Review Paper

International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 13; Issue: 1; January 2026
Website: www.ijrrjournal.com
E-ISSN: 2349-9788; P-ISSN: 2454-2237

Use of Agent-Based AI Applications in Research Institutions

Peter Schlecht[1], Tobias Oberdieck[2], Enrico Moch[3]

[1]CEO, Department of Economics, GrandEdu GmbH, Germany
[2]Chief Executive Officer, GrandEdu GmbH, Germany
[3]Academic Director, Department of Economics, GrandEdu Research School, Germany

Corresponding Author: Dr. Peter Schlecht, Email: Peter.Schecht@GrandEdu.de

DOI: https://doi.org/10.52403/ijrr.20260119

ABSTRACT

Research institutions face increasing pressure to ensure methodological rigour, transparency, and governance compliance under conditions of growing complexity and limited human resources. While generative artificial intelligence is often discussed as a tool for efficiency gains, its use in academic contexts raises fundamental concerns regarding responsibility, decision authority, and the risk of implicit automation of scholarly judgement. This paper addresses these challenges by proposing a governance-aware, agent-based approach to AI-supported research processes. The study develops a structured framework for integrating agent-based AI support into scientific workflows without delegating epistemic or managerial responsibility. Instead of treating AI as a monolithic system, the proposed approach decomposes support functions into specialised agents that perform clearly bounded tasks such as structural checks, methodological consistency analysis, citation validation, and formal compliance review. Human actors retain full control over interpretation, prioritisation, and decision-making at all stages of the research process. Methodologically, the paper follows a design-oriented research approach. It introduces a reference process for AI-supported academic work, a role- and agent-based interaction model, a multi-agent system architecture, and a governance and operating model that ensures transparency, accountability, and institutional control. The applicability of the approach is demonstrated through an illustrative use case covering the development of an exposé, manuscript preparation, Internal Review, and submission readiness. The contribution of the paper is threefold. First, it provides a practically implementable model for AI-supported research workflows that preserves scholarly autonomy. Second, it translates abstract governance principles into concrete organisational and technical design choices. Third, it offers research institutions a transferable blueprint for deploying generative AI as a supportive infrastructure rather than a decision-making authority. The paper concludes by discussing limitations and outlining avenues for future research on institutional AI governance in academia

Keywords: *AI governance, managerial decision support, organisational design, human-in-the-loop systems, digital transformation in research organisations, responsible AI management, process governance, strategic use of artificial intelligence*

INTRODUCTION

Universities and research institutions are increasingly faced with the challenge of ensuring scientific quality, methodological rigour and regulatory compliance under conditions of growing complexity. In particular, structuring processes such as the development of exposés, the selection of suitable research methods, the preparation of manuscripts and Internal Review require a considerable amount of coordination and verification. At the same time, the time resources of researchers, supervisors and reviewers are limited, while the requirements for transparency, traceability and documentation continue to increase (Mintzberg, 1994; Mintzberg et al., 1998). In this context, artificial intelligence is often discussed as a tool for increasing efficiency. In scientific practice, however, there is a risk that support functions and decision-making functions will become mixed up. Generative AI systems in particular produce texts, suggestions or evaluations whose origin, reasoning and limitations are not always immediately apparent. Without clear embedding, there is a risk of methodological inconsistencies, insufficient source verification or an implicit delegation of scientific responsibility to technical systems (Bender et al., 2021; Floridi et al., 2018; Rai et al., 2019).

Against this backdrop, an agent-based approach is gaining in importance. Instead of using AI as a monolithic tool, specialised AI agents with clearly defined roles, tasks and responsibilities are being designed. Each agent supports a narrowly defined sub-process, such as structural review, method comparison, citation consistency or formal journal compliance, without making scientific decisions itself. Decisions remain explicitly with human actors; AI acts as a structuring, reviewing and exploratory assistant (Huang & Rust, 2021; Moch, 2025a; Rai et al., 2019).

This paper examines the use of agent-based AI applications in research institutions using the example of the GrandEdu Research School. The aim is to show how such a system can be implemented in practice without undermining scientific autonomy, accountability and governance. The focus is not on algorithmic performance optimisation, but on designing a comprehensible, auditable and institutionally compatible support and decision-making framework (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019). Methodologically, the work follows a design science research approach according to Hevner et al. (2004). Several conceptual and technical design components are being developed: a process-oriented reference model for the use of agent-based AI throughout the research life cycle, a modular multi-agent architecture with clearly defined roles, interfaces and control points, and a governance and operating model that integrates data protection, responsibilities and organisational control mechanisms (Hevner et al., 2004; Peffers et al., 2007).

The central research question is:

How can agent-based AI applications be implemented in research institutions in such a way that they effectively support scientific work processes without compromising decision-making responsibility, methodological quality and governance?

This paper addresses a structural tension that has become central to the deployment of AI in research organisations. On the one hand, there is a growing demand for scalable support and standardised quality assurance. On the other hand, scientific responsibility cannot be delegated without undermining core principles of academic work. In many practical settings, this tension is not resolved through additional guidelines, but through implicit shifts of judgement to technical systems driven by opacity, convenience or workflow integration. The present work therefore proposes a decision-negative design logic for AI support. AI components are deliberately constrained to preparatory, auditing and structuring functions, while interpretative weighting, prioritisation and approval remain exclusively human. This logic is operationalised through governance-enforced agent orchestration, understood as a process- and role-bound activation of specialised agents with explicit control points, clearly defined functional boundaries and comprehensive traceability. The contribution of the work lies not in the formulation of another governance framework, but in an architecture-embedded implementation of governance requirements that renders responsibility allocation verifiable within the workflow.

