

Enabling reproducibility in data science

Die Session „Enabling reproducibility in data science“ wurde vom [German Reproducibility Network](#) (GRN) durchgeführt und thematisierte die Bedeutung von Reproduzierbarkeit für eine offene und robuste Wissenschaftslandschaft aus verschiedenen Blickwinkeln.

Antonia C. Schrader, Referentin im [Helmholtz Open Science Office](#), führte zunächst in die Thematik ein und erläuterte, dass der Begriff „Reproduzierbarkeit“ je nach Fachgebiet unterschiedlich definiert wird. Grundsätzlich kann jedoch gesagt werden, dass Forschung reproduzierbar ist, wenn es möglich ist, (unabhängig) dieselben Ergebnisse mit denselben Daten und/oder demselben Code/Analyseverfahren zu erzielen, die vom ursprünglichen Forscher oder Forscherteam verwendet wurden. Antonia Schrader betonte in ihrem Vortrag, dass Reproduzierbarkeit die Zusammenarbeit und Transparenz in der Wissenschaft fördert und die Wiederverwendbarkeit wissenschaftlicher Ergebnisse unterstützt. Es wurde deutlich, dass dies in engem Zusammenhang mit den Bemühungen um Open Science und den damit verbundenen kulturellen Wandel von Wissenschaft und der -kommunikation steht.

Anschließend ging Ulf Tölch, Leiter der Gruppe „[Systemische Perspektiven in der translationalen Biomedizin](#)“ am Berlin Institute of Health (BIH) an der Charité, auf die Bedeutung von nachvollziehbaren Datensätzen in der Forschung ein. Hierzu stellte er ein Statement des GRN vor, indem die Mitglieder des Netzwerkes fünf Maßnahmen vorschlugen, die erforderlich sind, um die Zugänglichkeit und Reproduzierbarkeit der von der Wissenschaft erzeugten Forschungsdaten zu verbessern. Das Statement ist hier veröffentlicht: <https://doi.org/10.5281/zenodo.6840403>

In der nachfolgenden Präsentation von Peter Steinbach, Berater und Teamleiter bei [Helmholtz AI](#), wurde Reproduzierbarkeit im Zusammenhang mit maschinellem Lernen und künstlicher Intelligenz betrachtet. Maschinelles Lernen ist in vielen wissenschaftlichen Bereichen allgegenwärtig. Praktiker:innen haben jedoch Schwierigkeiten, jede Neuerung auf dem Markt des maschinellen Lernens auf ihre Daten anzuwenden und dabei vergleichbare Ergebnisse zu erzielen. In seinem Vortrag berichtete Peter Steinbach von seinen Beobachtungen zur Reproduzierbarkeit von Ergebnissen des maschinellen Lernens und zeigte auf, wie Forschende die damit verbundenen Herausforderungen bewältigen.

Um Forschung reproduzierbar zu machen, ist die Art und Weise wissenschaftlicher Kommunikation ausschlaggebend; Sie muss den Austausch von Informationen über die Ergebnisse und ihre Herkunft zu ermöglichen. Im letzten Vortrag dieser Session skizzierte Guido Scherp, Leiter der [Abteilung Open-Science-Transfer](#) am Leibniz-Informationszentrum Wirtschaft, die Funktionen, Merkmale und Rollen eines idealen Forschungskommunikationssystems, das reproduzierbare Forschung unterstützt. Der Vortrag spiegelte den Stand eines entsprechenden Diskussionspapiers wider, an dem derzeit mehrere internationale Reproduzierbarkeitsnetzwerke unter der Leitung des GRN beteiligt sind.

Die Session wurde in englischer Sprache durchgeführt und die Folien unter einer CC BY 4.0 International Lizenz auf Zenodo veröffentlicht.