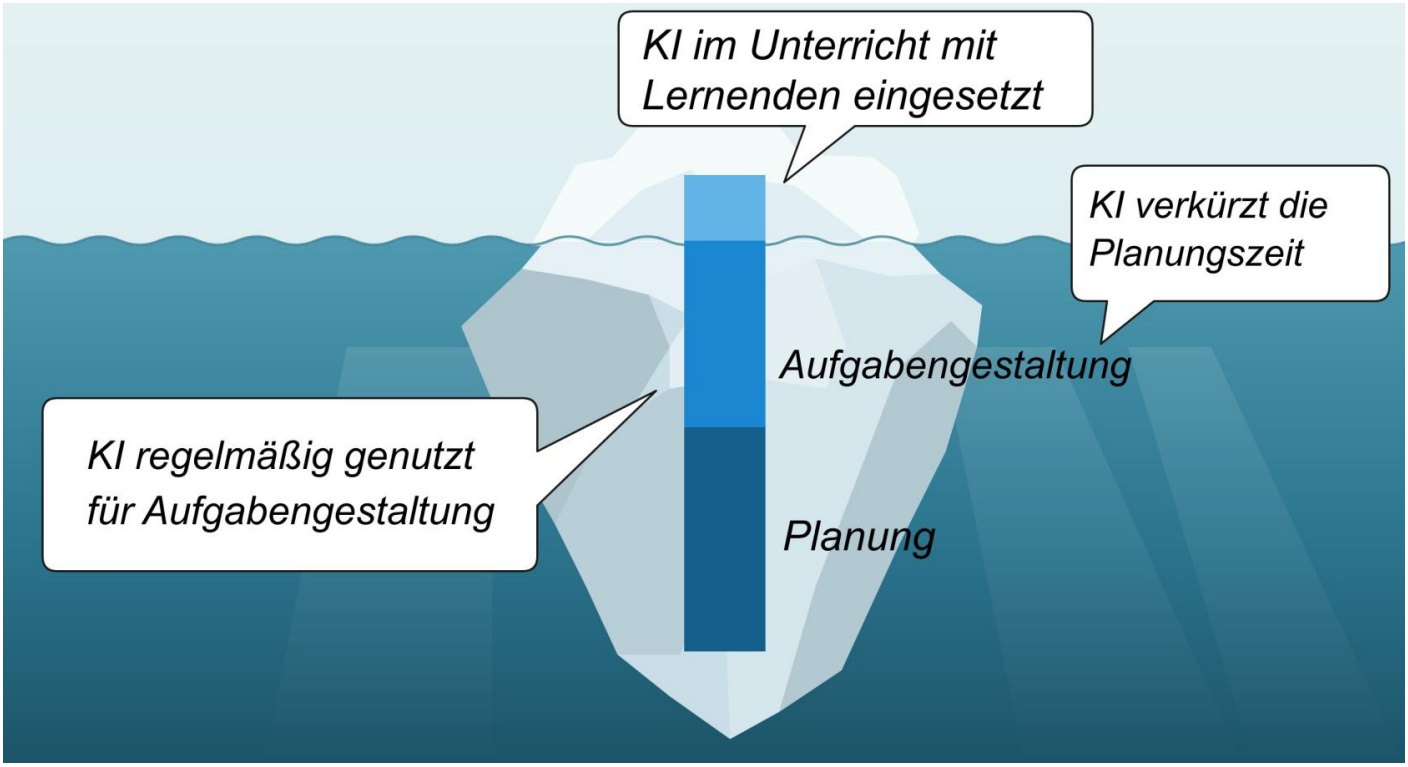


Abb. 2: „Unterschied zwischen der KI-Nutzung zur Vorbereitung und der Nutzung im Unterricht“ (KI-generiertes Bild, Google Gemini)

Was tatsächlich passiert, ist Entlastung von der Arbeitsbelastung, und hier sind die Nachrichten gut. In einer Studie in England erhielten 259 Lehrkräfte an 68 Schulen ChatGPT und einen Leitfaden für die Planung von Naturwissenschaftsunterricht für 11- bis 13-Jährige. Sie wendeten etwa 56 Minuten pro Woche für die Planung auf, gegenüber 81 Minuten in der Vergleichsgruppe. Unabhängige Gutachter, die nicht wussten, aus welcher Gruppe eine Unterrichtsstunde stammte, fanden keine Hinweise auf einen Qualitätsunterschied. In einer großen US-Befragung schätzten Lehrkräfte, die KI mindestens wöchentlich nutzen, dass sie fast sechs Stunden pro Woche einsparen – ihre eigene Schätzung, keine Messung.