The contribution of this work is threefold. First, it presents a practical reference model that systematically integrates agent-based AI into scientific processes. Second, it describes a concrete system and software architecture that takes into account both technical feasibility and regulatory requirements. Third, the paper formulates design principles for research institutions that want to institutionally embed AI-supported assistance systems (Hevner et al., 2004; Peffers et al., 2007; Rai et al., 2019). Against this backdrop, the contribution of this paper must be clearly distinguished from existing AI governance approaches. Existing work on AI governance and human-in-the-loop concepts often address key normative principles such as transparency, accountability and control at an abstract level and repeatedly refer to the difficulty of translating these principles into operational structures (Floridi et al., 2018; Rai et al., 2019; Ransbotham et al., 2017). Corresponding frameworks formulate requirements for the use of AI, but mostly remain at the level of guidelines, catalogues of principles or governance objectives (Floridi et al., 2018; Rai et al., 2019). In contrast, this paper does not start with normative justification, but with the operational translation of these principles. The contribution lies not in the development of another governance framework, but in the systematic translation of existing governance requirements into concrete process logics, role models and technical architectures for scientific work. Thanks to the agent-based structure, governance is not postulated, but implemented technically and organisationally. The work thus closes a gap between the normative AI governance discussion and practically implementable, institutionally controllable AI support in scientific work processes.

CONTEXT AND REQUIREMENTS OF THE GRANDEDU RESEARCH SCHOOL

As a research-oriented institution, the GrandEdu Research School is designed to provide structured support for scientific work throughout the entire research life cycle (Figure 1). The focus is not only on formal qualifications, but also on ensuring methodological quality, argumentative rigour and compliance with publication requirements. The processes considered range from the initial development of topics and exposés to the choice of methods and manuscript preparation to Internal Review, revision and external submission.

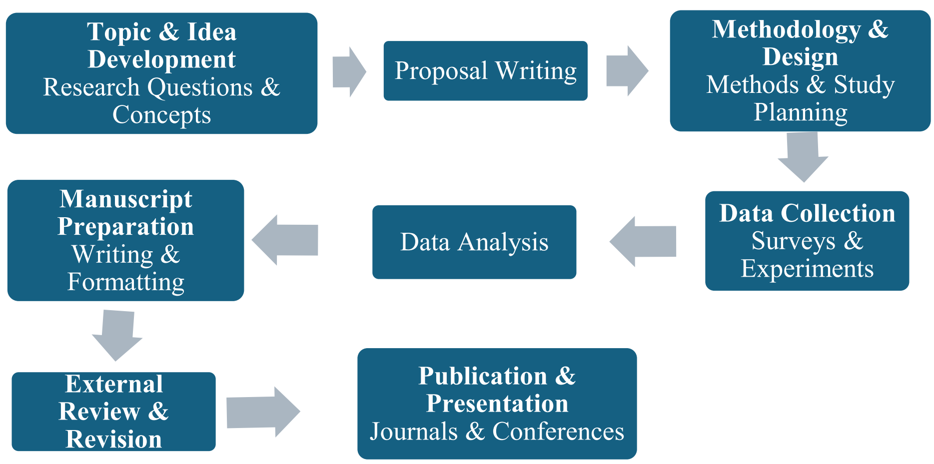

Figure 1: Research life cycle Author's own illustration (for illustrative purposes only).

Research processes and typical bottlenecks

In practice, scientific work processes follow a recurring pattern:
(1) Development of an exposé with research question, theoretical framework and methodological outline,
(2) Preparation of a manuscript,
(3) Internal feedback and revision,
(4) Preparation for submission to a target journal.

Similar bottlenecks repeatedly occur during these phases. Exposés often lack clarity in defining the research question or in matching the question to the method. Manuscripts contain inconsistent lines of argumentation, insufficiently substantiated statements or formal deviations from journal guidelines. Internal Reviews are time-consuming and heavily dependent on individual expertise, resulting in varying quality and limited scalability.

Stakeholders and understanding of roles

Various groups of actors are involved in the processes described: researchers, supervisors, Internal Reviewers and administrative units. What all roles have in common is that they bear responsibility for scientific quality and decision-making authority. In this context, technical systems, including AI-based tools, are not intended to be decision-making systems, but solely as supporting infrastructure (Floridi et al., 2018; Rai et al., 2019). A central design principle of the GrandEdu Research School is therefore the clear separation between support and decision-making. Evaluations, approvals, methodological specifications and publication decisions always remain with human actors. Support systems may prepare, structure, review and suggest alternatives, but may not make independent decisions or implicitly anticipate them (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019).

Quality requirements and institutional standards

Institutional quality requirements encompass several dimensions. Methodological quality emerges from the disciplined integration of several core requirements. At its foundation lies methodological consistency, understood as a transparent and intelligible linkage between the research question, the theoretical framing and the selected methodological approach. Argumentative rigour constitutes a second pillar. Central claims must be developed in a logically coherent manner and anchored in relevant scholarly literature. Formal criteria form a third dimension of quality assurance. Compliance with citation standards, structural conventions and journal-specific requirements is not ancillary, but constitutive of academic work (Hevner et al., 2004; Mintzberg et al., 1998; Peffers et al., 2007). Beyond these dimensions, scientific practice is bound by requirements of transparency and traceability. Revisions, feedback processes and the rationale underlying substantive decisions should remain systematically documented. This documentation serves not only to safeguard quality, but also to enable revision processes and to ensure institutional accountability (Floridi et al., 2018; Rai et al., 2019).

Governance and compliance requirements

The use of technical support systems is subject to additional governance requirements. These include data protection, access controls, role and rights concepts, and the avoidance of opacity in automated suggestions. In the case of generative AI systems in particular, it must be ensured that content is not adopted without being checked and that the origin, purpose and limitations of AI contributions remain recognisable (Bender et al., 2021; Floridi et al., 2018; Rai et al., 2019). For the GrandEdu Research School, this means that any form of AI-supported assistance must be embedded in a clearly defined organisational framework. This framework includes technical controls, organisational rules, and explicit responsibilities. The goal is not maximum automation, but controlled, traceable support (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019).

