

Beyond Static Checklists: A Dynamic, Modular Toolkit for Automated Metadata Assessment

Hamed Hemati¹, Alicia Janz¹, Stefan Sandfeld¹

¹Institute for Advanced Simulation - Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich

Motivation

Q: How do we verify research when publications are scaling faster than human reviewers?



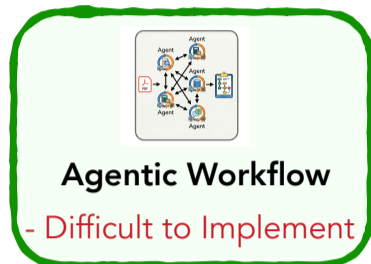
Manual

- Expensive



LLM-Based

- Hallucination



Agentic Workflow

- Difficult to Implement



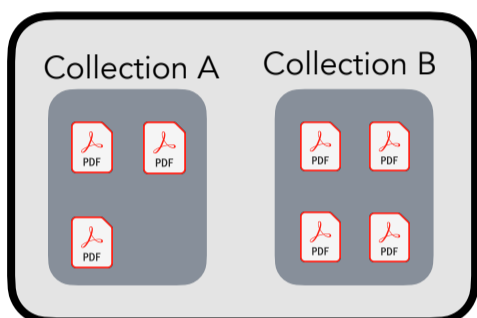
Example: Machine Learning Reproducibility Checklist

- Does it provide the checkpoints of the trained models?
- Does it provide theoretical proofs for all claims?
- Is the space complexity of the proposed method $O(N^2)$?

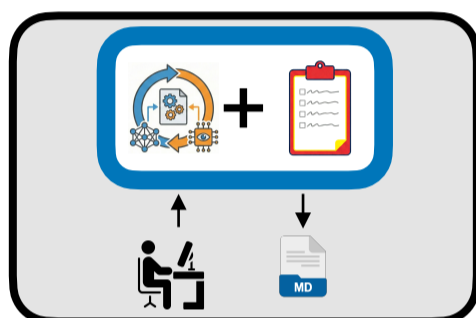
⋮

Toolkit Components

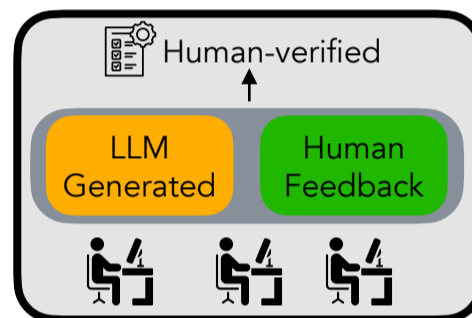
1. Data Collection



2. Review Process



3. Human Verification

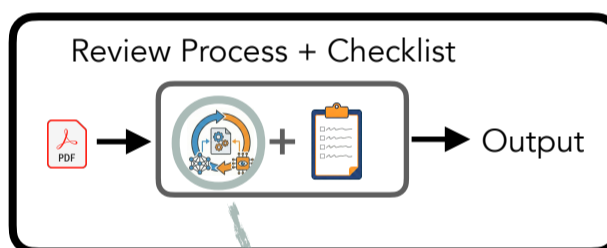


4. Analysis



Review Process

Researchers can obtain various outputs for the same collection and checklist using different review processes.

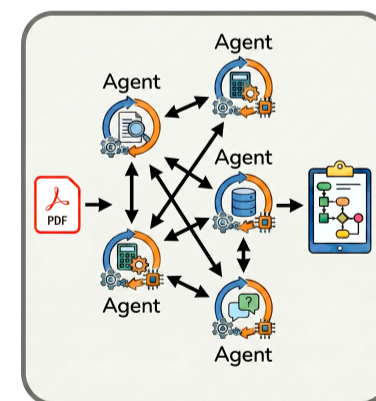
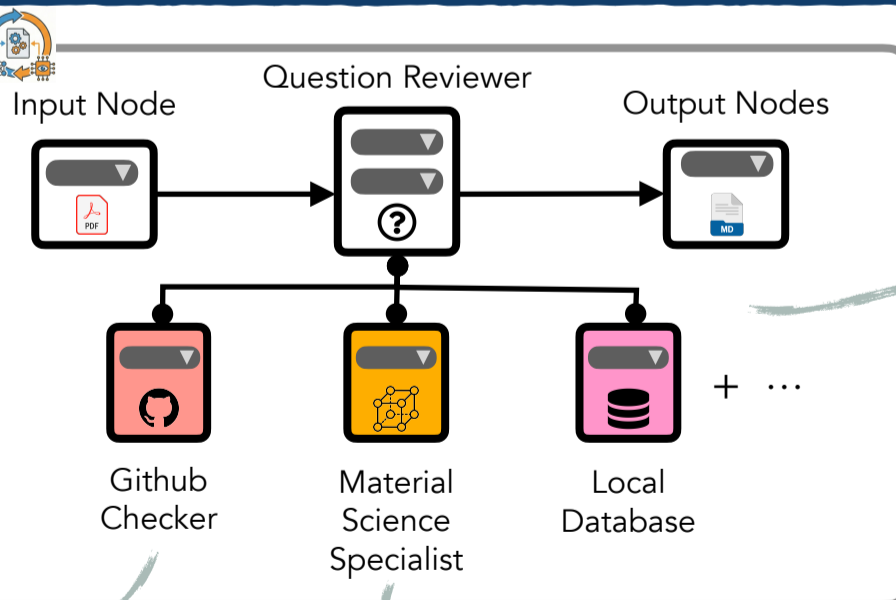
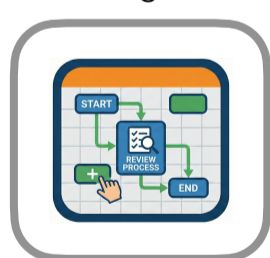


Important Features

- Review processes customization
- Workflow transparency
- Ability of extension
- Explanation for outputs
- Local and remote execution

A node-based review process designer that allows adding new tools and capabilities as "agents."

Dynamic Process Designer

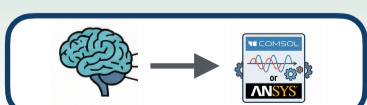


Model-Driven Agentic Process

Freedom of choice in backbone model selection for each component

- + Local and Remote
- + Customized level of reasoning

Calling external tools for "claim verification."



- New components can be added without hard rewiring.
- Nondeterministic question reviewing process

SCAN ME

