Curriculum for

Certified Professional for Software Architecture (CPSA)®
Advanced Level

Module
SWARC4AI

Software Architecture for AI Systems

2024.1-rev1-EN-20250208

iSAQB®

International Software Architecture
Qualification Board

Table of Contents

List of Learning Goals . . . 2

Introduction: General information about the iSAQB Advanced Level . . . 5

What is taught in an Advanced Level module? . . . 5

What are Advanced Level (CPSA-A) graduates qualified for? . . . 5

Requirements for CPSA-A certification . . . 5

Essentials . . . 6

What does the module “SWARC4AI” convey? . . . 6

Curriculum Structure and Recommended Durations . . . 6

Duration, didactics and further details . . . 6

Prerequisites . . . 7

Structure of the curriculum . . . 7

Supplementary Information, Terms, Translations . . . 7

1. Introduction to software architecture-relevant concepts for artificial intelligence . . . 8

1.1. Terms and Concepts . . . 8

1.2. Learning Goals . . . 8

1.3. References . . . 9

2. Compliance, security, alignment . . . 10

2.1. Terms and Concepts . . . 10

2.2. Learning Goals . . . 10

2.3. References . . . 12

3. Design and development of AI systems . . . 13

3.1. Terms and Concepts . . . 13

3.2. Learning Goals . . . 13

3.3. References . . . 15

4. Data management and data processing for AI systems . . . 17

4.1. Terms and Concepts . . . 17

4.2. Learning Goals . . . 17

4.3. References . . . 17

5. Important quality features for the operation of AI systems . . . 19

5.1. Terms and Concepts . . . 19

5.2. Learning Goals . . . 19

5.3. References . . . 21

6. System architectures and platforms for generative AI systems . . . 22

6.1. Terms and Concepts . . . 22

6.2. Learning Goals . . . 22

6.3. References . . . 25

7. Case studies and practical projects . . . 26

7.1. Learning Goals . . . 26
References . . . 27

© (Copyright), International Software Architecture Qualification Board e. V. (iSAQB® e. V.) 2024

The curriculum may only be used subject to the following conditions:

1. You wish to obtain the CPSA Certified Professional for Software Architecture Foundation Level® certificate or the CPSA Certified Professional for Software Architecture Advanced Level® certificate. For the purpose of obtaining the certificate, it shall be permitted to use these text documents and/or curricula by creating working copies for your own computer. If any other use of documents and/or curricula is intended, for instance for their dissemination to third parties, for advertising etc., please write to info@isaqb.org to enquire whether this is permitted. A separate license agreement would then have to be entered into.
2. If you are a trainer or training provider, it shall be possible for you to use the documents and/or curricula once you have obtained a usage license. Please address any enquiries to info@isaqb.org. License agreements with comprehensive provisions for all aspects exist.
3. If you fall neither into category 1 nor category 2, but would like to use these documents and/or curricula nonetheless, please also contact the iSAQB e. V. by writing to info@isaqb.org. You will then be informed about the possibility of acquiring relevant licenses through existing license agreements, allowing you to obtain your desired usage authorizations.

Important Notice

We stress that, as a matter of principle, this curriculum is protected by copyright. The International Software Architecture Qualification Board e. V. (iSAQB® e. V.) has exclusive entitlement to these copyrights.

The abbreviation "e. V." is part of the iSAQB’s official name and stands for "eingetragener Verein" (registered association), which describes its status as a legal entity according to German law. For the purpose of simplicity, iSAQB e. V. shall hereafter be referred to as iSAQB without the use of said abbreviation.

List of Learning Goals

- LG 1-1: Classify artificial intelligence and machine learning, data science, deep learning, generative AI
- LG 1-2: Specify typical general use cases for AI
- LG 1-3: Know possible applications in different industries and end-user applications
- LG 1-4: Identify risks in the application of AI
- LG 1-5: Understanding differences to traditional software
- LG 1-6: Decide on solutions to problems using AI or classic software development
- LG 1-7: Know roles and their tasks as well as their cooperation in the context of AI
- LG 1-8: Identify and prioritize AI use cases
- LG 1-9: Know the strengths and limitations of AI
- LG 1-10: Know the concept of the Productivity J-Curve in conjunction with AI technology
- LG 2-1: Know the influence of data protection laws on the implementation and use of AI
- LG 2-2: Understand the objectives and regulations of the EU AI Act and their impact on the development process and architecture
- LG 2-3: Perform classification of AI systems according to the EU AI Act risk level
- LG 2-4: Classify copyright issues of AI-generated content
- LG 2-5: Overview of types and degrees of openness and types of licenses of free ML models
- LG 2-6: Understand strategies for compliance with the EU AI Act and potential challenges
- LG 2-7: Document models and data sets for traceability and transparency
- LG 2-8: Know security pitfalls and types of attack on ML models
- LG 2-9: Know and apply strategies for AI risk minimization
- LG 2-10: Apply options for protection against attacks (AI security)
- LG 2-11: Understanding the basic issues and facets of AI safety
- LG 2-12: Understand ethical problems of AI systems and know approaches to dealing with them
- LG 2-13: Overview of ethical guidelines
- LG 2-14: Know the core principles of AI governance and responsible AI for companies
- LG 2-15: Gain insight into the creation of regulatory sandboxes
- LG 2-16: Ensuring effective data management for the quality and security of data in AI applications
- LG 3-1: Understanding the life cycle of a machine learning or data science project
- LG 3-2: Knowing process models for the software development of AI systems
- LG 3-3: Know types of and requirements for data and typical ML problems
- LG 3-4: Understanding machine learning problems and their requirements
- LG 3-5: Use input data for various AI algorithms
- LG 3-6: Dealing with challenges such as non-determinism, data quality and concept and model drift
- LG 3-7: Knowing transfer learning or fine-tuning