Derivation of functional requirements for AI-supported assistance

Key functional requirements can be derived from the context described above. AI-supported assistance must be modular in order to address individual subtasks in a targeted manner. It must be explainable and verifiable in order to enable trust and acceptance. In addition, it must be integrable into existing processes without fundamentally changing or displacing them (Floridi et al., 2018; Huang & Rust, 2021; Rai et al., 2019). The ability to clearly separate different support functions is particularly relevant. Structural review, method comparison, citation consistency, and formal compliance are all independent tasks that require different rules, data sources, and control mechanisms. A monolithic AI system is unsuitable for this (Huang & Rust, 2021; Rai et al., 2019). These requirements form the basis for the agent-based AI support approach developed in the following chapters. The division into specialised, clearly defined AI agents is intended to enable a practical, governance-conscious and institutionally compatible implementation (Huang & Rust, 2021; Moch, 2025b; Rai et al., 2019).

THEORETICAL FRAMEWORK AND METHODOLOGICAL APPROACH

This chapter outlines the theoretical framework and methodological approach of the study. The aim is to conceptually classify the use of agent-based AI support in research institutions and to justify it in a methodologically comprehensible manner. The focus is not on the algorithmic performance of individual AI models, but on the design of organisationally compatible support structures (Hevner et al., 2004; Huang & Rust, 2021; Peffers et al., 2007).

Agent-based AI as an organisational design principle

Agent-based AI systems are characterised by the fact that they consist of several specialised, relatively autonomous units, each of which performs clearly defined tasks and interacts with each other in a coordinated manner. In contrast to monolithic AI applications, agent-based approaches allow for the functional decoupling of individual support services and the explicit assignment of responsibilities and boundaries (Huang & Rust, 2021; Rai et al., 2019). This approach is particularly suitable for the scientific context. Research processes consist of clearly distinguishable subtasks such as structuring, methodological fit testing, literature consistency or formal compliance each of which follows different rules and quality criteria. An agent-based structure makes it possible to support these subtasks separately without mixing them in a single system (Hevner et al., 2004; Huang & Rust, 2021; Peffers et al., 2007). It is crucial that the agents are not designed as independent decision-makers. Their function is limited to preparatory, testing and structuring activities. Evaluation, prioritisation and approval remain explicitly with human actors. Agent-based AI is thus understood as an organisational design principle, not as a substitute for scientific judgement (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019).

Human-in-the-loop and responsibility

The human-in-the-loop principle forms an essential theoretical foundation of this work. This principle states that AI systems in decision-relevant contexts must be designed in such a way that human actors retain control, insight and decision-making authority at all times. In scientific work, this applies in particular to methodological specifications, argumentative evaluations and publication decisions (Floridi et al., 2018; Rai et al., 2019). The distinction between support and decision-making is not only technically relevant, but also normatively relevant. Scientific responsibility cannot be delegated. AI-supported suggestions can prepare or structure decision-making processes, but must not result in de facto decision automation. This distinction forms a central guideline for the design of the models and system components described below (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019).

Capability perspective on AI support

A capability-oriented perspective is taken for the conceptual classification of AI support. AI is not viewed as an isolated tool, but rather as an organisational capability that enables certain achievements without independently pursuing goals or making decisions. This perspective allows AI support to be described functionally and systematically embedded in existing processes (Huang & Rust, 2021; Rai et al., 2019). In the context of the GrandEdu Research School, this means that AI agents function as specialised support capabilities: they increase the consistency, transparency and efficiency of scientific work processes without changing their normative foundations. The capability perspective thus forms a bridge between technical implementation and organisational responsibility (Floridi et al., 2018; Huang & Rust, 2021; Rai et al., 2019).

Methodological approach

Methodologically, the work follows a design-oriented research approach. The starting point is a clearly defined practical problem: the limited scalability and traceability of scientific support and review processes amid increasing quality and governance requirements. The aim is not to empirically measure effects, but to develop, justify and demonstrate viable design approaches (Hevner et al., 2004; Peffers et al., 2007). The procedure is divided into four steps. First, the organisational context is analysed and key requirements are derived. Second, conceptual models are developed that translate these requirements into structured process, role and system descriptions. Third, a concrete technical implementation is outlined that shows how agent-based AI support can be realised in practice. Fourth, a demonstration is carried out using an exemplary use case to highlight the applicability and limitations of the approach (Hevner et al., 2004; Peffers et al., 2007).

Evaluation logic and scope

The models developed are not evaluated in terms of statistical evidence of effectiveness, but rather according to qualitative criteria. These include, in particular, traceability, organisational connectivity, governance compliance and practical feasibility. The benefits of the approach are evident in the structured support of scientific processes, the reduction of avoidable inconsistencies, and the improved transparency of decisions and revisions (Hevner et al., 2004; Peffers et al., 2007; Rai et al., 2019). The scope of this work is deliberately limited. The concepts presented are aimed at research-oriented institutions with structured qualification and publication processes. No statements are made about empirical effects or performance improvements. Instead, the focus is on providing a robust, transferable design framework (Hevner et al., 2004; Peffers et al., 2007).

REFERENCE PROCESS FOR AGENT-BASED AI SUPPORT IN SCIENTIFIC WORK PROCESSES

This chapter describes the reference process that structures the use of agent-based AI support throughout the scientific work cycle. The aim is to define a clearly comprehensible process in which human decision-making authority, AI support and governance requirements are systematically coordinated. The process should not be understood as a rigid sequence, but rather as a modular framework that can be adapted to institutional needs (Floridi et al., 2018; Hevner et al., 2004; Rai et al., 2019).

