

Kurzfassung

Ohne qualitativ hochwertige Daten können Unternehmen nichts Wichtiges tun. (Redman, 2020) Diese zugespitzte Aussage wird unter Berücksichtigung der Renaissance der Künstliche Intelligenz (KI) noch bedeutender: In Kombination mit Maschinelles Lernen (ML) und Analytik wird eine völlig neue Phase der Wertschöpfung eingeleitet, in der der strategische Umfang mit Daten Gewinner und Verlierer im Wettrennen um die Marktführerschaft trennt. (Kagermann u. a., 2018, S. 7) Daher ist die systematische Entwicklung von für KI-Anwendungen geeigneten Datenlagen von essentieller Bedeutung.

Durch die zunehmende Vernetzung von Produkten und deren Ausstattung mit Sensoren werden immer mehr Daten verfügbar, die potentiell auch von KI analysiert werden können. Dies geschieht insbesondere im Bereich Smart Services.

Für die Entwicklung und den erfolgreichen Betrieb von Smart Services als Bestandteil von Smart Product Service Systemen (SPSS) ist eine qualitativ und quantitativ hinreichende Menge an Daten erforderlich. Dies gilt insbesondere dann, wenn datengetriebene Software-Komponenten statistische Methoden aus dem Bereich der KI einsetzen: Die Qualität der Trainingsdaten bestimmt direkt die Qualität des trainierten KI-Modells. Allerdings kann die Qualität der Trainingsdaten mit bestehenden Methoden erst in späten Produktentwicklungsphasen beurteilt werden, da der Trainingserfolg erst bekannt ist, wenn das KI-Modell trainiert wurde. Daher beschäftigt sich die vorliegende Forschungsarbeit mit folgender Leitfrage: *Wie kann die Verfügbarkeit qualitativ hinreichender Daten mit Relevanz für den Smart Service praxistauglich gesteigert werden, um die Qualität der datengetriebenen Softwarebestandteile eines Smart Product Service System (SPSS) zu steigern?*

Die Beantwortung der Leitfrage folgt der in L. T. Blessing und Chakrabarti (2009) vorgestellten *Design Research Methodology (DRM)*. Eine umfassende deskriptive Studie I resultiert in einem Referenzmodell mit insgesamt 93 Einfluss-, neun Schlüssel- und neun Erfolgsfaktoren. Auf Basis des Referenzmodells werden ein Wirkmodell und eine Anforderungsliste abgeleitet, welche im Rahmen der präskriptiven Studie zur Erarbeitung einer Funktionsarchitektur genutzt werden.

Das auf Basis von Funktionsarchitektur und Anforderungsliste resultierende Lösungskonzept wird Semantische Datenspezifikation für KI-basierte Smart Services (SemDaServ) genannt und adressiert auf Basis des Engineering Operating System (EOS) nach Lünemann u. a. (2017, S. 319) die Dimensionen Prozesse und Organisation, virtuelle und physische Artefakte sowie Werkzeuge und IT-Systeme, um die Aktivität der semantischen Datenspezifikation für KI-basierte Smart Services zu unterstützen. SemDaServ besteht aus drei übergeordneten Prozessschritten (Wissensbedarf klären, Informationsbedarf klären, Datenbedarf spezifizieren), sechs Artefakten (Datenkategorienmodell, Informationsmodell, Datenmodell, Wissensmatrix, Datenentscheidungsmatrix, Datenspezifikationscheckliste) sowie der Beschreibung zum Einsatz von IT-Werkzeugen zur Erstellung der Artefakte (u.a. Modellierungswerkzeuge, Data Science Werkzeuge, Enterprise IT). Der SemDaServ-Ansatz wird prototypisch anhand eines Smart Home Office Lightning Service realisiert.

Basierend auf der prototypischen Realisierung erfolgt die deskriptive Studie II zur Evaluation des SemDaServ-Ansatz. Die qualitative Evaluation auf Basis von zwei Workshops, acht Experteninterviews und zwei Fallstudien ergibt, dass der SemDaServ-Ansatz ein nützlicher und in der Praxis anwendbarer Ansatz zur Steigerung der Verfügbarkeit qualitativ hinreichender Daten mit Relevanz für KI-basierte Smart Services ist. Weitere Forschung (insbesondere A/B-Vergleiche) ist erforderlich, um Effektivität und Effizienz des SemDaServ-Ansatz zu quantifizieren und soziale Aspekte bei der Interaktion der den SemDaServ-Ansatz durchführenden Akteure zu berücksichtigen.

Abstract

Without high-quality data, companies can't do anything important. (Redman, 2020) This pointed statement becomes even more significant when the renaissance of artificial intelligence (AI) is taken into account: Combined with machine learning (ML) and analytics, a whole new phase of value creation is being triggered, where strategic scale with data separates winners and losers in the race for market leadership. (Kagermann et al., 2018, S. 7) Therefore, the systematic development of data layers suitable for AI applications is essential.

Due to the increasing networking of products and their equipment with sensors, more and more data is becoming available that can potentially also be analyzed by AI. This is happening in particular in the area of smart services. A qualitatively and quantitatively sufficient amount of data is required for the development and successful operation of smart services as a component of Smart Product Service Systems (SPSS). This is especially true when data-driven software components use statistical methods from the field of AI: The quality of the training data directly determines the quality of the trained AI model. However, the quality of the training data can only be assessed with existing methods in late product development phases, since the training success is only known once the AI model has been trained. Therefore, this dissertation addresses the following guiding question: *How can the availability of qualitatively sufficient data relevant to smart service be increased in a practical way to improve the quality of the data-driven software components of a smart product service system (SPSS)?*

The answer to the guiding question follows the Design Research Methodology (DRM) presented in L. T. Blessing and Chakrabarti (2009). A comprehensive descriptive study I results in a reference model with a total of 93 influence factors, nine key factors and nine success factors. Based on the reference model, an impact model and a requirements list are derived, which are used in the prescriptive study to develop a functional architecture.

The resulting solution concept based on functional architecture and requirements list is called Semantic Data Specification for AI-based Smart Services (SemDaServ). SemDaServ addresses the dimensions of processes and organization, virtual and physical artifacts, as well as tools and IT systems based on Lünemann et al. (2017, S. 319) Engineering Operating System (EOS) to support the activity of semantic data specification for AI-based smart services. SemDaServ consists of three overarching process steps (Clarify Knowledge Need, Clarify Information Need, Specify Data Need), six artifacts (Data Category Model, Information Model, Data Model, Knowledge Matrix, Data Decision Matrix, Data Specification Checklist), and the description for using IT tools to create the artifacts (including modeling tools, data science tools, enterprise IT). The SemDaServ approach is prototypically realized using a Smart Home Office Lightning Service.

Based on the prototypical realization, descriptive study II is conducted to evaluate the SemDaServ approach. The qualitative evaluation based on two workshops, eight expert interviews and two case studies shows that the SemDaServ approach is a useful and practically applicable approach to increase the availability of qualitatively sufficient data with relevance for AI-based smart services. Further research (especially A/B comparisons) is needed to quantify the effectiveness and efficiency of the SemDaServ approach and to consider social aspects in the interaction of the actors implementing the SemDaServ approach.