- LG 3-8: Selecting design patterns for AI systems
- LG 3-9: Define the task of an ML model
- LG 3-10: Understand and specify inputs and outputs for the functioning of an ML system
- LG 3-11: Know metrics for measuring the performance of ML models
- LG 3-12: Integrating ML models into existing systems
- LG 3-13: Designing user interfaces for AI systems
- LG 3-14: Understanding performance metrics such as latency and throughput in AI systems
- LG 3-15: Managing scalability for increased data volumes
- LG 3-16: Understanding robustness in AI systems and applying strategies to increase robustness
- LG 3-17: Classify the reliability and availability of AI systems
- LG 3-18: Understanding the reproducibility and testability of AI results
- LG 3-19: Know security, data protection and compliance requirements
- LG 3-20: Classify explainability and interpretability in AI systems
- LG 3-21: Recognizing bias in data and models
- LG 3-22: Knowing fault tolerance in AI systems
- LG 4-1: Acquiring and labeling data
- LG 4-2: Know common platforms for publicly accessible data
- LG 4-3: Overview of relevant tools for data labeling
- LG 4-4: Designing efficient data pipelines and architectures
- LG 4-5: Know strategies for data aggregation, cleansing, transformation, enrichment and augmentation
- LG 4-6: Overview of tools for data engineering pipelines
- LG 4-7: Know the options for storing data
- LG 5-1: Know (hardware) requirements for training and inference
- LG 5-2: Know the trade-offs of different model architectures with regard to quality characteristics
- LG 5-3: Adjust different quality features of an ML model
- LG 5-4: Understanding the costs, power consumption and sustainable use of AI (Green IT)
- LG 5-5: Know (hardware) requirements for training and inference
- LG 5-6: Model training, parameters, metrics and results tracking
- LG 5-7: Evaluate ML models and AI systems based on them
- LG 5-8: Know types of drift and possible causes and solutions for them
- LG 5-9: Overview of CI/CD pipelines, model management and deployment strategies for AI models
- LG 5-10: Know platforms for model provision
- LG 5-11: Classify tools for the creation of POCs of AI systems
- LG 5-12: Knowing the deployment options of AI models

- LG 5-13: List advantages and disadvantages of SaaS and self-hosting
- LG 5-14: Overview of SaaS AI solutions
- LG 5-15: Know embedded deployments of ML models
- LG 5-16: Understanding monitoring with regard to AI-specific requirements
- LG 5-17: Overview of sample tools for monitoring
- LG 5-18: Understanding user feedback and methods and tools for collecting user feedback
- LG 5-19: Know methods for using feedback for model training
- LG 5-20: [OPTIONAL] Understanding the MLOps pipeline using a practical example
- LG 5-21: [OPTIONAL] Make build vs. buy decisions for MLOps systems/components
- LG 5-22: [OPTIONAL] Know MLOps tools and end-to-end platforms
- LG 6-1: Overview of integration levels of AI
- LG 6-2: Know libraries, interfaces and tools for the integration of AI models
- LG 6-3: Integrate AI systems into the overall architecture of an IT landscape
- LG 6-4: Overview of relevant quality features for AI systems
- LG 6-5: [OPTIONAL] Know evaluation frameworks for AI systems
- LG 6-6: [OPTIONAL] Discuss a case study with an imaginary professionalism
- LG 6-7: Fundamental understanding of generative AI
- LG 6-8: Understanding how LLMs work
- LG 6-9: Understand known patterns in the use of LLMs
- LG 6-10: Knowing use cases for RAG (Retrieval-Augmented Generation)
- LG 6-11: Knowing and understanding selected RAG techniques
- LG 6-12: Know types of prompt engineering
- LG 6-13: Overview of agentic workflows
- LG 6-14: Know a selection of design patterns for generative AI systems
- LG 6-15: Know techniques for evaluating LLM applications
- LG 6-16: Overview of known LLMs and selection criteria
- LG 6-17: Understanding the importance of cost management for GenAI applications
- LG 6-18: Give examples of common libraries, interfaces and tools related to LLM applications
- LG 6-19: [OPTIONAL] Know agentic AI software architectures, AI agent architecture components and types of AI agent architectures
- LG 7-1: [OPTIONAL] Apply the acquired knowledge in real-life scenarios using case studies and practical projects

Introduction: General information about the iSAQB Advanced Level

What is taught in an Advanced Level module?

The module can be attended independently of a CPSA-F certification.

- The iSAQB Advanced Level offers modular training in three areas of competence with flexible training paths. It takes into account individual inclinations and specializations.
- The certification takes the form of a term paper. The assessment and oral examination are carried out by experts appointed by the iSAQB.

What are Advanced Level (CPSA-A) graduates qualified for?

CPSA-A graduates can:

- design medium-sized to large IT systems independently and methodically
- assume technical and content-related responsibility in IT systems of medium to high criticality
- Conceptualize, design and document measures to achieve quality requirements and support development teams in the implementation of these measures
- Manage and carry out architecture-related communication in medium to large development teams

Requirements for CPSA-A certification

- Successful training and certification as a Certified Professional for Software Architecture, Foundation Level® (CPSA-F)
- At least three years of full-time professional experience in the IT sector; collaboration on the design and development of at least two different IT systems
    - Exceptions are allowed on application (e.g., collaboration on open source projects)
- Training and further education within the scope of iSAQB Advanced Level training courses with a minimum of 70 credit points from at least three different areas of competence
- Successful completion of the CPSA-A certification exam

CPSA-A®
Certified Professional for Software Architecture
Advanced Level

Essentials

What does the module “SWARC4AI” convey?

The module presents participants with modern software architecture concepts for AI systems as a means of designing powerful, scalable and integrable AI solutions. At the end of the module, participants will be familiar with the key principles of software architecture for AI systems and will be able to apply them to the design and implementation of machine learning and generative AI systems. With the help of the modeling techniques and architecture tools taught, they will be able to seamlessly integrate AI components into existing software systems. The course covers both machine learning systems and generative AI and teaches how these can be combined with classic software systems. Participants learn what the architecture for such hybrid systems must look like in order to ensure scalability, maintainability and extensibility.

Curriculum Structure and Recommended Durations

Contents	Recommended minimum duration (min)
Introduction to concepts relevant to software architecture for artificial intelligence	120
Compliance, security, alignment	120
Design and development of AI systems	320
Data management and data processing for AI systems	90
Important quality features for the operation of AI systems	160
System architectures and platforms for generative AI systems	160
Case studies and practical projects	110
Total	1080

Duration, didactics and further details

The duration of a training course on the SWARC4AI module should be at least 3 days, but can be longer. Providers may differ in terms of duration, didactics, type and structure of the exercises and the detailed course structure. In particular, syllabus leaves the type of examples and exercises completely open. Licensed training courses on SWARC4AI contribute to admission to the final Advanced Level certification examination (following credit points):

Methodical Competence:	10 Points
Technical Competence:	20 Points
Communicative Competence:	0 Points

Prerequisites

Participants should have the following knowledge and/or experience:

- A basic understanding of artificial intelligence, machine learning and data science
- A basic understanding of software architecture, DevOps and the design of software systems and APIs
- A basic knowledge of the Python programming language and its use for AI problems
- Overview of common libraries such as scikit-learn, TensorFlow and PyTorch

Knowledge in the following areas can be helpful for understanding some concepts

- Experience with the command line interface on Linux systems
- Knowledge from the iSAQB CPSA Foundation level training for a general understanding of software architecture, design patterns and methodologies

Structure of the curriculum

The individual sections of the curriculum are described according to the following structure:

- **Terms/concepts**: Essential core terms of this topic.
- **Teaching/practice time**: Defines the minimum amount of teaching and practice time that must be spent on this topic or its practice in an accredited training course.
- **Learning goals**: Describes the content to be conveyed including its core terms and principles.