Basic structure of the reference process

The reference process is based on typical phases of scientific work in research institutions:

(1) Initialisation and clarification of context,

(2) Structuring and preliminary methodological review,

(3) Content development,

(4) Internal Review and revision,

(5) Approval and external submission.

Each phase is characterised by clearly defined goals, inputs, outputs and responsibilities. AI support is used exclusively where structuring, reviewing or exploratory activities are required. Decision-making and approval points remain consistently reserved for human actors.

Phase 1: Initialisation and clarification of context

The process begins with the formal and content-related initialisation of a research project. Inputs include, for example, a topic proposal, an initial exposé or an existing manuscript. The aim of this phase is to clearly define the context, objectives and framework conditions (Hevner et al., 2004; Peffers et al., 2007). AI support takes on preparatory tasks here: structural review of the input document, identification of missing core elements, and formal classification (e.g., document type, discipline, target journal). The results are output as structured notes without prioritising evaluations or recommendations. The decision on relevance and further action is made by the responsible human actor (Floridi et al., 2018; Rai et al., 2019).

Phase 2: Structuring and methodological preliminary review

The second phase focuses on the fit between the research question, theoretical framework and methodological orientation. Inconsistencies often arise here that are difficult to identify in the early stages. Specialised AI agents support this phase by comparing explicit and implicit assumptions, identifying methodological inconsistencies and presenting possible alternatives in a structured manner. It is important that these suggestions are not formulated as recommendations with a decision-making character, but as options with explanatory approaches. The selection or rejection remains entirely up to humans (Floridi et al., 2018; Hevner et al., 2004; Peffers et al., 2007; Rai et al., 2019).

Phase 3: Content development and consistency check

The third phase involves the actual drafting of the scientific text. Here, the focus is on argumentative rigour, coherence between sections and the correct embedding of literature. AI support is used in this phase to check consistency: repetitions, argumentative leaps, unsubstantiated statements or formal deviations are identified and transparently reported. The AI does not interfere with the content design, but provides structured review notes that can be checked in a targeted manner (Hevner et al., 2004; Mintzberg et al., 1998; Rai et al., 2019).

Phase 4: Internal Review and revision

An Internal Review is conducted before external submission. This phase is particularly resource-intensive, as it places high demands on experience, diligence and time. Agent-based AI support can take over standardised checks here, such as compliance with formal guidelines, citation rules or documentation requirements. In addition, critical questions that typically arise in peer reviews can be simulated. This simulation does not replace a review, but serves to prepare and structure human feedback (Floridi et al., 2018; Hevner et al., 2004; Rai et al., 2019).

Phase 5: Approval and external submission

The final phase is characterised by clear decision and approval points. Here, it is determined whether a document will be submitted, revised or deferred. In this phase, the use of AI support is deliberately minimal. Technical systems can only provide checklists or confirm formal completeness. The approval decision itself is explicitly human and is documented (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019).

Cross-section: control, documentation and traceability

Support contributions, notes and revisions are documented across all phases. The aim is not monitoring, but traceability. Each AI support is assigned to a clearly defined purpose, phase and responsible role. This prevents AI contributions from implicitly exerting decision-making power or evading control. The reference process thus forms the organisational and backbone for the use of agent-based AI support. It creates the conditions for integrating technical systems into scientific work processes in a controlled, transparent and responsible manner (Floridi et al., 2018; Hevner et al., 2004; Rai et al., 2019).

ROLE AND AGENT MODEL FOR AI-SUPPORTED RESEARCH SUPPORT

The role and agent model forms the operational basis for the reference process described in Chapter 4. It serves to clearly structure the interaction between human actors and AI-supported assistance and to clearly assign responsibilities. The aim is to embed AI support in such a way that it complements scientific work processes without assuming decision-making powers, evaluation authority or responsibility (Floridi et al., 2018; Rai et al., 2019). The central starting point of the model is the strict functional separation between human roles and technical support units. Human actors, in particular researchers, supervisors, Internal Reviewers and administrative roles, bear full responsibility for scientific decisions, methodological specifications and approvals. AI systems, on the other hand, are designed exclusively as supporting entities whose tasks are clearly limited, comprehensible and verifiable (Floridi et al., 2018; Rai et al., 2019). The human roles are clearly differentiated according to their respective areas of responsibility. Researchers are responsible for the content, argumentation and scientific integrity of the work. Supervisors are responsible for checking methodological consistency, classifying the project and making strategic decisions in the research process. Internal Reviewers are responsible for quality assurance assessments prior to external submission. Administrative roles ensure compliance with institutional standards, formal requirements and regulatory requirements. This role structure remains fully intact regardless of the use of technical support (Hevner et al., 2004; Peffers et al., 2007). Against this background, AI agents are not understood as autonomous actors, but as specialised support units, each covering a narrowly defined area of responsibility. Each agent is functionally limited to a clearly defined activity and generates structured notes, test results or alternatives without evaluating, prioritising or translating them into decisions. This prevents AI support from implicitly taking on a decision-making character (Floridi et al., 2018; Rai et al., 2019).

A central component of the model is an agent for structure and completeness checking, which examines incoming documents for formal coherence, missing core elements or inconsistent structures. This agent does not interfere with content, but only provides structured information on formal quality. In addition, a method and fit agent support the comparison between the research question, theoretical framework and methodological orientation. It identifies potential inconsistencies and presents alternative methodological options without making recommendations or weightings (Hevner et al., 2004; Peffers et al., 2007).

Another agent focuses on argumentative consistency. It analyses the logical structure of a text, identifies inconsistencies, redundancies or unsubstantiated statements and makes these transparent. The assessment of relevance and the decision on possible changes remain entirely with the human being. This interaction is complemented by a literature and citation agent that checks formal citation rules, consistency of references and obvious gaps in the literature review without generating new sources or setting content priorities (Mintzberg et al., 1998; Rai et al., 2019).

