

## Zur neuen Bedeutung von Daten in Data Science und künstlicher Intelligenz

Lukas Höper,<sup>1</sup> Susanne Podworny,<sup>2</sup> Sven Hüsing,<sup>1</sup> Carsten Schulte,<sup>1</sup> Yannik Fleischer,<sup>2</sup>  
Rolf Biehler,<sup>2</sup> Daniel Frischemeier,<sup>3</sup> Hülya Malatyali<sup>1</sup>

In vielen Bereichen des alltäglichen Lebens entstehen große Mengen an Daten. Diese werden teilweise automatisch gesammelt wie etwa Standortdaten durch entsprechende Sensoren bei der GPS-Nutzung oder Daten, die beim Nutzen von digitalen Artefakten zum Teil aktiv und bewusst hinterlassen werden. Um diese Daten nutzbar zu machen, werden datengetriebene Algorithmen etwa aus dem Bereich der künstlichen Intelligenz oder speziell aus dem maschinellen Lernen eingesetzt. Die Systeme, bei denen diese Technologien zum Sammeln und Verarbeiten von Daten eingesetzt werden, prägen unser Leben und sind dabei noch schwer zu verstehen. Seit einiger Zeit werden verstärkt Unterrichtsansätze und -materialien entwickelt, in denen künstliche Intelligenz erklärt werden soll. Zum Verstehen und Beurteilen der entsprechenden Methoden und Systeme reicht es aber nicht aus, (nur) die Algorithmen zu erläutern. Zum einen ist es notwendig, deutlich herauszustellen, dass es sich bei Daten stets um Modelle der Realität handelt. Zum anderen müssen, um die Relevanz der Daten und Validität der Ergebnisse beurteilen zu können, auch Aspekte wie Bias in Daten und Fairness von datengetriebenen Systemen thematisiert werden. Bei vielen Ansätzen findet ein wichtiger Bereich wenig Beachtung: Die Rolle der Daten als Grundlage für Verfahren der künstlichen Intelligenz, insbesondere des maschinellen Lernens.

Beim Problemlösen mit klassischen Algorithmen stehen Kompetenzen zum Verstehen und Analysieren des Problems, zum Finden einer Lösung und schließlich zum Formalisieren dieser Lösung im Mittelpunkt, sodass als Produkt eines solchen Problemlöseprozesses in der Regel ein implementierter Algorithmus steht. Ein solcher Algorithmus kann etwa auf verschiedene Eingangsdaten angewendet werden und liefert ein zuvor bestimmtes Ergebnis, wobei die Korrektheit dieser Ergebnisse erklärt und nachgewiesen werden kann.

In diesem moderierten Austausch sprechen wir über datengetriebenes Problemlösen im Vergleich zum Problemlösen mit klassischen Algorithmen, um darauf aufbauend den Datenbegriff in diesem Kontext zu greifen und resultierende Folgerungen für die Bedeutung von Daten festzuhalten, die im Informatikunterricht im Kontext von Data Science und künstlicher Intelligenz berücksichtigt werden sollten.

<sup>1</sup> Universität Paderborn, Didaktik der Informatik, Fürstenallee 11, D-33102 Paderborn, {lukas.hoeper,sven.huesing,carsten.schulte,hulya.malatyali}@upb.de

<sup>2</sup> Universität Paderborn, Didaktik der Mathematik, Warburger Str. 100, D-33098 Paderborn, {podworny,yanflei,biehler}@math.upb.de

<sup>3</sup> Universität Münster, Didaktik der Mathematik, Fliegerstr. 21, D-48149 Münster, dfrische@uni-muenster.de