This section therefore also outlines the skills to be acquired in corresponding training courses.

Supplementary Information, Terms, Translations

To the extent necessary for understanding the curriculum, we have added definitions of technical terms to the iSAQB glossary and complemented them by adding referenes to the literature.

1. Introduction to software architecture-relevant concepts for artificial intelligence

Duration: 120 min	Practice time: 0 min

1.1. Terms and Concepts

AI, Generative AI, Machine Learning, Symbolic AI, Evolutionary Algorithms, Statistical Learning, Deep Learning, LLMs, Hallucination, Bias, Jagged Technological Frontier, Data Preparation, Model Training, Model Evaluation, Feature Engineering, Data Science, Data Engineering

1.2. Learning Goals

LG 1-1: Classify artificial intelligence and machine learning, data science, deep learning, generative AI

Participants will know how artificial intelligence is defined and how machine learning, data science, deep learning and generative AI are categorized within it.

LG 1-2: Specify typical general use cases for AI

Participants will know how to specify typical general use cases for AI. This includes use cases for e.g. image recognition & generation, language processing, prediction, personalization and anomaly detection.

LG 1-3: Know possible applications in different industries and end-user applications

Participants will be familiar with the possible applications in various industries, e.g. marketing, medicine, robotics and content creation. They will also gain an overview of the possible uses of AI in end-user applications. These include, for example, voice assistants (chatbots) and recommendation systems (recommender engines).

LG 1-4: Identify risks in the application of AI

Participants can identify risks that arise when using AI. These can be the following risks, for example:

- Hallucinations
- Bias
- (un)fairness
- societal risks such as deepfakes, AI-enabled cyberattacks, safety risks in critical systems, social manipulation, intellectual property issues, etc.

LG 1-5: Understanding differences to traditional software

Participants understand the differences between AI systems and traditional software:

- Data-driven (with ML) - data-centric instead of code-centric development
- Probabilistic results (non-deterministic behavior)
- Statistical validation
- Experimental design: Support for rapid iteration and testing of models.
- Model complexity and interpretability: ML models, especially deep learning models, can be extremely complex. The "black box" character makes them difficult to interpret and explain.

- Debugging and testing are more complex due to non-determinism.
- Model decay: Continuous monitoring and retraining are necessary to maintain model performance.
- AI-specific regulation of industries and at EU level must be taken into account.
- Interoperability: Seamless integration into existing systems and technology stacks.

LG 1-6: Decide on solutions to problems using AI or classic software development

Participants can decide and explain whether and why a problem should be solved using either AI or classic software development.

LG 1-7: Know roles and their tasks as well as their cooperation in the context of AI

The participants know typical roles and their tasks in these contexts. The roles include in particular

Data Scientist, Data Analyst, Data Engineer, Machine Learning Engineer, MLOps Engineer, AI Architect, Data Architect, Business Intelligence (BI) Developer, Data Governance Specialist and ML Researcher.

LG 1-8: Identify and prioritize AI use cases

Participants will be able to identify and prioritize AI use cases.

LG 1-9: Know the strengths and limitations of AI

Participants understand the strengths and limitations of AI and are familiar with the so-called "Jagged Technological Frontier".

LG 1-10: Know the concept of the Productivity J-Curve in conjunction with AI technology

This phenomenon helps to understand why companies can initially experience a drop in productivity when implementing AI, but this can be followed by productivity gains as it develops further.

1.3. References

[Roser 2022], [Brynjolfsson et al.], [Burkov 2019], [Géron 2022], [Kelleher 2015], [Vaughan 2020], [Bahree 2024], [Harvard et al. 2024], [Dell'Acqua 2022], [Visengeriyeva, JTF], [Agrawal et al.], [Tan et al.], [Chong et al.], [Hall et al. 2023], [Huyen 2022], [Wang et al. 2024]

2. Compliance, security, alignment

Duration: 120 min	Practice time: 30 min

2.1. Terms and Concepts

EU AI Act, data protection, copyright, license, open (source), AI security, jailbreak, adversarial attack, data poisoning, model inversion & extraction, AI safety, AI ethics, AI alignment, model and data documentation, transparency obligation, human oversight, data governance, AI governance, AI systems by risk levels (prohibited, high-risk, limited risk, low-risk)

2.2. Learning Goals

LG 2-1: Know the influence of data protection laws on the implementation and use of AI

Participants are familiar with data protection laws such as the GDPR and know how they affect the collection, processing and storage of data by AI systems.

LG 2-2: Understand the objectives and regulations of the EU AI Act and their impact on the development process and architecture

The participants understand the objectives and regulations of the EU AI Act and know what influence this has on the development and use of AI systems. They also understand the requirements of the EU AI Act (Trustworthy AI) and what influence these requirements have on the development process and the architecture of the software system. In particular, they know the influence on some of the following aspects:

- Risk management system (risk minimization)
- Data quality and data governance (quality management system)
- Creation and maintenance of comprehensive technical documentation for the AI system
- Automatic recording/logging of events in the AI system.
- Provision of clear and understandable information for users
- Implementation of measures for human supervision
- Ensuring an appropriate level of accuracy and robustness
- Implementation of cyber security measures

LG 2-3: Perform classification of AI systems according to the EU AI Act risk level

The participants know the classification of AI systems according to the EU AI Act risk levels (prohibited, high-risk, limited-risk, low-risk) and know which regulatory requirements apply in each case.

LG 2-4: Classify copyright issues of AI-generated content

The participants understand the copyright issues for AI-generated content, know the effects on certain software license models and can devise potential solutions for these issues.

