

Feasibility Study: Optimizing Data Management for Your AI Project Success

Make Your Data AI-Ready!

Effective data management is crucial for successful AI training. Poor data management can lead to errors, biases, and unreliable models, but investing in professional services guarantees high-quality data, regulatory compliance, and optimized workflows, saving time and reducing risks. Fraunhofer FIT, with its decades of expertise in data management, AI research, and industry collaboration, is uniquely qualified to conduct tailored feasibility studies.

You will obtain:

- **Data Collection:** Identifying potential data sources (e.g., sensors, databases, APIs) and determining how to collect relevant data aligned with your AI objectives.
- **Data Cleaning:** Evaluating strategies to handle missing data, remove duplicates, resolve inconsistencies, and manage outliers.
- **Data Labeling:** Assessing methods for annotating data points, ensuring label accuracy, and addressing potential class imbalances for supervised learning tasks.
- **Data Transformation:** Identifying techniques for normalizing, standardizing, and encoding features, as well as considering feature engineering and dimensionality reduction where necessary.
- **Data Integration:** Evaluating approaches for merging datasets from various sources, resolving schema conflicts, and ensuring data consistency.
- **Data Segmentation:** Analyzing how to appropriately split the data into training, validation, and test sets to optimize model performance and prevent overfitting.
- **Data Quality Assurance:** Reviewing processes to ensure the dataset's completeness, accuracy, consistency, and relevance to the AI task at hand.
- **Data Storage and Management:** Identifying secure storage solutions, version control systems, and scalable options for managing growing data volumes.

- **Data Privacy and Security:** Evaluating strategies to anonymize sensitive data, encrypt information, and ensure compliance with legal and ethical standards.
- **Validation of Prepared Data:** Assessing the steps necessary to test data pipelines, validate functionality, and conduct a final quality review to prepare the data for AI model training.

The outcome of this feasibility study is a comprehensive, structured approach to preparing high-quality data for AI applications. This ensures that your data is optimized, secure, and compliant with industry standards—laying a solid foundation for reliable, efficient, and unbiased AI model training.

Contact

Dr. Christina Gillmann
Head of Data Management and Engineering Group
Phone +49 2241 143-745
christina.gillmann@fit.fraunhofer.de

Department
Data Science and Artificial Intelligence

Fraunhofer Institute for
Applied Information Technology FIT
Schloss Birlinghoven
53757 Sankt Augustin | Germany
www.fit.fraunhofer.de

Machbarkeitsstudie: Optimierung des Datenmanagements für den Erfolg Ihres KI-Projekts

Machen Sie Ihre Daten KI-bereit!

Effektives Datenmanagement ist entscheidend für erfolgreiches KI-Training. Mangelhaft verwaltete Daten können zu Fehlern, Verzerrungen und unzuverlässigen Modellen führen. Die Investition in professionelle Dienstleistungen stellt hingegen sicher, dass Ihre Daten qualitativ hochwertig, rechtskonform und optimal aufbereitet sind – dies spart Zeit und minimiert Risiken. Mit jahrzehntelanger Erfahrung in Datenmanagement, KI-Forschung und Industriekooperation ist das Fraunhofer FIT bestens qualifiziert, maßgeschneiderte Machbarkeitsstudien durchzuführen.

Sie erhalten:

- **Datensammlung:** Identifikation potenzieller Datenquellen und Bestimmung geeigneter Methoden zur Erhebung relevanter Daten im Einklang mit Ihren KI-Zielen.
- **Datenannotierung:** Bewertung von Methoden zur Datenbeschriftung, Sicherstellung der Label-Genauigkeit und Lösung von Klassenungleichgewichten für überwachtes Lernen.
- **Datenumwandlung:** Auswahl geeigneter Techniken zur Normalisierung, Standardisierung und Kodierung von Merkmalen sowie gegebenenfalls Durchführung von Feature Engineering und Dimensionsreduktion.
- **Datenintegration:** Analyse von Ansätzen zur Zusammenführung von Datensätzen aus verschiedenen Quellen, Behebung von Schema-Konflikten und Sicherstellung der Datenkonsistenz.
- **Datenaufteilung:** Untersuchung optimaler Strategien zur Aufteilung der Daten in Trainings-, Validierungs- und Testsets zur Optimierung der Modellleistung
- **Datenqualitätsprüfung:** Sicherstellung der Vollständigkeit, Genauigkeit, Konsistenz und Relevanz des Datensatzes für die gegebene KI-Aufgabe.

- **Datenverwaltung & Speicherung:** Identifikation sicherer Speicherlösungen, Versionskontrollsysteme und skalierbarer Optionen zur Verwaltung wachsender Datenmengen.
- **Validierung der aufbereiteten Daten:** Überprüfung und Testen der Datenpipelines, Funktionsvalidierung und abschließende Qualitätskontrolle zur optimalen Vorbereitung der Daten für das KI-Training.

Das Ergebnis dieser Machbarkeitsstudie ist ein umfassender, strukturierter Ansatz zur Bereitstellung hochwertiger Daten für KI-Anwendungen. Dies gewährleistet, dass Ihre Daten optimiert, sicher und konform mit Industriestandards sind – und damit eine solide Grundlage für zuverlässiges, effizientes und unverzerrtes KI-Training bilden.

Kontakt

Dr. Christina Gillmann
Gruppenleiterin Datenmanagement und Engineering
Telefon +49 2241 143-745
christina.gillmann@fit.fraunhofer.de

Abteilung
Data Science und Künstliche Intelligenz

Fraunhofer-Institut für
Angewandte Informationstechnik FIT
Schloss Birlinghoven
53757 Sankt Augustin | Germany
www.fit.fraunhofer.de