



Quality-Driven Machine Learning-based Data Science Pipeline Realization: a software engineering approach

Giordano d'Aloisio
University of L'Aquila (Italy)

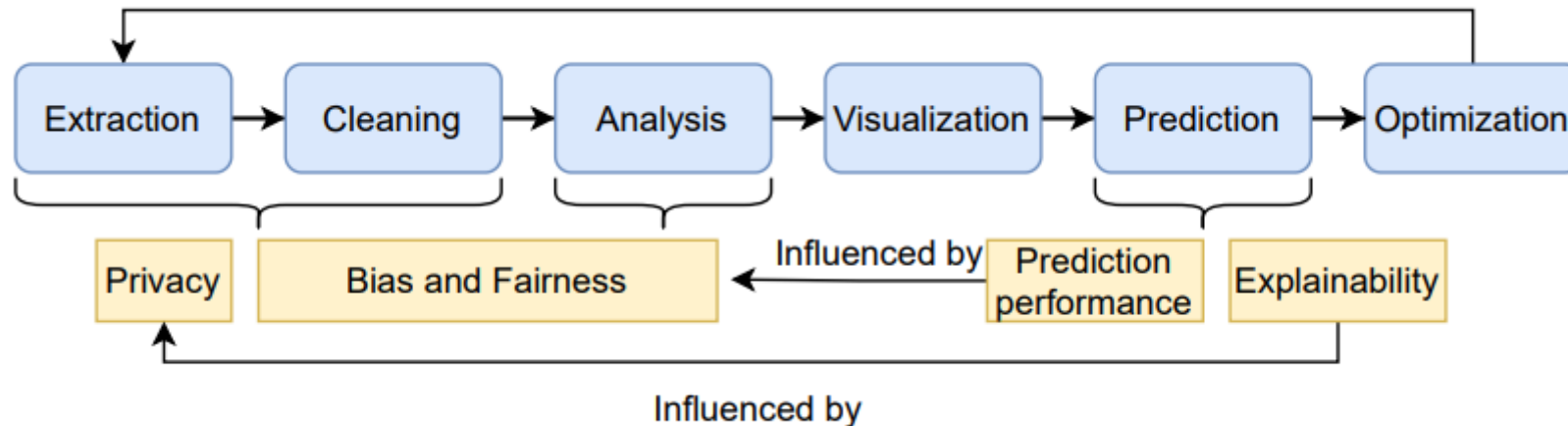


PARTNER

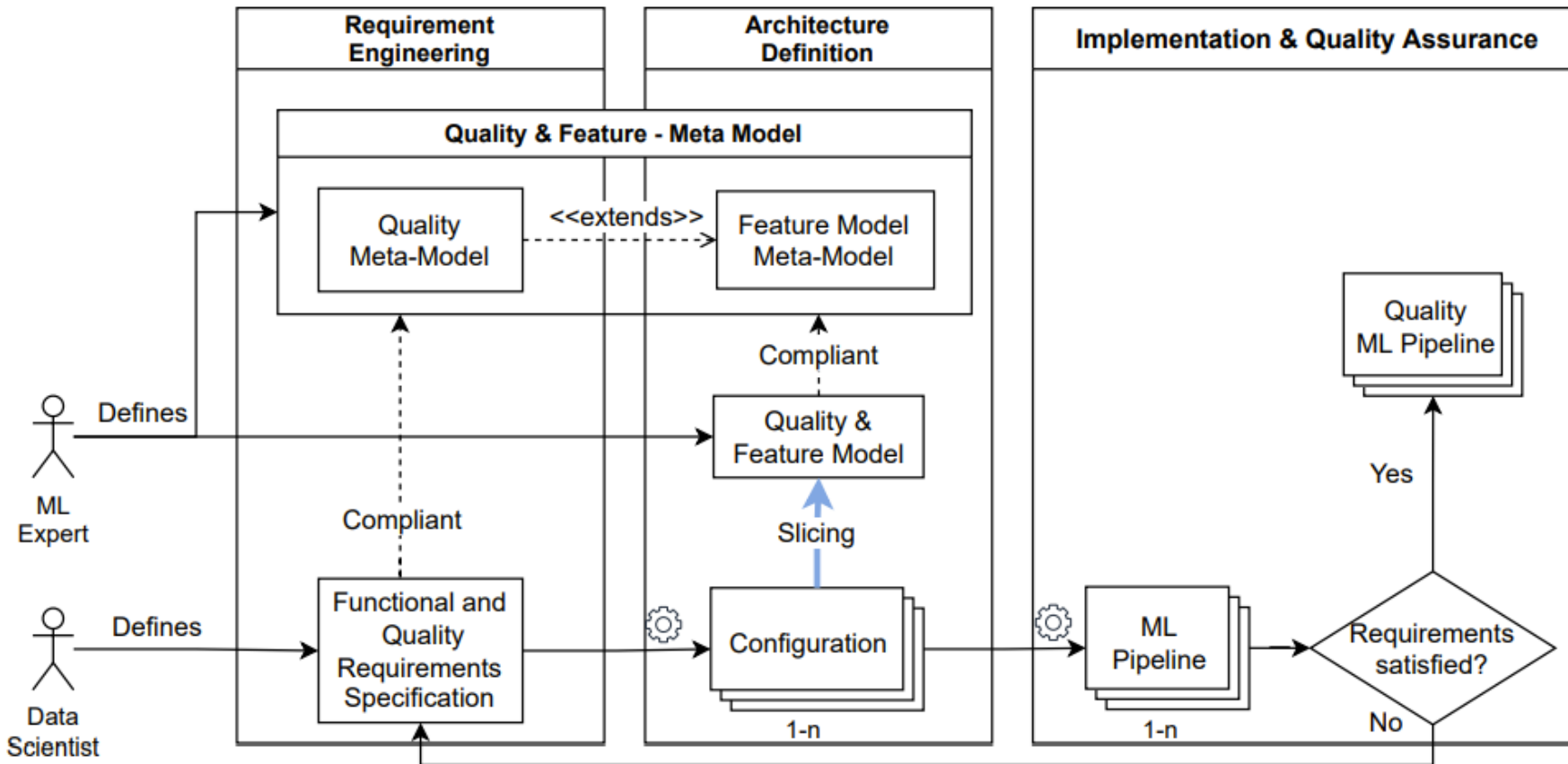


Introduction

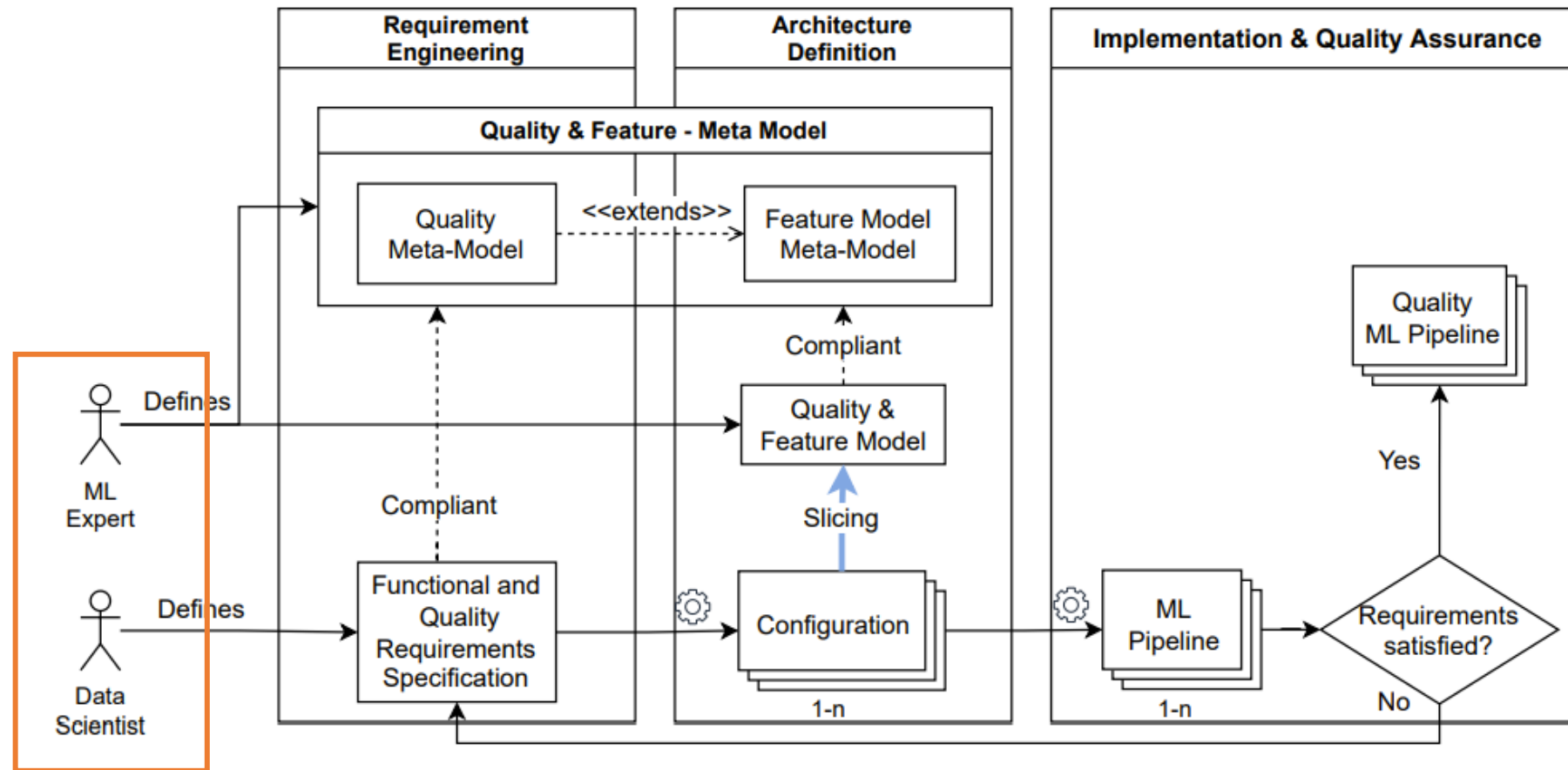
- ML systems are becoming a widely used instrument, applied in many domains
- They can be defined as a set of one or more pipelines which takes raw data and returns trained ML models
- Over the years many methods have been defined to help users in making ML pipelines. But we think that quality attributes must also be considered during their development