LG 2-5: Overview of types and degrees of openness and types of licenses of free ML models

The participants gain an overview of different types and degrees of openness of free ML models. This concerns, for example, the disclosure of data and model parameters. In addition, they know different types of licenses of free ML models and their impact on the AI system.

LG 2-6: Understand strategies for compliance with the EU AI Act and potential challenges

The participants understand the basic statements of the EU AI Act (in particular transparency obligations) and know strategies for compliance and possible challenges.

LG 2-7: Document models and data sets for traceability and transparency

Participants know how to effectively document models and data sets to ensure traceability and transparency.

LG 2-8: Know security pitfalls and types of attack on ML models

Participants will be familiar with possible security pitfalls and typical types of attack on ML models. This applies, for example:

- LLM jailbreaks through prompt engineering
- Adversarial Attacks
- Data Poisoning
- Model Inversion & Extraction.

LG 2-9: Know and apply strategies for AI risk minimization

Participants know how to develop and apply strategies to minimize AI risks. In particular, the participants know the following strategies:

- Strengthening robustness through extensive testing
- Fault-tolerant AI systems
- Transparent development
- Explainable AI

LG 2-10: Apply options for protection against attacks (AI security)

Participants will be familiar with various options for protecting against attacks (AI security) and, in particular, they will be familiar with options for integrating security standards into the architecture and can take these into account for the design of the system.

LG 2-11: Understanding the basic issues and facets of AI safety

The participants understand the basic problems of AI safety and know the various facets of it. In particular, participants will be familiar with specific problems such as "AI model risks by poisoning" and "bias".

LG 2-12: Understand ethical problems of AI systems and know approaches to dealing with them

Participants are aware of the ethical problems that AI systems can bring with them and know approaches and possibilities for resolving these problems. This includes, for example, AI alignment (and its limits) and the creation of their own AI guidelines.

LG 2-13: Overview of ethical guidelines

The participants review important ethics guidelines such as the "EU Ethics Guidelines for trustworthy AI" and the "Google AI Ethics Guidelines".

LG 2-14: Know the core principles of AI governance and responsible AI for companies

The participants are familiar with the most important documents on AI governance in order to develop the core principles of AI governance and responsible AI for the company. This applies in particular to the following documents:

- OECD AI Principles, https://oecd.ai/en/ai-principles
- The Asilomar AI Principles, https://futureoflife.org/open-letter/ai-principles/
- The IEEE Ethically Aligned Design framework, https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

LG 2-15: Gain insight into the creation of regulatory sandboxes

Participants will gain an insight into the creation of regulatory sandboxes to promote innovation and the possible legal consequences of non-compliance with the provisions of the AI Act.

LG 2-16: Ensuring effective data management for the quality and security of data in AI applications

Participants will know how effective data management ensures the quality and security of data in AI applications.

2.3. References

[Engler et al.], [Nist], [Hall et al. 2023], [Masood et al. 2023], [CSIRO et al. 2023], [Pruksachatkun et al. 2023], [Chen et al. 2022], [ATLAS], [Visengeriyeva, Ethics], [EU AI Act], [Hotz, Best Practices], [Bhajaria 2022], [Jarmul 2023], [Molnar 2024]

3. Design and development of AI systems

Duration: 320 min	Practice time: 45 min

3.1. Terms and Concepts

CI/CD, MLOps, CRISP-ML(Q), AI application, AI engineering, ML model development, ML infrastructure, performance, robustness, reliability, fault tolerance, model-as-service, model-as-dependency, precompute, model-on-demand, hybrid serving, multi-agent system (MAS)

3.2. Learning Goals

LG 3-1: Understanding the life cycle of a machine learning or data science project

Participants have an understanding of the life cycle of a machine learning or data science project. In particular, they are familiar with the following phases:

- Exploratory Data Analysis
- Data cleansing and preparation
- Feature Engineering
- Model training and selection
- POC
- Deployment
- Maintenance

LG 3-2: Knowing process models for the software development of AI systems

Participants will be familiar with typical process models for the software development of AI systems. These are, for example

- CRISP-ML(Q)
- Team Data Science Process
- GenAI Life Cycle.

LG 3-3: Know types of and requirements for data and typical ML problems

Participants know different types of data and typical ML problems as well as use cases for utilizing the data. In addition, the participants have an understanding of various requirements for the data, e.g. the presence of correct labels of various types.

LG 3-4: Understanding machine learning problems and their requirements

The participants understand the various machine learning problems such as

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

and know what requirements they have.

LG 3-5: Use input data for various AI algorithms

Participants know which data is required for AI algorithms and have a good understanding of the need for validation and knowledge of the typical data breakdown into training, validation and test data. In addition, participants have a good understanding of the input data required for various AI algorithms such as

- Neural networks as numerical vectors and matrices or tensors
- One-Hot-Encodings
- embeddings

are required.

LG 3-6: Dealing with challenges such as non-determinism, data quality and concept and model drift

Participants understand how to deal with challenges such as non-determinism, data quality and concept and model drift.

LG 3-7: Knowing transfer learning or fine-tuning

The participants are familiar with transfer learning and fine-tuning as a way of incorporating the pre-trained basic models into existing use cases.

LG 3-8: Selecting design patterns for AI systems

Participants will know which design patterns exist for AI systems and how to select suitable patterns. This concerns the following design patterns:

- ML Systems Topology Patterns
- Pipeline Architecture Patterns
- Model Training Pattern
- Model Serving Patterns
- Model Deployment Patterns

LG 3-9: Define the task of an ML model

The participants know how to define the use cases / tasks of an ML model, such as the classification of images or fraud detection.

LG 3-10: Understand and specify inputs and outputs for the functioning of an ML system

Participants understand which inputs and outputs are required for the functioning of an ML system and can specify them.

LG 3-11: Know metrics for measuring the performance of ML models

Participants are familiar with various metrics for measuring the performance of ML models, such as:

- Precision, Recall
- F1
- Accuracy

and know how to define evaluation criteria for performance assessment.

LG 3-12: Integrating ML models into existing systems

Participants understand how ML models can be integrated into existing systems and are familiar with the interfaces and integration points.

LG 3-13: Designing user interfaces for AI systems

Participants know how user interfaces should be designed to enable effective interactions with the ML system and optimize the user experience.

LG 3-14: Understanding performance metrics such as latency and throughput in AI systems

Participants understand the importance of key performance indicators such as latency and throughput in AI systems and know how these can be optimized.

