

## Big Data Analytics

Erforschung von Methoden zur Auswertung großer Datenmengen unter strenger Wahrung von Privacy- und DSGVO-Anforderungen

<b>Programm / Ausschreibung</b>	COIN, Aufbau, COIN Aufbau 7. Ausschreibung	<b>Status</b>	abgeschlossen
<b>Projektstart</b>	01.08.2018	<b>Projektende</b>	31.01.2022
<b>Zeitraum</b>	2018 - 2022	<b>Projektlaufzeit</b>	42 Monate
<b>Keywords</b>	Big Data; Privacy; Datenschutz; Machine Learning; DSGVO		

### Projektbeschreibung

Durch die fortschreitende Digitalisierung fast aller Arbeits- und Lebensbereiche und den damit verbundenen neuen Möglichkeiten in Bezug auf Produkte und Dienstleistungen hat auch das Thema Privacy an Bedeutung gewonnen. Trends wie personalisierte Dienstleistungen und Produkte, personalisierte Medizin, sowie Dienste, die sich durch den Verkauf von personalisierter Werbung, oder gar der Kundendaten selbst finanzieren, sind dabei ins Rampenlicht gerückt. Zusätzlich wurde in den letzten Jahren der Datenschutz gestärkt, am prominentesten durch die europäische General Data Protection Regulation (GDPR), in Österreich umgesetzt durch die DSGVO, die ab Mai 2018 in Kraft tritt.

Den Datenschutzbemühungen stehen eine Vielzahl von Interessen in Forschung und Wirtschaft gegenüber, die auf der Bereitstellung von personenbezogenen Daten beruhen. Die Qualität dieser Daten ist dabei oft wesentlich. Wie bereits akademisch nachvollziehbar demonstriert wurde, verzerren Anonymisierungsverfahren im Allgemeinen sehr stark und wirken sich nachteilig auf die Qualität aus. Die bisher typischerweise als Ersatz genutzte Pseudonymisierung kann nach der DSGVO hingegen nicht mehr als Datenschutzmaßnahme eingesetzt werden.

Im Rahmen dieses Projekts werden daher Methoden erforscht, diese negative Effekte auf die Ergebnisse von Big-Data-Analysen abschätzbar zu machen und einzudämmen. Dabei sind unterschiedliche Rahmenbedingungen zu beachten, die jeweils andere Methoden benötigen, je nachdem ob es sich bei den Analysen um Trendanalysen, oder exakte Auswertungen auf Kohorten- oder Individualbasis handelt.

Ein wichtiger Aspekt der DSGVO ist die informationelle Selbstbestimmung: Dazu gehört das Recht der nachträglichen Rücknahme der Zustimmung, sowie das Recht auf Transparenz und letztlich das Recht auf Datenlöschung. Datensubjekte haben das Recht auf Auskunft, welche ihrer Daten wofür verwendet werden. Prozesse müssen dies berücksichtigen – eine komplexe Herausforderung in Hinblick auf intelligente Algorithmen. Es werden daher Methoden entwickelt um Transparenz zu gewährleisten, ohne dadurch neue Gefahren für den Datenschutz zu generieren, sowie Methoden zur Löschung von Daten aus komplexen datenverarbeitenden Systemen. Hierzu werden die Effekte von Datenlöschung auf die Ergebnisse von Machine-Learning-Algorithmen quantifiziert.

Aus akademischer Sicht sind diese Forschungsergebnisse nicht nur im informatischen Bereich wichtig, sondern auch für Analysen in anderen Gebieten, bspw. im medizinischen Bereich. In kommerzieller Hinsicht können die Resultate den Standortnachteil heimischer Firmen gegenüber internationalen Konkurrenten, speziell aus China und den USA, mindern. Zusätzlich fließen die Ergebnisse in den neuen Studiengang „Data Science und Business Analytics“ ein, der im Wintersemester 2018 an der FH St. Pölten starten wird. Dieses Projektvorhaben wird im Sinne der forschungs- und praxisgeleiteten Lehre wichtige Beiträge zur Qualität dieses Studiengangs leisten.

### **Projektpartner**

- Hochschule für Angewandte Wissenschaften St. Pölten Forschungs GmbH