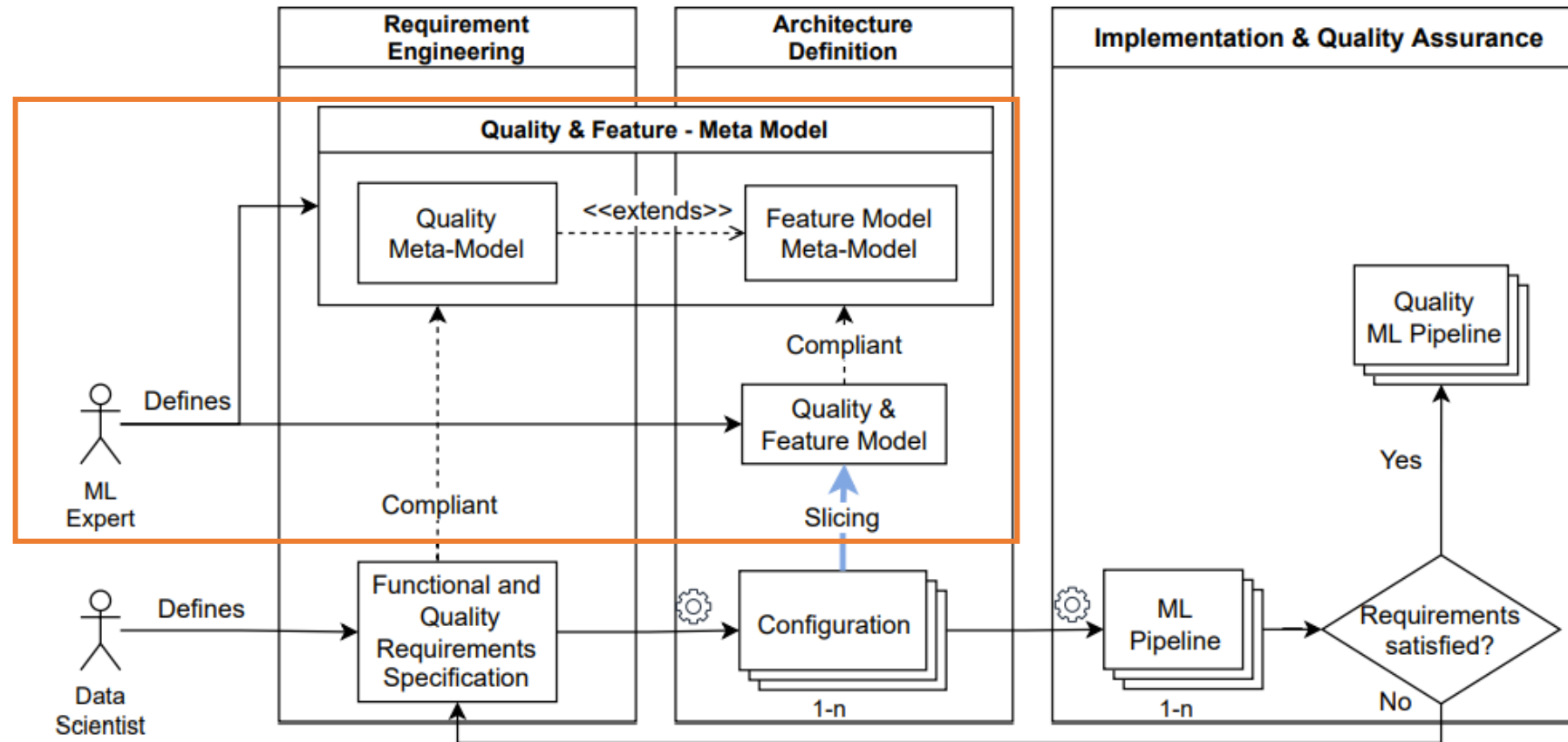
Proposed approach



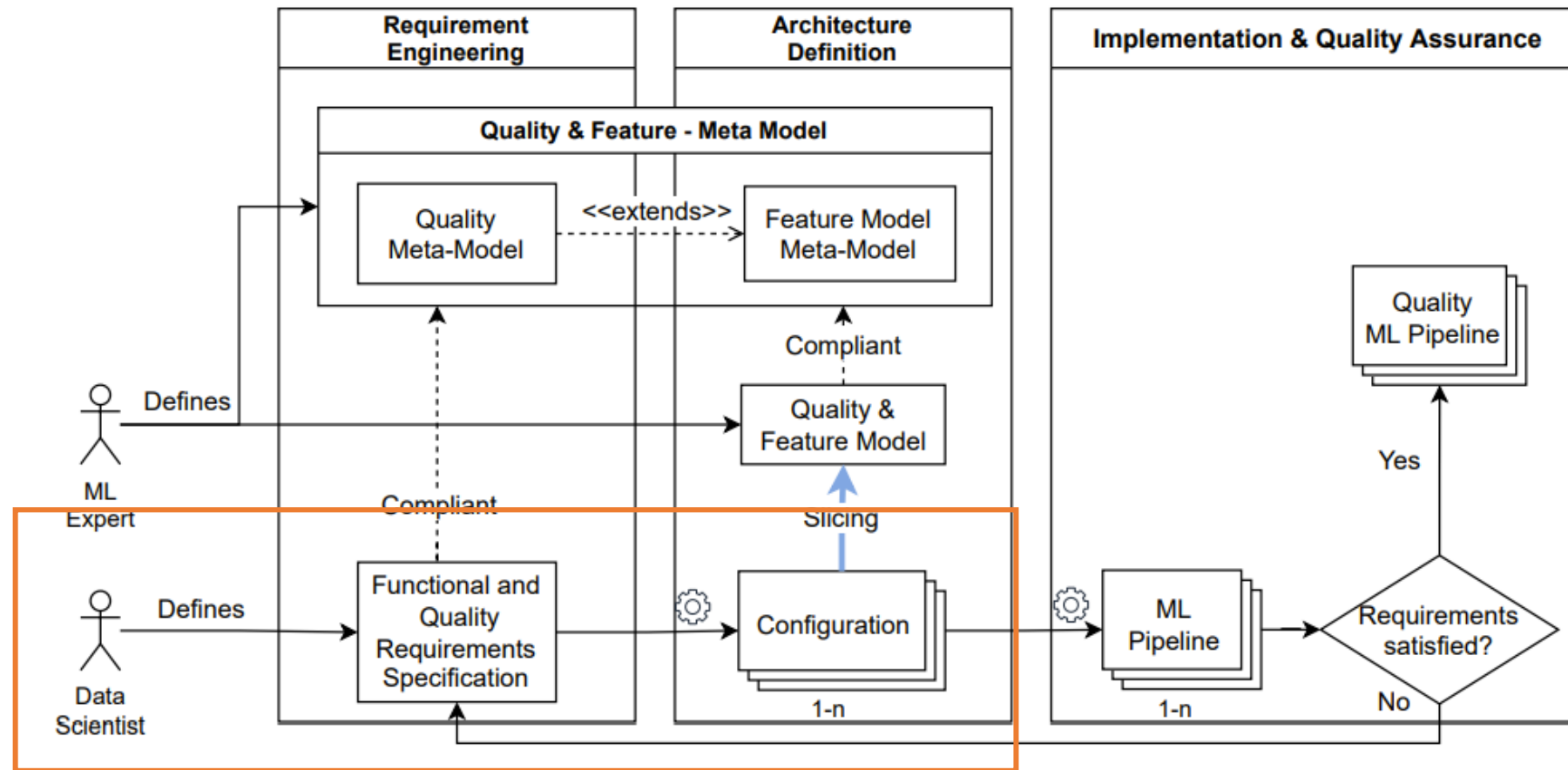
Proposed approach



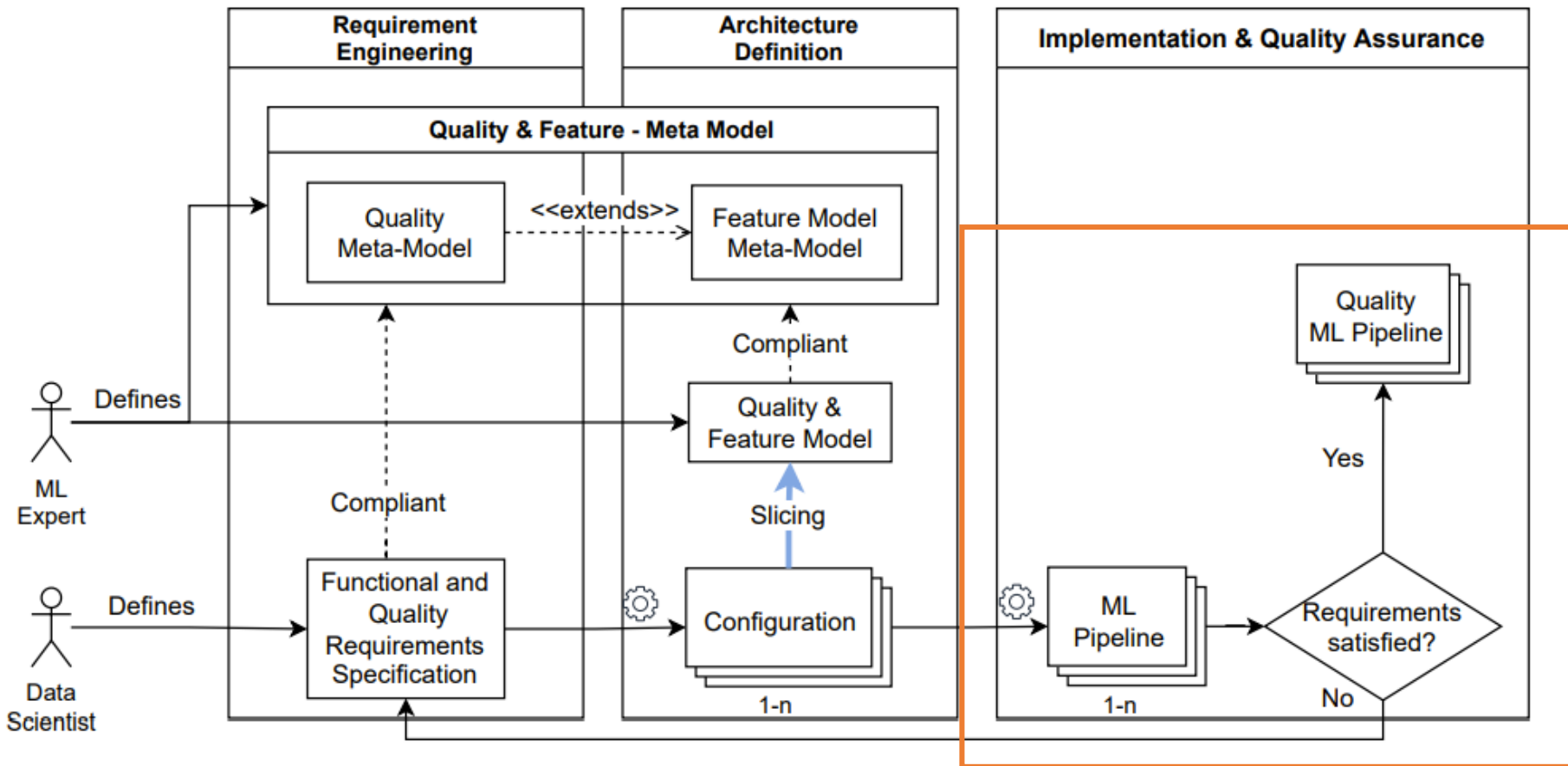
Proposed approach



Proposed approach



Proposed approach



Next goals and challenges

1. The project is in its beginning, and we plan to present the extended Quality and Feature Meta-Model as the first main result.
2. To define *quality* in ML systems we have selected from the literature the following attributes: *computational complexity, prediction correctness, fairness, explainability, privacy*, but also other attributes can be included depending on the stakeholders' constraints.
3. The next goal will be to verify that the generated ML pipelines satisfies the given specification.

Thank you for your attention!