LG 3-15: Managing scalability for increased data volumes

Participants understand the importance of scalability for increased data volumes and know how to develop AI systems that can handle increasing data volumes without losing performance.

LG 3-16: Understanding robustness in AI systems and applying strategies to increase robustness

Participants have an understanding of what robustness means in AI systems and can apply strategies to increase robustness in different application contexts.

LG 3-17: Classify the reliability and availability of AI systems

Participants understand the concepts of reliability and availability and know how to build AI systems that are stable and constantly available.

LG 3-18: Understanding the reproducibility and testability of AI results

The participants know how important it is that AI results are reproducible and testable and know which methods can be used to ensure these properties.

LG 3-19: Know security, data protection and compliance requirements

Participants are familiar with the requirements for security, data protection and compliance and know how these are implemented in AI systems.

LG 3-20: Classify explainability and interpretability in AI systems

Participants understand the importance of explainability and interpretability in AI systems and know how to ensure this in order to promote trust and transparency.

LG 3-21: Recognizing bias in data and models

The participants know how bias in data and models can be recognized and reduced in order to ensure fairness and equal treatment in AI applications.

LG 3-22: Knowing fault tolerance in AI systems

Participants are familiar with the concepts of fault tolerance and can explain how AI systems remain functional despite errors or malfunctions.

3.3. References

[TU Berlin], [Bornstein et al.], [Crowe et al. 2024], [Lakshmanan et al.], [Alake], [Koc], [Cdteliot], [Visengeriyeva, AI Agents], [Zaharia et al.], [Savarese], [tdcox], [Studer et al.], [Hotz, Life Cycle], [Hotz, TDSP], [Saltz], [Serban], [Heiland et al. 2023], [Nahar et al.], [ML software architecture],

4. Data management and data processing for AI systems

Duration: 90 min	Practice time: 0 min

4.1. Terms and Concepts

Data acquisition, labeling, data pipeline, ETL processes, data aggregation, data cleansing, transformation, augmentation, files, RDBMS, NoSQL, data products, data contracts, data architectures: Data Warehouse, Data Lake, Data Mesh

4.2. Learning Goals

LG 4-1: Acquiring and labeling data

Participants gain an overview of various methods for acquiring and labeling data.

LG 4-2: Know common platforms for publicly accessible data

Participants can name common platforms for publicly accessible data.

LG 4-3: Overview of relevant tools for data labeling

The participants know relevant tools for data labeling such as CVAT, Amazon Mechanical Turk.

LG 4-4: Designing efficient data pipelines and architectures

The participants have an understanding of the design of efficient data pipelines and data architectures, and they understand how to consider data quality for storage solutions and their management. To this end, participants are familiar with architectural patterns for data engineering pipelines and ETL processes.

LG 4-5: Know strategies for data aggregation, cleansing, transformation, enrichment and augmentation

The participants know strategies for data aggregation, cleansing, transformation, enrichment and augmentation.

LG 4-6: Overview of tools for data engineering pipelines

Participants will gain an overview of relevant tools for data engineering pipelines such as Apache Spark and Flink.

LG 4-7: Know the options for storing data

Participants will be familiar with various options for storing data as well as their advantages and disadvantages. The options include the following technologies:

- CSV files
- Column-oriented files
- Relational and NoSQL databases
- Data warehouses
- Data Lakes

4.3. References

[Sarkis], [Serra], [Dehghani], [Reis et al.], [Bornstein et al.], [Ford et al.], [Bhajaria 2022], [Sanderson et al.], [Jones]

5. Important quality features for the operation of AI systems

Duration: 160 min	Practice time: 30 min

5.1. Terms and Concepts

Quality features, scalability, performance optimization, monitoring, logging, feedback, FinOps for AI platforms

5.2. Learning Goals

LG 5-1: Know (hardware) requirements for training and inference

The participants know the different (hardware) requirements for TPU, GPU or CPU for training and inference, for example.

LG 5-2: Know the trade-offs of different model architectures with regard to quality characteristics

Participants will be able to name examples of trade-offs between different model architectures with regard to quality characteristics. In particular for scaling, efficiency and storage load, participants should know the trade-offs as well as the advantages and disadvantages of important architectures such as RNNs and transformers.

LG 5-3: Adjust different quality features of an ML model

Participants will be familiar with ways to adjust various quality characteristics such as accuracy, efficiency and storage load of an ML model and to trade them off against each other. In particular, the participants know the following techniques:

- Quantization
- Pruning
- Distillation
- LoRA

LG 5-4: Understanding the costs, power consumption and sustainable use of AI (Green IT)

Participants gain an understanding of the costs, power consumption and sustainable use of AI (green IT). In particular, participants will know how to develop AI models and systems that work in a resource-efficient manner by using memory, computing power and storage space efficiently.

LG 5-5: Know (hardware) requirements for training and inference

The participants know the term MLOps for the automation of the life cycle of a data science project and the connection with DevOps.

LG 5-6: Model training, parameters, metrics and results tracking

The participants have an understanding of tracking in model training, parameters, metrics and results.

LG 5-7: Evaluate ML models and AI systems based on them

Participants will gain an overview of approaches for evaluating ML models and AI systems based on them.

LG 5-8: Know types of drift and possible causes and solutions for them

Participants are familiar with different types of drift, such as data drift or model drift, as well as possible causes and solutions.

LG 5-9: Overview of CI/CD pipelines, model management and deployment strategies for AI models

Participants will have an understanding of CI/CD pipelines, model management and deployment strategies for AI models.

LG 5-10: Know platforms for model provision

The participants are familiar with common platforms for model provision, such as Huggingface Hub.

LG 5-11: Classify tools for the creation of POCs of AI systems

Participants will be able to name common tools for creating POCs for AI systems, such as Gradio, and understand how they work conceptually.

LG 5-12: Knowing the deployment options of AI models

Participants are familiar with various deployment options for AI models. Participants will understand a selection of the following options:

- API Deployment
- Embedded Deployment
- Batch Prediction
- Streaming
- Containerization
- Serverless Deployment
- Cloud services

LG 5-13: List advantages and disadvantages of SaaS and self-hosting

The participants know the advantages and disadvantages of SaaS and self-hosting and can weigh up the pros and cons.

LG 5-14: Overview of SaaS AI solutions

Participants will gain an overview of well-known SaaS AI solutions, such as Azure OpenAI Services.

