

Warum sollen wir unseren Daten trauen? Soziale Erkenntnistheorie und die 'rechnenden Geisteswissenschaften'

Wenn wir über die Beziehung zwischen Daten und Erkenntnis nachdenken, übersehen wir leicht, dass Daten selbst implizite Erkenntnisansprüche enthalten können. Daten sind nicht immer gegeben, sie können auch das Ergebnis menschlichen Handelns sein: ein Bibliothekar gibt Metadaten ein. Oder sie beruhen auf den Rechnungen eines Computers, bspw. in 'topic models', die aus einem elektronisch verfügbaren Corpus abgeleitet werden. Der Glaube an die Korrektheit dieser Daten ist eine Sache des Vertrauens. Wenn wir in eine Bibliothek reisen, um ein Rarum einzusehen, so vertrauen wir auf die professionellen Fähigkeiten des Bibliothekars, der für die Erstellung der Metadaten im Katalog verantwortlich war. Wenn wir einen Computer nutzen, um 'topic models' zu berechnen, vertrauen wir der Korrektheit der Berechnungen und benutzen die Resultate im Glauben, dass sie das entsprechende Corpus wahrheitsgetreu (innerhalb der gegebenen Parameter) repräsentieren.

In den letzten Jahrzehnten ist epistemisches Vertrauen und damit zusammenhängend das Wissen aus dem Zeugnis anderer zu einem wichtigen Thema der Erkenntnistheorie, insbesondere der sogenannten "sozialen Erkenntnistheorie" geworden. Don Fallis hat für die Informationswissenschaft diese Debatten in nützlichen Kriterien zusammengefasst (Fallis 2004), von denen ich drei in meiner Präsentation heranziehen werde: wenn wir uns fragen, wieviel Vertrauen wir einer möglichen Wissensquelle entgegenzubringen bereit sind, sollen wir (1) die Autorität unseres Informanten, (2) die Anzahl voneinander unabhängiger Wissensquellen sowie (3) den Inhalt der Information, ihre Plausibilität und den Grad ihrer Bestätigung in Rechnung stellen. Bei der Beurteilung des Computers als eines Informanten ist (1) von überragender Wichtigkeit: wie können wir rechtfertigen, dass wir den Computer als epistemische Autorität für die Erarbeitung geisteswissenschaftlicher Forschung akzeptieren? Der Versuch der Beantwortung dieser Frage wird uns jedoch darauf führen, dass auch (2) von einigem Belang ist: wir können (und sollten wohl auch) unser Vertrauen in die epistemische Autorität des Computers dadurch steigern, dass wir mehr als einen benutzen: die Diskussion des 'Wissens aus dem Zeugnis eines Computers' führt also zu (3), zum Problem einer effektiven Methodologie für die Plausibilitätsprüfung computererzeugter Daten.

(1) Die epistemische Autorität von Computern steht nicht nur in den digitalen Geisteswissenschaften in Frage. Durch einen Computer erzeugte Beweise haben in der Philosophie der Mathematik zu einer intensiven Debatte geführt. In einem mittlerweile klassischen Aufsatz (Burge 1998) argumentiert Tyler Burge für die Unterscheidung zwischen einer widerlegbaren Berechtigung zu epistemischem Vertrauen ('entitlement') und der vollständigen Rechtfertigung von Wissensansprüchen, die auf computergenerierten Daten beruhen. Die widerlegbare Berechtigung zu solchem Vertrauen beruht darauf, dass Wissen aus dem Zeugnis anderer auch in anderen Bereichen der Mathematik vorausgesetzt wird: die Anwendung des Satzes des Pythagoras hat bei der Lösung eines Problems auch dann ihre Berechtigung, wenn wir nicht zuvor seinen Beweis nochmals nachvollzogen haben (Burge 1998, 7). In ähnlicher Weise können Kinder in grundlegenden Wahrnehmungsüberzeugungen gerechtfertigt sein, auch wenn sie zu den Tatsachen keinen kognitiven Zugang haben, die dieses Gerechtfertigtsein ermöglichen, weil sich diese Tatsachen erst in der philosophischen Reflexion erschließen (Burge 1998, 3). Eine solche 'widerlegbare Berechtigung' in Burges Sinn kann jedoch durch vernünftigen Zweifel widerlegt werden, Zweifel, die jedoch dadurch ausgeräumt werden können, dass wir die Quelle des Wissensanspruchs ausklammern (bzw.,

in Burges' Terminologie, die Perspektive der dritten Person einnehmen). In dieser Perspektive wird der Gebrauch des Computers in der Mathematik zu einer Erweiterung unserer rationalen Fähigkeiten (Burge 1998, 31).