Und ein Muster verdient einen Moment Aufmerksamkeit. In derselben US-Befragung **nutzten an Schulen mit einer KI-Richtlinie mehr Lehrkräfte KI – sieben von zehn gegenüber sechs von zehn – und sparten dadurch insgesamt etwa ein Viertel mehr Lehrkräftezeit.** Für die einzelne Lehrkraft, die KI jede Woche nutzt, war der Unterschied gering. Das ist die eigene Einschätzung der Lehrkräfte und keine Messung, und eine Befragung kann nicht zeigen, in welche Richtung der Zusammenhang verläuft: Gut organisierte Schulen machen möglicherweise einfach beides gut. Aber es ist das Gegenteil dessen, was die meisten erwarten – **klare Regeln und intensive Nutzung treten gemeinsam auf, nicht als Gegensätze.** -> 1.4 Konformer Einsatz von KI

„Es ist wichtig, sie als Ergänzung zu nutzen, aber auf keinen Fall ChatGPT die ganze Unterrichtsstunde gestalten zu lassen.“

Warum es nicht auf das Werkzeug ankommt

Wer sich ein Ranking der besten KI-Werkzeuge wünscht, wird von der Forschung zunächst enttäuscht – und beim zweiten Lesen davon befreit. Wenn Forschende viele einzelne Studien zusammenführen und fragen, was Klassenzimmer, in denen KI stark geholfen hat, von solchen unterscheidet, in denen sie kaum geholfen hat, lautet die Antwort nie die Marke der Software. Es ist das, was die Schülerinnen und Schüler damit *tun* sollten. Die folgenden Punkte fassen wichtige Forschungsergebnisse zusammen. Alle Ergebnisse beruhen jedoch auf nur wenigen Studien.

- In Mathematik erzielten Schülerinnen und Schüler, die KI nutzten, um eine Lösung zu überarbeiten oder auf ihrem eigenen Lösungsversuch aufzubauen, etwa zweieinhalbmal so viel Lernzuwachs wie diejenigen, die KI nur zur Unterstützung gewöhnlicher Aufgaben nutzten, etwa zum Üben oder zum Überprüfen von Antworten.
- Schülerinnen und Schüler, die KI gemeinsam in Gruppen nutzten, erzielten etwa 70 % mehr Lernzuwachs als diejenigen, die sie allein nutzten.
- Wo KI eingesetzt wurde, um Feedback zu Schülerarbeiten zu geben, gehörten die Lernzuwächse zu den größten. Der Einsatz von KI zur Gamification des Unterrichts brachte einen Zuwachs, der zu klein war, um als belegt zu gelten.

Pädagogik ist wichtiger als die eingesetzte Technologie. Wo Forschende die Studien danach sortiert haben, was die Schülerinnen und Schüler tatsächlich getan haben, war die Aktivität wichtiger als das Werkzeug. Interessanterweise zeigte sich dasselbe Muster bereits bei der vorherigen Generation digitaler Werkzeuge an weiterführenden Schulen. Die Ergebnisse hängen weitgehend davon ab, ob die Lehrkräfte für deren Einsatz geschult worden waren.

Gutes Feedback ist nicht dasselbe wie wirksames Feedback. In einem Experiment schrieben 70 Studierende einen argumentativen Aufsatz und überarbeiteten ihn eine Woche später. Sie wurden zufällig drei Gruppen zugeteilt: Eine erhielt Feedback von einer erfahrenen Lehrkraft, eine von einem Chatbot mit einfacher Anweisung und eine von demselben Chatbot, der angewiesen wurde, den Aufsatz Schritt für Schritt zu durchdenken. Bewertet nach demselben Raster, lieferte der schrittweise vorgehende Chatbot das **beste Feedback der drei.** Es war besser als das des einfachen Chatbots und besser als das der Lehrkraft. Dann kamen die Überarbeitungen, und **alle drei Gruppen verbesserten sich um etwa denselben Betrag.** Das beste Feedback der Studie brachte keine größere Verbesserung als das am schlechtesten bewertete. **Bessere Kommentare führten nicht zu einem besseren Aufsatz.**

Das verdient einen Moment Aufmerksamkeit, denn es stellt die verbreitete Annahme infrage, besseres Feedback führe zu besserem Lernen. Für sich genommen tut es das nicht. Feedback bewirkt nur dann etwas, wenn es **aufgenommen** wird: gelesen, verstanden, umgesetzt – und so gut verstanden, dass die Schülerin oder der Schüler es beim nächsten Mal auch ohne Hilfe richtig machen könnte. **Die Aufnahme ist der Engpass, nicht die Qualität**.

Mehr ist nicht immer besser. In einer zweiten Studie mit 60 Schüler:innen der Jahrgangsstufe 11 und 240 Textstücken produzierte die KI weit mehr inhaltsbezogene Kommentare als die Lehrkräfte. Es waren die Kommentare der Lehrkräfte, die die Schüler:innen häufiger umsetzten. Konfrontiert mit dreißig präzisen Korrekturen, hört eine Schülerin oder ein Schüler womöglich auf zu diagnostizieren und beginnt stattdessen zu sortieren – oder fügt sich einfach. Wer den Regeln folgt, sagt damit im Grunde, dass die Überarbeitung unnötig war. Es ist wieder das Navi, nur auf einer anderen Ebene. Es gibt einem wirklich detailliertes Feedback, wie „In 200 Metern links abbiegen" – nur für den eigenen Text.