LG 5-15: Know embedded deployments of ML models

Participants will be familiar with various options and standards for embedded deployments of ML models.

LG 5-16: Understanding monitoring with regard to AI-specific requirements

The participants understand the need for monitoring, also with regard to AI-specific requirements such as drift tracking. Participants are familiar with relevant metrics such as accuracy, precision, recall and F1 score, MAE, MSE, perplexity, latency, throughput and resource utilization and understand why these are relevant for monitoring.

LG 5-17: Overview of sample tools for monitoring

The participants know example tools for monitoring. This includes both general tools, such as Prometheus & Grafana, as well as ML-specific tools, such as MLflow.

LG 5-18: Understanding user feedback and methods and tools for collecting user feedback

Participants understand the benefits of user feedback for further model training. In addition, participants are familiar with various methods and tools for collecting user feedback, such as choosing between multiple answers and flagging in Gradio.

LG 5-19: Know methods for using feedback for model training

Participants are familiar with various methods of using feedback for model training, such as RLHF, RLAIF and DPO.

LG 5-20: [OPTIONAL] Understanding the MLOps pipeline using a practical example

Using a practical example, participants will learn what an MLOps pipeline can look like and what insights it offers into parameters, metrics, etc.

LG 5-21: [OPTIONAL] Make build vs. buy decisions for MLOps systems/components

Participants will be able to make build vs. buy decisions for MLOps systems/components.

LG 5-22: [OPTIONAL] Know MLOps tools and end-to-end platforms

The participants are familiar with well-known MLOps tools and end-to-end platforms, such as:

- Domino Data Lab, h2o.ai, DVC, activeloop, aporia, argo, arize, bentoML, comet ML, DagsHub, Databricks MLOps Stacks, Feast, Kedro, Kubeflow, Metaflow, MLflow, MLRun, prefect, PrimeHub, Weights & Biases, WhyLabs, zenML, KNIME, RapidMiner, NVIDIA AI Enterprise, watsonx.ai
- OpenSource: MLFlow, Weights & Biases, ClearML
- PaaS: AWS SageMaker, Azure ML.

5.3. References

[Chen et al. 2022], [Treveil et al. 2020], [Haviv et al. 2023], [Osipov 2022], [Tan Wei Hao et al. 2024], [Wilson 2022], [Salama et al.], [Kumara et al.]

6. System architectures and platforms for generative AI systems

Duration: 160 min	Practice time: 30 min

6.1. Terms and Concepts

Generative AI, LLMs, MLflow, Managed MLflow, Azure Machine Learning, Metaflow, Generative AI, LLM, (Stable) Diffusion, Vector DB, Embedding, RNN, Transformer, RAG, Agentic Workflows etc.

6.2. Learning Goals

LG 6-1: Overview of integration levels of AI

Participants are familiar with the various levels of AI integration. These include the following levels:

- Applications (e.g. coding assistants)
- AI engineering (e.g. prompt engineering)
- ML model development (e.g. pytorch)
- ML infrastructure (e.g. vector DBs).

LG 6-2: Know libraries, interfaces and tools for the integration of AI models

The participants know some examples of common libraries, interfaces and tools for the integration of AI models.

LG 6-3: Integrate AI systems into the overall architecture of an IT landscape

Participants know how integrate AI systems into IT landscapes at a strategic level. For example, they can use the strategic design of DDD (especially context maps) to determine and document the type and degree of integration of AI systems.

LG 6-4: Overview of relevant quality features for AI systems

The participants understand the quality characteristics that are particularly relevant for AI systems. This includes the following quality characteristics in particular:

- Reliability
- Scalability
- Efficiency
- Security
- Maintainability
- Interpretability

LG 6-5: [OPTIONAL] Know evaluation frameworks for AI systems

Participants are familiar with common evaluation frameworks such as LangSmith or LangFuse to deal with indeterminacy and errors in AI systems.

LG 6-6: [OPTIONAL] Discuss a case study with an imaginary professionalism

Participants will practise and discuss a case study with an imaginary technicality to weigh up integration options for AI into an existing software landscape.

LG 6-7: Fundamental understanding of generative AI

The participants have a basic understanding of generative AI such as LLMs and stable diffusion.

LG 6-8: Understanding how LLMs work

Participants understand how LLMs work and can categorize the associated terminology: Token, Embedding, RNN, Transformer, Attention.

LG 6-9: Understand known patterns in the use of LLMs

The participants understand known patterns for the use of LLMs. This includes the following patterns:

- RAG and retrieval strategies
- Function calling
- Finetuning
- Assistants
- Agents

LG 6-10: Knowing use cases for RAG (Retrieval-Augmented Generation)

The participants know typical use cases for RAG such as "Talk to your documents/database/API".

LG 6-11: Knowing and understanding selected RAG techniques

The participants know a selection of common RAG techniques such as:

- Simple RAG
- Context Enrichment Techniques
- Fusion Retrieval
- Intelligent Reranking
- Query Transformations
- Adaptive Retrieval
- Iterative retrieval
- Ensemble Retrieval
- Knowledge Graph Integration

LG 6-12: Know types of prompt engineering

The participants know different types of prompt engineering, such as

- Few-Shot-Learning
- Chain-of-Thought
- Role-Playing

as well as general best practices for prompting.

LG 6-13: Overview of agentic workflows

Participants know what agentic workflows are and are familiar with the terms reflection, tool usage, planning and multi-agent collaboration.

LG 6-14: Know a selection of design patterns for generative AI systems

The participants know which design patterns exist for generative AI systems. They know a selection of the following patterns:

- AI Query Router [Simple Router; Ranking-based Router; Learning-based Router
- Layered Caching Strategy Leading to Fine-Tuning
- Multiplexing AI Agents
- Fine-Tuning LLMs for Multiple Tasks
- Blending Rules-Based and Generative Approaches
- Utilizing Knowledge Graphs with LLMs
- Swarm of Generative AI Agents
- Modular Monolith LLM Approach with Composability
- Memory Cognition for LLMs
- Red and Blue Team Dual-Model Evaluation

LG 6-15: Know techniques for evaluating LLM applications

The participants know several techniques for evaluating LLM applications. These can be, for example, the following techniques:

- Scoring
- Human feedback
- Comparative Evaluation
- Model Based Evaluation

LG 6-16: Overview of known LLMs and selection criteria

The participants are familiar with well-known LLMs such as GPT, Claude, Gemini, Llama, Mistral or Luminous, and the selection criteria for a suitable LLM.