In addition, a formalities and compliance agent is used to check journal-specific and institutional requirements and point out deviations. The aim is to reduce formal rejection risks without evaluating content. In later stages of the process, a review simulation agent can be used to formulate typical critical queries from peer review procedures. These queries are explicitly marked as simulations and serve to prepare human reviews, not to replace them (Floridi et al., 2018; Hevner et al., 2004; Rai et al., 2019). The interaction of the AI agents is not autonomous, but takes place via an orchestrated process that is initiated and controlled by human roles. Agents do not work hierarchically and their results are not automatically aggregated or prioritised. Instead, all information is presented in parallel and transparently so that human actors can consciously select, weight or reject it. This orchestration prevents agent-based results from reinforcing each other or consolidating in an uncontrolled manner (Floridi et al., 2018; Rai et al., 2019).

A central element of the model is the explicit assignment of responsibility and traceability. Each use of an AI agent is assigned to a specific process step, a defined purpose and a responsible human role. Results are documented in versions and can be checked or revised at any time. This ensures that the origin of each piece of support remains traceable and that a clear distinction is maintained between human decision-making and technical assistance (Floridi et al., 2018; Hevner et al., 2004; Rai et al., 2019).

Overall, the role and agent model ensures that AI-supported assistance does not act as a substitute for scientific judgement, but rather as a structuring, reviewing and relieving infrastructure. It creates the organisational conditions for integrating agent-based AI into scientific work processes in a controlled, responsible and institutionally compatible manner (Floridi et al., 2018; Huang & Rust, 2021; Rai et al., 2019).

MULTI-AGENT ARCHITECTURE AND CONCRETE SOFTWARE SOLUTION

This chapter describes the concrete technical implementation of agent-based AI support and translates the previously developed organisational framework into an operationally comprehensible system architecture. The aim is to present an implementation that is technically feasible, institutionally controllable and governance-compliant. The description deliberately goes beyond a purely conceptual sketch and reveals the central technical and organisational mechanisms without delving into code or product details (Floridi et al., 2018; Hevner et al., 2004; Rai et al., 2019).

The starting point for implementation is the separation between the scientific decision-making level and the technical support level. All binding decisions remain at the human level. The technical architecture is designed in such a way that AI components cannot develop independent process logic, but only respond to explicit human triggers (Floridi et al., 2018; Moch, 2025a; Rai et al., 2019).

The operational data flow always begins with conscious initiation by a human role. A document is not processed automatically, but is actively entered into the system and assigned to a defined process step. This assignment determines which support functions are permitted. On this basis, the orchestration layer generates a limited request to one or more specialised AI agents. Each agent receives only the document parts necessary for its task and a clearly defined analysis assignment. There is no free exploration or independent context enrichment (Floridi et al., 2018; Rai et al., 2019).

Orchestration acts as a central control authority. It does not decide on content, but on responsibilities. It ensures that agents are only activated individually, not recursively and not in self-reinforcing loops. Results are not merged or evaluated, but returned separately to the human level. This prevents technical systems from generating implicit prioritizations or decision proposals (Floridi et al., 2018; Rai et al., 2019). Each AI agent is implemented in a functionally minimalist manner. For example, a method agent only checks the formal fit between the research question, theoretical framework and methodological orientation without making any recommendations. An argumentation agent identifies logical inconsistencies or unsubstantiated statements without interpreting or weighting them. A citation agent checks the formal consistency of existing references without generating new sources. This functional limitation is technically ensured by allowing each agent to generate only predefined output types, such as structured notes or audit marks (Floridi et al., 2018; Huang & Rust, 2021; Rai et al., 2019).

One element of the implementation is the logging and traceability concept. Every interaction with an AI agent is systematically documented. The triggering process step, the responsible human role, the purpose of the request, the document context used and the output generated are all stored. These logs are not used for performance monitoring, but for institutional auditability. They make it possible to reconstruct retrospectively what technical support was used at what stage and on what basis human decisions were made. The technical architecture also provides for a clear separation between document storage and model use. Scientific content remains in a controlled document environment with versioning and access restrictions. AI-models do not access this data persistently, but are given temporary, purpose-specific contexts. Once a processing step is complete, the context is discarded. This reduces the risk of uncontrolled data use and supports compliance with data protection requirements. For the specific software implementation, a service-based architecture is recommended, in which language models are integrated via secure interfaces.

These can be operated both cloud-based and on-premise. The decisive factor is not the provider, but the ability to control access rights, logging and orchestration independently of the model provider. The architecture is therefore deliberately designed to be model-agnostic and allows individual components to be exchanged without changing the governance structure. The implementation contains explicit control points where AI support ends. A human decision is required before any further processing, especially before internal approvals or external submissions. Technical systems can only check for formal completeness in these phases, but cannot evaluate content. This limitation is not only organisational, but also technically anchored, with certain agents being deactivated for certain process phases. Overall, the operational implementation shows that agent-based AI support can only be used responsibly if technical architecture and organisational governance are intertwined. The solution presented is not designed to achieve maximum automation, but rather maximum control, traceability and institutional connectivity. It thus meets the requirements of a scientific environment in which responsibility and judgement cannot be delegated.

APPLICATION OF MULTI-AGENT ARCHITECTURE IN AN EXEMPLARY RESEARCH PROCESS

This chapter demonstrates the operational application of the multi-agent architecture described in Chapter 6 using an exemplary research process at the GrandEdu Research School. The aim is to highlight not only the process itself, but also the specific implementation logic used to embed AI support in scientific work processes in a controlled, purpose-specific and governance-compliant manner. The use case under consideration begins with the submission of an exposé by a researcher. The submission is made via an institutional document environment with versioning and role-based access control. At this point, it is already determined that the document is part of a formal research process and is therefore fundamentally eligible for AI-supported assistance. However, processing does not take place automatically. The use of AI support is explicitly approved by the researcher or the responsible supervisor and assigned to a specific process step.