(2) Aber selbst wenn wir diesen Überlegungen zustimmen würden, können sie nur eine allgemeine Bereitschaft zum Vertrauen in die epistemische Autorität eines Computers begründen. Sie geben jedoch keine Antwort auf die Frage, ob dieses Vertrauen bedingungslos sein muss. Burge konzediert, dass Programmiersprachen und die in ihnen geschriebenen Programme selbst Ausdruck mathematischer Wahrheiten seien. Sie seien aber so zuverlässig wie vom Menschen erarbeitete mathematische Wahrheiten (Burge 1998, 8). Hardwarefehler können mit Schreibfehlern bei der Abfassung eines Beweises gleichgesetzt werden (Burge 1998, 8f). Weil aber Computer von rationalen Wesen gebaut und programmiert werden, sind es insgesamt deren Rationalität und die Unveränderlichkeit der im Computer ausgenutzten Naturgesetze, die aus der Sicht von Burge die letztendliche Rechtfertigung unseres Vertrauens in die Ergebnisse begründen, die von solchen Maschinen geliefert werden. Aus dem Blickwinkel des Ingenieurs erscheint diese Einschätzung als sehr optimistisch. Arkoudas und Bringsjord (2007) unterscheiden sechs unterschiedliche Ebenen, auf denen bei der Generierung von Computerergebnissen Fehler unterlaufen können. Jenseits der allgemeinen Berechtigung, der epistemischen Autorität von Computern zu vertrauen ist die konkrete Rechtfertigung unserer Wissensansprüche und damit zusammenhängend das Vertrauen, das wir dem Computer entgegenbringen sollten, also auch ein technisches Problem. Ähnlich wie in Experimenten der Naturwissenschaften sind auch die digital humanities gehalten, der Überprüfung der Replizierbarkeit von Ergebnissen erhöhtes Augenmerk zu schenken.

(3) Deswegen werde ich abschließend einen konkreten Anwendungsfall erörtern: die Verwendung von algebraischen Programmen zu heuristischen Zwecken in der Mathematik und die daraus resultierenden Probleme, wie sie in Durán et al. 2014 geschildert werden. Wollen wir computergeneriertes Wissen als gerechtfertigt auszeichnen, sind konkrete methodische Richtlinien zu befolgen: dies betrifft bspw. den Gebrauch mehrerer Implementationen, den Einsatz von Open Source Software, um im Zweifel die Fehlerfindung zu erleichtern oder die Publikation von Code, der in einem Projekt eingesetzt wurde, um die Replizierbarkeit von Ergebnissen zu erleichtern. Nur dann sind wir nicht nur widerlegbar berechtigt, sondern gerechtfertigt, dem in Daten implizierten Wissen zu vertrauen und dürfen es zur Erarbeitung weiterreichender interpretatorischer Überlegungen und Erkenntnisse nutzen.

Literatur

Konstantine Arkoudas, Selmer Bringsjord, "Computers, Justification, and Mathematical Knowledge", in: *Minds & Machines* (17) 2007, 185–202

Tyler Burge, "Computer Proof, Apriori Knowledge, and Other Minds: The Sixth Philosophical Perspectives Lecture", in: *Noûs* (32) 1998, supplement, 1-37

Antonio J. Durán, Mario Pérez, and Juan L. Varona, "The Misfortunes of a Trio of Mathematicians Using Computer Algebra Systems: Can We Trust in Them?", in: *Notices of the AMS*, (61) 2014, 1249-1252

Don Fallis, "On Verifying the Accuracy of Information: Philosophical Perspectives", in: *Library Trends* (52) 2004, 463-487