LG 6-17: Understanding the importance of cost management for GenAI applications

The participants understand the importance of cost management for GenAI applications.

LG 6-18: Give examples of common libraries, interfaces and tools related to LLM applications

The participants know some examples of common libraries, interfaces and tools in connection with LLM applications such as OpenAI-API or LangChain.

LG 6-19: [OPTIONAL] Know agentic AI software architectures, AI agent architecture components and types of AI agent architectures

The participants know agentic AI software architectures, AI agent architecture components, types of AI agent architectures.

6.3. References

[Koc], [Dibia 2025], [Gradient Flow], [bornstein-radovanic], [Bahree 2024], [Spirin et al.], [Foster 2023], [Parnin]

7. Case studies and practical projects

Duration: 110 min	Practice time: 110 min

7.1. Learning Goals

LG 7-1: [OPTIONAL] Apply the acquired knowledge in real-life scenarios using case studies and practical projects

References

This section contains references that are cited in the curriculum.

A

- [Agrawal et al.] A. Agrawal, J. Gans, A. Goldfarb: Prediction Machines: The Simple Economics of Artificial Intelligence https://www.predictionmachines.ai/
- [Alake] R. Alake: ML Pipeline Architecture Design Patterns (With 10 Real-World Examples) https://neptune.ai/blog/ml-pipeline-architecture-design-patterns
- [ATLAS] ATLAS - Adversarial Threat Landscape for Artificial-Intelligence Systems. https://github.com/mitre/advmlthreatmatrix

B

- [Bahree 2024] Bahree, A.: Generative AI in Action https://www.manning.com/books/generative-ai-in-action
- [Bhajaria 2022] N. Bhajaria: Data Privacy - A runbook for engineers https://www.manning.com/books/data-privacy
- [Bornstein et al.] M. Bornstein, J. Li, M. Casado: Emerging Architectures for Modern Data Infrastructure https://a16z.com/emerging-architectures-for-modern-data-infrastructure/
- [bornstein-radovanic] M. Bornstein and R. Radovanovic: Emerging Architectures for LLM Applications https://a16z.com/emerging-architectures-for-llm-applications/
- [Brynjolfsson et al.] Brynjolfsson, E.: The Productivity J-Curve: How Intangibles complement General Purpose Technologies https://www.nber.org/system/files/working_papers/w25148/w25148.pdf
- [Burkov 2019] Burkov, A.: The Hundred-Page Machine Learning Book https://themlbook.com/

C

- [Cdteliot] AI Agents: Understanding Their Impact and Functions https://www.perplexity.ai/page/ai-agents-understanding-their-bL1Mg8FeStyUB4o9u3HT5Q
- [Chen et al. 2022] C. Chen, N. R. Murphy, K. Parisa, D. Sculley, T. Underwood: Reliable Machine Learning https://www.oreilly.com/library/view/reliable-machine-learning/9781098106218/
- [Chong et al.] J. Chong, Y. C. Chang: How to Lead in Data Science https://www.manning.com/books/how-to-lead-in-data-science
- [Crowe et al. 2024] R. Crowe, H. Hapke, E. Caveness, D. Zhu: Machine Learning Production Systems https://learning.oreilly.com/library/view/machine-learning-production/9781098156008/
- [CSIRO et al. 2023] CSIRO, Q. Lu, J. Wittle, X. Xu, L. Xhu: Responsible AI: Best Practices for Creating Trustworthy AI Systems https://www.oreilly.com/library/view/responsible-ai-best/9780138073947/

D

- [Dehghani] Z. Dehghani: Data Mesh https://learning.oreilly.com/library/view/data-mesh/9781492092384/
- [Dell’Acqua 2022] Fabrizio Dell’Acqua et al.: Paper: “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality”

https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf

- [Dibia 2025] V. Dibia with C. Wang: Multi-Agent Systems with AutoGen https://www.manning.com/books/multi-agent-systems-with-autogen

E

- [Engler et al.] M. Engler, N. Dhamani: Generative AI. Misuse and Adversarial Attacks. https://learning.oreilly.com/library/view/introduction-to-generative/9781633437197/OEBPS/Text/05.html
- [EU AI Act] EU AI Act https://artificialintelligenceact.eu/de/ai-act-explorer/

F

- [Ford et al.] N. Ford, M. Richards, P. Sadalage, Z. Dehghani Software Architecture: The Hard Parts. https://learning.oreilly.com/library/view/software-architecture-the/9781492086888/
- [Foster 2023] D. Foster: Generative Deep Learning, 2nd Edition https://www.oreilly.com/library/view/generative-deep-learning/9781098134174/

G

- [Géron 2022] Aurélien Géron: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/
- [Gradient Flow] LLM Routers Unpacked https://gradientflow.com/llm-routers-unpacked/

H

- [Hall et al. 2023] P. Hall, J. Curtis, P. Pandey: Machine Learning for High-Risk Applications https://www.oreilly.com/library/view/machine-learning-for/9781098102425/
- [Harvard et al. 2024] Harvard Business Review, E. Mollick, D. De Cremer, T. Neeley, P. Sinha: Generative AI: The Insights You Need. (Generative AI Use Cases) https://learning.oreilly.com/library/view/generative-ai-the/9781647826406/
- [Haviv et al. 2023] Y. Haviv, N. Gift: Implementing MLOps in the Enterprise https://www.oreilly.com/library/view/implementing-mlops-in/9781098136574/
- [Heiland et al. 2023] L. Heiland, M. Hauser, J. Bogner: Design Patterns for AI-based Systems: A Multivocal Literature Review and Pattern Repository. 2023 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN). IEEE, 2023.
- [Hotz, Best Practices] N. Hotz: 15 Data Science Documentation Best Practices https://www.datascience-pm.com/documentation-best-practices/
- [Hotz, Life Cycle] N. Hotz: What is a Data Science Life Cycle? https://www.datascience-pm.com/data-science-life-cycle/
- [Hotz, TDSP] N. Hotz: What is TDSP https://www.datascience-pm.com/tdsp/
- [Huyen 2022] C. Huyen: Designing Machine Learning Systems https://www.oreilly.com/library/view/designing-machine-learning/9781098107956/

J

- [Jarmul 2023] K. Jarmul: Practical Data Privacy https://www.oreilly.com/library/view/practical-data-privacy/9781098129453/