This approval activates the orchestration layer. It does not interpret the content of the document, but only the process context. On this basis, a decision is made as to which AI agents are permissible for this process step. For the exposé phase, these are typically the structure and completeness agent as well as the method and suitability agent. Other agents, such as those for formal journal compliance, remain technically deactivated in this phase. The structure and completeness agent receives only the formal outline and defined core elements of the exposé. It checks whether all the necessary components are present and provides structured feedback, for example on missing sections or inconsistent structure. At the same time, the method and fit agent analyse the explicitly formulated research question, the theoretical framework and the planned methodology. It identifies potential tensions or inconsistencies and presents alternative methodological options without making recommendations. Both agents work in isolation and generate separate outputs.

The results are returned to the researcher via the orchestration layer. At this point, the technical effect ends. There is no aggregation, weighting or prioritisation of the suggestions. The researcher decides independently which aspects are relevant and whether revisions should be made. This decision is documented, as is the use of the AI agents themselves.

After the exposé has been revised, a manuscript is created and re-entered into the document environment. Here, too, the next use of AI support is explicitly approved. For this phase, the orchestration activates the argumentation and consistency agents as well as the literature and citation agents. Both agents receive only those parts of the document that are necessary for their respective tasks. A complete context transfer does not take place.

The argumentation agent identifies logical inconsistencies, redundancies or unsubstantiated statements and returns them as structured audit notes. The citation agent checks formal citation rules and the consistency of existing references. It does not generate new sources or evaluate content relevance. Here, too, the results are output separately and not automatically integrated into the text.

Before Internal Review, the formalities and compliance agent is also activated. This checks compliance with institutional standards and journal-specific requirements. The use of this agent is technically limited to this process phase. Earlier or later activations are excluded in order to avoid a creeping formalisation of creative phases. Optionally, a review simulation agent can be used in the review preparation phase. It is activated only at the explicit request of a human actor. The queries generated are marked as simulations and are treated organisationally not as evaluations but as preparatory material. Automatic derivation of decisions from these results is technically impossible.

Every interaction with an AI agent is logged throughout the entire process. The process step, the triggering role, the purpose of the support, the context provided and the outputs generated are recorded. These logs are auditable and enable subsequent reconstruction of the support process without themselves having any decision-making character. Approval for external submission marks the final checkpoint. At this stage, AI support is technically limited to formal completeness checks. The final decision on submission is made and documented exclusively by human actors. This concludes the technical support process.

The demonstration shows that the multi-agent architecture is not only conceptually viable, but also operationally feasible. The decisive factor here is not the performance of individual AI agents, but the consistent coupling of technical support to process context, role responsibility and explicit control points. The use case illustrates how AI support can be used effectively without delegating scientific responsibility or automating decision-making logic.

EVALUATION

The evaluation confirms that the proposed architecture consistently implements the decision-negative design logic by restricting AI components to preparatory and auditing functions, while excluding interpretative or approving roles. The evaluation of the developed artefact follows the logic of design science research and does not aim at statistical generalisability, but rather at analytical traceability, theoretical connectivity and practical resilience. Accordingly, the artefact is tested against defined evaluation criteria derived from the problem definition, the theoretical framework and the normative requirements of scientific governance. The central evaluation criterion is the traceability of scientific decisions. The artefact consistently separates AI-supported analysis from human decision-making. Agents are functionally limited to clearly defined support tasks, such as structural testing, method comparison or formal consistency checks. There is no aggregation, weighting or prioritisation of content by the AI. Compared to the unstructured use of generative AI, this separation significantly increases the transparency of the process of creating scientific texts. Closely related to this is the preservation of scientific responsibility. The artefact implements the human-in-the-loop principle not only declaratively, but also procedurally. All normative, interpretative and decision-relevant steps remain explicitly with the responsible researchers. AI outputs are not decisive in nature, but serve exclusively as hints or prompts for review. This prevents responsibility from being implicitly shifted to technical systems.

Another key evaluation criterion concerns governance and liability clarity. Through the role-based activation of individual agents and the complete logging of all interactions, the use of AI can be clearly assigned within the organisation. Each AI contribution can be assigned to a specific functional context, thereby fulfilling the requirements for auditability and institutional control. A comparison with the fundamental governance requirements of scientific organisations shows that the artefact particularly fulfils requirements for auditability and accountability. In terms of error and hallucination resilience, the artefact addresses key risks of generative AI. The deliberate limitation of context, the exclusion of recursive agent activations, and the lack of authority of AI-generated sources reduce the risk of systematic misjudgements. Nevertheless, the complete elimination of such risks cannot be guaranteed. The evaluation therefore shows that the artefact significantly limits these risks, but does not completely eliminate them. This corresponds to realistic expectations of socio-technical systems in a scientific context. Revision and auditability are another key evaluation criterion. By logging all agent activations and outputs, AI contributions can be verified ex post. This creates a prerequisite for systematically supporting scientific quality assurance, Internal Reviews or external audits. Finally, the practical applicability of the artefact in everyday research is evaluated. The modular structure of the model allows it to be integrated into existing research and publication processes without fundamentally changing them. The demonstration in the use case shows that the agent-based approach does not create additional decision-making levels, but rather complements existing processes in a structured manner.

In summary, the evaluation shows that the developed artefact meets the central requirements for governance-compliant, responsible use of agent-based AI in research institutions. In particular, the evaluation confirms the traceability, accountability and auditability of the approach. At the same time, it becomes clear that the contribution of the artefact lies not in the automation of scientific decisions, but in the structured support of human judgement. The evaluation therefore demonstrates analytical validity and governance robustness of the artefact, while empirical effectiveness is deliberately left for future research.

