

Machine Learning Roles

After describing the principles and their resulting instantiation of components, we identify necessary roles in order to realize MLOps in the following. MLOps is an interdisciplinary group process, and the interplay of different roles is crucial to design, manage, automate, and operate an ML system in production. In the following, every role, its purpose, and related tasks are briefly described:

- Principal Investigator **R1** Business Stakeholder (similar roles: Product Owner, Project Manager). The business stakeholder defines the business goal to be achieved with ML and takes care of the communication side of the business, e.g., presenting the return on investment (ROI) generated with an ML product.
- ML Architect **R2** Solution Architect (similar role: IT Architect). The solution architect designs the architecture and defines the technologies to be used, following a thorough evaluation. * Data Scientist **R3** Data Scientist (similar roles: ML Specialist, ML Developer). The data scientist translates the business problem into an ML problem and takes care of the model engineering, including the selection of the best-performing algorithm and hyperparameters.
- Data Architect **R4** Data Engineer (similar role: DataOps Engineer). The data engineer builds up and manages data and feature engineering pipelines. Moreover, this role ensures proper data ingestion to the databases of the feature store system.
- ML/Data Engineer **R5** Software Engineer. The software engineer applies software design patterns, widely accepted coding guidelines, and best practices to turn the raw ML problem into a well-engineered product.
- MLOps Engineer **R6** DevOps Engineer. The DevOps engineer bridges the gap between development and operations and ensures proper CI/CD automation, ML workflow orchestration, model deployment to production, and monitoring.
- **R7** ML Engineer/MLOps Engineer. The ML engineer or MLOps engineer combines aspects of several roles and thus has cross-domain knowledge. This role incorporates skills from data scientists, data engineers, software engineers, DevOps engineers, and backend engineers (see Figure 3). This cross-domain role builds up and operates the ML infrastructure, manages the automated ML workflow pipelines and model deployment to production, and monitors both the model and the ML infrastructure.