

INTRODUCTION

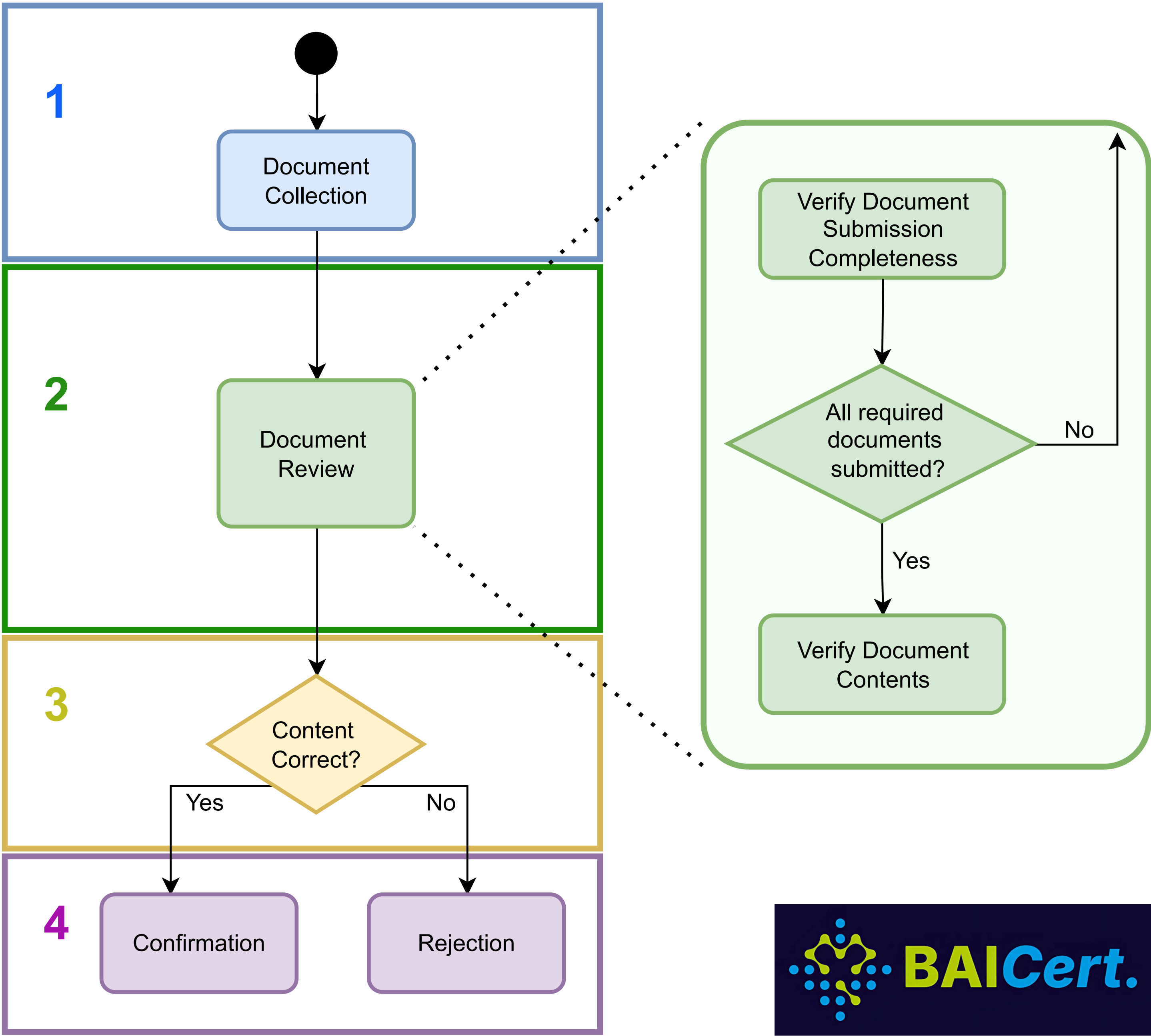
Audit and administrative processes in certification are often characterized by bureaucratic inefficiencies and limited automation. Experts frequently spend considerable time manually extracting and processing information from documents to produce certificates, audit reports, or compliance statements.

The project BAICert addresses these challenges by developing an **AI-driven document management system** that reduces manual workload while ensuring **transparency and explainability** of decisions made by artificial intelligence. By processing even **low-quality scans and images**, the system meets increasing regulatory requirements, mitigates labor shortages, and helps reduce bureaucratic hurdles – ultimately boosting regional competitiveness through more efficient, intelligent workflows.

PROJECT OBJECTIVES

- ▶ Information extraction from scans and photos, including low-quality images
- ▶ Document type classification with anomaly detection
- ▶ Explainable AI assistance for certification decisions and reporting
- ▶ Large Language Model (LLM) for automated text generation (e.g. in customer interaction)
- ▶ Improvement of existing standards, development of new approaches
- ▶ Deployment of open-source software

WORKFLOW OF THE REVIEW PROCESS



PROJECT SCOPE

- ▶ Project duration: 04.2025 – 03.2028
- ▶ Utilization of real datasets from cooperation partner
- ▶ Validation of prototypes in real-world settings through a cooperation partner
- ▶ MLOps-based quality assurance for sustainable AI models

FUNDAMENTAL PROJECT TASKS

Data Collection & Organization

- Heterogenous input: Data of varying quality and formats (scans, photos)
- Structured organization of input documents

Text Extraction

Convert images / text to machine-readable format

- Image Preprocessing
- Optical Character Recognition (OCR)

Understand meaning & intention of text

- Extraction of key information
- Natural Language Processing (NLP)

Classification & Anomaly Detection

Classification of document type

Identification of documents with atypical layouts

Automation of Certification Decisions

Automation of certification approval procedures

- Approval proposal for experts
- Consideration of the specific case and legal framework
- Focus on explainability and transparency

Text Generation

- Assisted Meta-Form completion
- Automated customer communication
- Large Language Models (LLM)

RESEARCH FOCUS

- ▶ Reliable data extraction from low-quality document images
- ▶ Robust document type classification
- ▶ Explainability & interpretability of AI models

ACKNOWLEDGEMENTS

This research is supported by the **Bavarian Collaborative Research Programme BayVFP** of the Bavarian State Ministry for Economic Affairs, Regional Development and Energy, funding line “Digitalization”, funding area “Information and Communication Technology”.

CONTACT

Laboratory Imaging and Data Science
Prof. Dr. Jürgen Friel
Prof. Dr. Filippo Riccio

Collaboration Partner:
OmniCert Umweltgutachter GmbH