DISCUSSION

The results of this work show that agent-based AI support can make a substantial contribution to scientific work processes if it is not only conceptually constrained, but also implemented in a technically and organisationally controlled manner. The proposed architecture and its operational application demonstrate that governance should not be understood as an abstract principle, but as an integral property of system design. A central point of discussion concerns the added value of the proposed approach in comparison to purely conceptual governance frameworks. While many existing works address governance, human-in-the-loop principles or responsibility at a normative level, they often leave the resolution of responsibility conflicts to organisational practice. The decision-negative design logic developed in this work responds to this limitation by structurally excluding AI systems from interpretative, prioritising and approving functions. Governance is therefore not enforced ex post through guidelines or oversight mechanisms, but ex ante through process-bound activation rules, technical constraints and role-specific control points. In this way, governance is transformed from a declarative requirement into an operational characteristic of the system itself.

The operational implementation further illustrates that the benefit of agent-based AI support lies less in automation than in the structured relief of human actors. The functional specialisation of agents and the deliberate avoidance of aggregation, prioritisation or evaluation address typical risks associated with generative systems without negating their supportive potential. The architecture thus enforces a clear separation between advice and decision-making, not only at the organisational level, but also at the technical level. Another relevant aspect concerns the institutional connectivity of the approach. The demonstrated implementation logic shows that the use of AI support does not necessarily lead to a technologisation of scientific practice. On the contrary, the requirement for explicit approvals, documented purposes and clearly defined control points promotes reflection on existing processes and responsibilities. In this sense, the architecture does not merely support organisational practice, but actively structures it. At the same time, the effectiveness of the approach is closely tied to the organisational maturity of the respective research institution. Implementation presupposes that processes, roles and quality criteria are formalised at least to a basic degree. Where such structures are absent, agent-based AI support cannot fully unfold its strengths. This dependency does not represent a technical shortcoming, but rather underlines that AI governance cannot be meaningfully addressed in isolation from organisational framework conditions.

The discussion further highlights a deliberate distinction from approaches that position AI systems as epistemic co-authors or implicit decision-makers. The architecture proposed here explicitly contradicts such narratives by systematically preventing any form of autonomy. While this normative positioning limits the scope of application, it simultaneously strengthens scientific integrity and accountability. The contribution of the work therefore lies less in maximising technical performance than in minimising epistemic and organisational risks.

Finally, the close linkage between architecture and use case illustrates that implementation itself constitutes a moment of insight. The need to technically operationalise governance requirements forces explicit decisions regarding responsibilities, boundaries and control mechanisms. These decisions are not neutral, but exert a lasting influence on institutional practice. The central scientific contribution of the work thus lies precisely in the translation of normative principles into operational structures.

Overall, the discussion underscores that agent-based AI support can only be used responsibly when implementation and governance are inseparably connected. The work demonstrates that such a connection can be realised in practice and offers a robust and transferable approach for research institutions that seek to deploy AI not as a substitute for scientific judgement, but as a controlled and accountable infrastructure.

LIMITATIONS OF THE WORK

This paper has several deliberate and conceptually justified limitations that are central to the classification of its results. These limitations are not weaknesses in the strict sense, but result from the chosen objectives, methodological approach and normative orientation of the paper. Their explicit and designation serves the purpose of scientific transparency and clear delimitation of the scope of application. A first limitation concerns the design-oriented nature of the work. The contribution does not aim to empirically measure the effectiveness of agent-based AI systems, but rather to develop a robust organisational and technical design framework. Accordingly, no quantitative statements are made about efficiency gains, quality improvements or productivity effects. The demonstration of the approach serves to illustrate practical feasibility, not statistical validation. Statements about empirical effects are reserved for future research.

Closely related to this is a second limitation regarding the evaluation of technical performance. The paper makes no statements about the quality, accuracy or reliability of specific AI models. The proposed architecture is deliberately model-agnostic. It assumes that technical systems can be prone to errors and addresses this uncertainty through organisational limitations, control and human responsibility. Statements about the progress or shortcomings of individual model generations are beyond the scope of this paper. A further limitation arises from the dependence on institutional framework conditions. The approach developed is aimed at research institutions with clearly defined processes, roles and quality standards. Its effectiveness is limited in highly informal, individualised or unstructured research environments. The architecture only demonstrates its benefits where organisational rules can be made explicit and enforced. Transferability is therefore less a technical issue than an organisational one.

Furthermore, the approach has a clear normative position. The work explicitly assumes that scientific responsibility cannot be delegated and that AI systems have no epistemic authority. This normative assumption deliberately limits the scope of application of agent-based AI support. Approaches that aim at further automation of scientific decisions are not addressed. This limitation is also an expression of a conscious scientific ethical stance. Another limitation concerns interdisciplinarity. Although the approach is fundamentally generic in nature, its elaboration is based on typical processes in research-oriented qualification and publication contexts. Disciplines with significantly different publication cultures, such as artistic research or highly experimental laboratory research, are only covered indirectly. Full adaptation to such contexts requires additional specification.

Finally, it should be noted that the work does not make any statements about acceptance by individual stakeholder groups. The actual use of agent-based AI support depends largely on trust, experience and institutional culture. These social and cultural factors are taken into account conceptually, but not examined empirically. Here, too, there is a deliberate limitation of the scope of the investigation.

In summary, it can be said that the limitations of the work are closely linked to its objectives. The paper does not claim to be technologically complete or empirically generalisable. Instead, it offers a clearly defined, normatively reflected and organisationally compatible design framework. The explicit identification of the limitations helps to place the paper in a realistic context and makes it usable as a basis for further empirical, technical or institutional research.