- [Jones] A. Jones: Driving Data Quality with Data Contracts https://learning.oreilly.com/library/view/driving-data-quality/9781837635009/

K

- [Kelleher 2015] John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy: Fundamentals of Machine Learning for Predictive Data Analytics https://mitpress.mit.edu/9780262029445/fundamentals-of-machine-learning-for-predictive-data-analytics
- [Koc] V. Koc: Generative AI Design Patterns: A Comprehensive Guide https://towardsdatascience.com/generative-ai-design-patterns-a-comprehensive-guide-41425a40d7d0
- [Kumara et al.] I. Kumara, R., D. Di Nucci, W. J. Van Den Heuvel, D. A. Tamburri: Requirements and Reference Architecture for MLOps:Insights from Industry https://www.techrxiv.org/doi/full/10.36227/techrxiv.21397413.v1

L

- [Lakshmanan et al.] V. Lakshmanan, S Robinson, M. Munn: Machine Learning Design Patterns https://learning.oreilly.com/library/view/machine-learning-design/9781098115777/

M

- [Masood et al. 2023] A. Masood, H. Dawe: Responsible AI in the Enterprise https://www.oreilly.com/library/view/responsible-ai-in/9781803230528/
- [ML software architecture] ML software architecture https://appliedaiinitiative.notion.site/ML-software-architecture-790b9f5fcfcf408884287acb82f4d75e
- [Molnar 2024] C. Molnar: Interpretable Machine Learning, 2nd ed. https://christophm.github.io/interpretable-ml-book/

N

- [Nahar et al.] N. Nahar, et al.: A meta-summary of challenges in building products with ml components–collecting experiences from 4758+ practitioners. 2023 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN). IEEE, 2023.
- [NirDiamant] RAG Techniques https://github.com/NirDiamant/RAG_Techniques
- [Nist] NIST AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework

O

- [Osipov 2022] C. Osipov: MLOps Engineering at Scale https://www.manning.com/books/mlops-engineering-at-scale

P

- [Parnin] Building Your Own Product Copilot: Challenges, Opportunities, and Needs https://arxiv.org/pdf/2312.14231v1
- [Pruksachatkun et al. 2023] Y. Pruksachatkun, M. Mcateer, S. Majudmar: Practicing Trustworthy Machine Learning https://www.oreilly.com/library/view/practicing-trustworthy-machine/

9781098120269/

R

- [Reis et al.] J. Reis, M. Housley: Fundamentals of Data Engineering https://learning.oreilly.com/library/view/fundamentals-of-data/9781098108298/
- [Roser 2022] Roser, Max: Brief History of AI: https://ourworldindata.org/brief-history-of-ai

S

- [Salama et al.] K. Salama, J. Kazmierczak, D. Schut: Practitioners guide to MLOps: A framework for continuous delivery and automation of machine learning. https://services.google.com/fh/files/misc/practitioners_guide_to_mlops_whitepaper.pdf
- [Saltz] J. Saltz: The GenAI Life Cycle https://www.datascience-pm.com/the-genai-life-cycle/
- [Sanderson et al.] C. Sanderson, M. Freeman: Data Contracts https://learning.oreilly.com/library/view/data-contracts/9781098157623/
- [Sarkis] A. Sarkis: Training Data for Machine Learning https://learning.oreilly.com/library/view/training-data-for/9781492094517/
- [Savarese] S. Savarese: How AI Agents Will Revolutionize the AI Enterprise https://blog.salesforceairesearch.com/how-ai-agents-will-revolutionize-the-ai-enterprise/
- [Serban] A. Serban, J. Visser: "Adapting software architectures to machine learning challenges." 2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE, 2022.
- [Serra] J. Serra: Deciphering Data Architectures https://learning.oreilly.com/library/view/deciphering-data-architectures/9781098150754/
- [Spirin et al.] N. Spirin, M. Balint: Mastering LLM Techniques: LLMOps https://developer.nvidia.com/blog/mastering-llm-techniques-llmops/
- [Studer et al.] S. Studer et al.: Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology https://arxiv.org/abs/2003.05155

T

- [Tan et al.] D. Tan, A. Leung, D. Colls: Effective Machine Learning Teams https://learning.oreilly.com/library/view/effective-machine-learning/9781098144623/
- [Tan Wei Hao et al. 2024] B. Tan Wei Hao, S. Padmanabhan, V. Mallya: Design a Machine Learning System (From Scratch) https://www.manning.com/books/design-a-machine-learning-system-design-from-scratch
- [tdcox] MLOps Roadmap 2024 - DRAFT https://github.com/cdfoundation/sig-mlops/blob/main/roadmap/2024/MLOpsRoadmap2024.md
- [Treveil et al. 2020] M. Treveil, N. Omont, C. Stenac, K. Lefevre, D. Phan, J. Zentici, A. Lavoillotte, M. Miyazaki, L. Heidmann: Introducing MLOps https://www.oreilly.com/library/view/introducing-mlops/9781492083283/
- [TU Berlin] Architecture of Machine Learning Systems (TU Berlin, SS 2024): https://mboehm7.github.io/teaching/ss24_amls/index.htm

V

- [Vaughan 2020] Vaughan, D.: Analytical Skills for AI and Data Science (AI Use Cases) https://learning.oreilly.com/library/view/analytical-skills-for/9781492060932/
- [Visengeriyeva, JTF] Visengeriyeva, L.: Defining Jagged Technological Frontier:https://www.perplexity.ai/page/defining-jagged-technological-iF8sDPVFQEKSdd2oyytztA
- [Visengeriyeva, AI Agents] Visengeriyeva, L.: AI Agents vs. Traditional Models https://www.perplexity.ai/page/ai-agents-vs-traditional-model-JFf4gKT0RySW_Ehvbxho2g
- [Visengeriyeva, Ethics] Model Governance, Ethics, Responsible AI (Linksammlung) https://github.com/visenger/Awesome-ML-Model-Governance

W

- [Wang et al. 2024] C. Wang et al.: Quality Assurance for Artificial Intelligence: A Study of Industrial Concerns, Challenges and Best Practices https://arxiv.org/pdf/2402.16391
- [Wilson 2022] B. Wilson: Machine Learning Engineering in Action https://www.manning.com/books/machine-learning-engineering-in-action

Z

- [Zaharia et al.] M. Zaharia et al.: The Shift from Models to Compound AI Systems https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/