

de.NBI im Kontext von Methoden der Künstlichen Intelligenz

verabschiedet 21.01.2020

Zusammenfassung: *Das Deutsche Netzwerk für Bioinformatik Infrastruktur (de.NBI) unterstützt die KI-Offensive der Bundesregierung. Die Sammlung, Bewertung und der geschickte Einsatz von Trainingsdaten ist der Schlüssel für die Entwicklung neuer Methoden der künstlichen Intelligenz und die Übertragung in neue Anwendungsbereiche wie die Biomedizin. de.NBI ist ideal positioniert, in diesen Bereichen Beiträge zu leisten.*

Nach Jahrzehnten relativer Stille um das Thema KI haben Durchbrüche im Deep Learning den Verfahren der Künstlichen Intelligenz zurück in die öffentliche Sichtbarkeit verholten. Weniger in der Wahrnehmung ist, dass die spektakulären Fortschritte in der KI nur durch große Mengen an verfügbaren Daten zum Lernen ermöglicht werden. Gerade im internationalen Vergleich fällt auf, dass erfolgreiche Anbieter KI-basierter Dienste gezielt große Datenmengen sammeln, um dadurch kompetitive Vorteile zu gewinnen. Die vielen positiven Errungenschaften und neuen Produkte und Dienstleistungen müssen andererseits gegen den Datenschutz abgewogen werden. Wann ist das Sammeln von Daten ethisch angezeigt, wann ist es bedenklich? Wie können Daten für die Allgemeinheit nutzbar werden, ohne ihren Eigentümern zu schaden?

Wie sensibel Daten sind, ist auch in verwandten Forschungsprojekten und Infrastrukturen ein wesentliches Thema, wie zum Beispiel in der BMBF-geförderten Medizin-informatikinitiative (MII). Die informationelle Selbstbestimmung von Bürgern und Patienten muss mit dem Nutzen für die Allgemeinheit sorgfältig ausbalanciert werden, und ein verlässlicher und transparenter rechtlich-ethischer Rahmen für die Datennutzung muss geschaffen werden. Ebenso muss die Wahl der technischen Infrastruktur (Stichwort: Cloud) im Lichte dieses rechtlich-ethischen Rahmens und der nötigen Sicherheitsanforderungen betrachtet werden.

In Gesundheitsforschung und Biotechnologie müssen die notwendigen große Datensätzen zunächst in Experimenten und Studien gesammelt werden. Sie werden dann mit bioinformatischen Methoden und Algorithmen aufbereitet, um daraus KI-Modelle zu trainieren. Weitere Daten werden von Experten „per Hand“ aus wissenschaftlichen Publikationen extrahiert. Sowohl für die Algorithmen als auch für die Datenpflege benötigt es besonderer Expertise, um erfolgreich zu sein.

Deep Learning und KI brauchen große Mengen an Daten, Bioinformatik-Werkzeuge, und Experten der Anwendungsdomäne, um Durchbrüche in den Lebenswissenschaften erzielen zu können.

Diese Expertise bietet de.NBI. In einer kürzlich unternommenen Erhebung unter den Zentren innerhalb von de.NBI haben fast alle Zentren entweder bereits laufende Projekte genannt oder auf konkrete Planungen verwiesen. Aus de.NBI heraus gab es vielfältige Projektvorschläge innerhalb der neuen KI-zentrierten Fördermaßnahme *Computational Life Sciences* des BMBF. Die verwendeten Methoden sind vielfältig. Es kommen sowohl klassische Methoden wie Hidden-Markov-Modelle und Support-Vektor-Maschinen, aber auch neuere Deep-Learning-Ansätze zum Einsatz.

Die Anwendungsgebiete sind ebenso breit gefächert. Deep Learning kann Bilder aus der Mikroskopie (z.B. Schnitte durch Krebsgewebe) vergleichbar gut wie Experten klassifizieren, aus molekularen Daten kann KI viel früher als bisher das Vorliegen einer Blutvergiftung erkennen, die Erkennung seltener Erkrankungen aus Genomdaten ist schneller möglich, oder die mikrobielle Zusammensetzung des Stuhlgangs kann zur Früherkennung von Darmkrebs benutzt werden.

Die Währung des KI-Zeitalters sind qualitativ hochwertige Daten. Forschungsinfrastrukturen wie de.NBI stellen diese in einer transparenten, demokratisch kontrollierten und direkt nutzbaren Form zur Verfügung (FAIR-Prinzipien). An den Standorten werden viele Petabytes solcher Daten vorgehalten. Darüber hinaus entwickelt de.NBI die Werkzeuge und Infrastrukturen (z.B. die de.NBI-Cloud), um KI-Anwendungen zu den Daten zu bringen.

Wir benötigen eine Kombination von KI-Expertise und Anwendungsexpertise, um KI einsetzbar zu machen. Alle der oben genannten Beispiele benötigen eine Kombination beider Expertisen. Wir halten es für wichtig und notwendig, bei zukünftigen Förderprogrammen sowohl Grundlagenforschung zur KI zu unterstützen, wie auch die **praktische Anwendung und notwendige Datenquellen** nicht aus dem Auge zu verlieren. Aus unserer Sicht werden hier Fortschritte nur durch **interdisziplinäre Zusammenarbeit** erzielt. Viele der geplanten Vorhaben in KI und den Lebenswissenschaften können daher von der Zusammenarbeit mit de.NBI und seinen europäischen Partnern (ELIXIR) profitieren - und tun es bereits in zahlreichen Projekten. Beispiele dafür sind die *Million European Genomes Alliance* (MEGA+), das *European Lab for Learning and Intelligent Systems* (ELLIS) oder die *Kompetenzzentren für Künstliche Intelligenz* des BMBF.

KI wird die Forschung in den Lebenswissenschaften transformieren. Dies wird durch spezialisierte Infrastrukturen wie de.NBI begünstigt. Gleichzeitig vereinfacht KI auch das Bereitstellen klassischer Dienste wie auch und insbesondere die Kuratierung von wissenschaftlichen Daten.

de.NBI kann als eine Infrastruktur, bei der Daten aus vielen Quellen verarbeitet werden und fachspezifische Expertise versammelt ist, entscheidende Beiträge leisten, um KI in den Lebenswissenschaften in Deutschland auf international kompetitivem Niveau zu ermöglichen.