CONCLUSION AND OUTLOOK

This work has shown that agent-based AI support can make a substantial contribution to scientific work processes when it is consistently designed as organisational infrastructure rather than as an epistemic substitute. The point of departure was the increasing complexity of scientific work under rising quality and governance requirements. Against this background, a design framework was developed that systematically constrains technical support, renders its operation transparent and embeds it institutionally. The focus of the paper is not the automation of scientific decisions, but their deliberate preparation, structuring and safeguarding. The proposed multi-agent architecture enables a clear separation of support functions and their targeted use without delegating decision-making authority or responsibility. In this way, the work demonstrates how AI can be productively employed in scientific contexts without undermining the normative foundations of scientific practice.

The results must be interpreted in light of the limitations discussed above. The approach does not claim to provide empirical evidence of effectiveness or technological completeness. Instead, it highlights that the benefits of AI support depend less on model performance than on institutional design. Governance structures, role clarity and documented accountability emerge as central prerequisites for responsible use. At the same time, the analysis makes clear that agent-based AI support is not a universal solution. Its added value is particularly evident in structured research environments characterised by defined processes, quality standards and review mechanisms. In less formalised contexts, the approach reaches its limits. This limitation does not represent a deficiency, but reflects conceptual clarity and normative consistency.

Beyond the specific research context examined, the proposed architecture defines a transferable design logic for decision-sensitive domains in which AI support must remain strictly subordinate to human responsibility, such as peer review, regulatory assessment or quality assurance processes.

The outlook points to several directions for future work. At the institutional level, the approach offers a basis for more systematic anchoring of quality and governance structures while preserving individual scientific autonomy. At the conceptual level, further research may explore how agent-based support profiles can be differentiated across disciplinary contexts. At the methodological level, empirical studies are needed to examine acceptance, practical use and long-term effects of such systems.

Overall, the paper provides a realistic, responsible and transferable framework for AI support in scientific work processes. It demonstrates that technological innovation and institutional responsibility are not opposing forces, but mutually dependent conditions for maintaining scientific quality and integrity.

Declaration by Authors

Acknowledgement: None

Source of Funding: None

Conflict of Interest: No conflicts of interest declared.

About the Author

Dr. Peter Schlecht is Managing Director of GrandEdu GmbH, a leading continuing education institution that specialises in preparing professionals for degrees such as the Bachelor Professional, Master Professional and Master of Business Administration (MBA). In his role at GrandEdu, he is committed to developing innovative educational strategies that bridge the gap between vocational and academic education and contribute to the sustainable qualification of skilled workers in Germany. During his doctorate, Dr. Schlecht worked as a research assistant at the Department of Economics at the Niederrhein University of Applied Sciences. His research focused on the intersection of tax law, financial education and labour market policy, which

today shapes his strategic work at GrandEdu.
https://orcid.org/0009-0002-5371-750X

Dr. Tobias Oberdieck is Managing Director of GrandEdu GmbH, a leading further education institution that specialises in preparing professionals for degrees such as the Bachelor Professional, Master Professional and Master of Business Administration [MBA]. In his role at GrandEdu, he drives innovative educational strategies to actively support the recruitment of skilled labour in Germany. During his doctorate, Dr. Oberdieck worked as a research assistant at the Department of Business Administration and Economics at the Niederrhein University of Applied Sciences in Germany. During this time, he deepened his expertise in the areas of education management, labour market research and economic education models, which he now brings to his work at GrandEdu. His research focuses on the interface between continuing vocational training, academic education and the challenges of the modern labour market.
https://orcid.org/0009-0002-2017-4991

Enrico Moch holds a doctorate in economics

and teaches as a lecturer at various universities, including the DHBW Ravensburg and as an assistant professor at the IIC University of Technology. As Academic Director of the GrandEdu Research School in Germany, he combines academic excellence with practice-oriented teaching. His research interests include the Austrian School of Economics, AI governance, technical data protection and the institutional governance of digital platforms. Dr Moch publishes regularly in specialist journals, is involved in interdisciplinary book projects and runs the podcast "GrandEdu Research School - On the trail of the economy". He is also actively involved in academic peer review and is committed to the transfer of knowledge between research and practice in order to provide long-term support for current economic and social issues.
https://orcid.org/0009-0005-4722-0961

REFERENCES

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922
2. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
3. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105. https://doi.org/10.2307/25148625
4. Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9
5. Mintzberg, H. (1994). The rise and fall of strategic planning. Free Press.
6. Mintzberg, H., Ahlstrand, B., & Lampel, J. (1998). Strategy safari: A guided tour through the wilds of strategic management. Free Press.
7. Moch, E. (2025a). Algorithms as judges: Constitutional limits to automated decision-making in public administration. East African Journal of Law and Ethics, 8(1), 178–188. https://doi.org/10.37284/eajle.8.1.3044
8. Moch, E. (2025b). Beyond efficiency: Artificial intelligence as a driver for resilient market organisations. East African Journal of Interdisciplinary Studies, 8(1), 441–453. https://doi.org/10.37284/eajis.8.1.3178

9. Moch, E. (2025c). Culture of shared intelligence: Perspectives on a coevolutionary future of humans, machines and society. International Journal of Research and Review, 12(10), 202–214. https://doi.org/10.52403/ijrr.20251020
10. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302
11. Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Toward human–AI hybrid governance. MIS Quarterly, 43(1), iii–ix.
12. Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. MIT Sloan Management Review in collaboration with The Boston Consulting Group

How to cite this article: Peter Schlecht, Tobias Oberdieck, Enrico Moch. Use of agent-based AI applications in research institutions. *International Journal of Research and Review*. 2026; 13(1): 202-218. DOI: *https://doi.org/10.52403/ijrr.20260119*

******