Wer es sagt, spielt eine Rolle. Wir wissen nicht genau, warum Kommentare von Lehrkräften häufiger aufgenommen werden. Das ist Interpretation, keine Messung, aber die wahrscheinlichste Erklärung ist, dass Feedback als etwas zwischen Menschen wahrgenommen wird. Was jemand sagt, der weiß, was man letztes Halbjahr gemacht hat, und sehen wird, was man damit macht, ist etwas anderes als derselbe Satz, auf Anfrage generiert.

Was das für die Praxis bedeutet. Eine zusammenfassende Analyse von Studien zu automatisiertem Schreibfeedback, die noch vor den heutigen Chatbots durchgeführt wurde, kam zu ähnlichen Ergebnissen. Feedback bewirkte mehr, wenn Schülerinnen und Schüler es auf einen neuen Text anwenden mussten, als wenn sie nur den vorliegenden Text überarbeiteten, und mehr, wenn es über mehrere Sitzungen statt nur einmalig eingesetzt wurde. Und fairerweise muss noch etwas gesagt werden: Verglichen mit gar keinem Feedback hilft KI-Feedback durchaus. In einer Studie eines deutschen Teams mit 459 Schülerinnen und Schülern der Sekundarstufe II verbesserte KI-generiertes Feedback ihre Überarbeitungen und ihre Motivation in moderatem Umfang.

Und dieselbe Aufspaltung zeigt sich auch im großen Maßstab. Ein Team führte 228 Studien zusammen: KI verbesserte klar und durchgängig, was Schüler:innen produzieren konnten, während jeder Zuwachs bei der Fähigkeit, das eigene Lernen zu planen, zu überwachen und zu korrigieren, klein genug war, um Zufall sein zu können. Am deutlichsten zeigt sich das beim Programmieren. Schüler:innen mit einer KI-Assistenz waren messbar schneller, und am Ende nicht besser als jene ohne. **Mehr zu schaffen und besser zu werden sind nicht dasselbe, und KI trennt die beiden voneinander.**

Europa als Spiegel

Europaweit reichen die in Befragungen ermittelten Anteile der Lehrkräfte, die KI nutzen, **von etwa neun von zehn in einer tschechischen Stichprobe bis zu etwa einer von sieben in Frankreich.** Die Befragungen unterscheiden sich in Methode und Zeitpunkt: Die tschechische Zahl stammt aus einer kleinen Befragung von 317 Lehrkräften aus dem Jahr 2025 und zählt jede Erfahrung mit KI, während die französische Zahl aus der internationalen Lehrkräftebefragung TALIS 2024 der OECD stammt, in der knapp die Hälfte der tschechischen Lehrkräfte angab, KI im Unterricht zu nutzen.

Frankreich hat eines der größten staatlichen KI-Programme des Kontinents aufgebaut: adaptive Übungssoftware in Mathematik und Französisch, kostenlos für alle Schülerinnen und Schüler im ersten Jahr der Sekundarstufe II. Dennoch nutzen französische Lehrkräfte KI selten, und nur 9 %

von ihnen haben überhaupt eine KI-Fortbildung erhalten, verglichen mit 38 % im OECD-Durchschnitt.

Norwegen hat das Navi-Problem – zumindest als Empfehlung – so in seine Politik überführt, wie es eine Taxifahrschule tun würde: Erst mit der Karte lernen, danach kommt das Gerät hinzu. Im Juni 2026 empfahl Norwegen, dass Schülerinnen und Schüler der Jahrgangsstufen 1 bis 7 in der Regel keinen Zugang zu KI erhalten sollten. In den Jahrgangsstufen 8 bis 10 kann KI schrittweise und behutsam eingeführt werden, sobald die Lehrkräfte die Kompetenz haben, dies anzuleiten. In der Sekundarstufe II sollen die Schülerinnen und Schüler lernen, KI zielgerichtet zu nutzen.

Der blinde Fleck

Ein Risiko von KI wird selten benannt, und auch das Bild führt dorthin: Das Wissen über die Karte der Stadt London **ist nicht von Dauer.** Eine Fahrerin, die fünf Jahre lang den Pfeilen folgt, behält die Stadt nicht im Kopf. Es verblasst, und sie bemerkt die aufgerissene Straße nicht mehr. Ihre eigene Expertise erhalten Sie, indem Sie genau die Dinge selbst tun, die KI am besten übernimmt: die Aufgabe formulieren, das Missverständnis vorausahnen und die Erklärung entwerfen. **Wo KI die Arbeit übernimmt, durch die sich Urteilsvermögen bildet, wird aus Entlastung Dequalifizierung.** In einem deutschen Gremium von E-Learning-Fachleuten erwarteten 65 von 100, dass generative KI Menschen wichtige Fähigkeiten kosten wird – gegenüber 44 im Jahr zuvor. In der Medizin wurde der Effekt bereits gemessen: Nachdem KI-Unterstützung bei der Darmspiegelung zur Routine geworden war, entdeckten Ärzte, die ohne sie arbeiteten, weniger Krebsvorstufen als zuvor.

Nutzen Sie das Navi. Bewahren Sie sich nur die Fähigkeit, die Karte zu lesen.

Revision #17
Created 2026-09-01 09:16:52 UTC by Alexander
Updated 2026-10-30 07:36:13 UTC by Weitzel