DHBW
Duale Hochschule
Baden-Württemberg
Ravensburg

Agentic AI Architecture Design

Workflow Automation in Compliance with Operational Excellence

Bachelor's Thesis

Wirtschaftsinformatik—Business Engineering
Duale Hochschule Baden-Württemberg
Ravensburg

Francisco Rodriguez Müller, on September 21, 2025
Mat. Nr.: 2775857, Course: RV-WWIBE122
Supervisor: Prof. Dr. Paul Kirchberg

Declaration of Authenticity

I hereby declare that I have written the present bachelor's thesis independently and that I have not used any sources or aids other than those indicated. All passages that were taken from published or unpublished works are clearly marked as such. This thesis has not been submitted to any other examination authority in the same or a similar form.

Thesis Title:
Agentic AI Architecture Design
Workflow Automation in Compliance with Operational Excellence

Ravensburg, September 21, 2025

(Place and Date) *(Signature)*

Acknowledgments

I would like to thank the *All for One Group SE* for providing the opportunity and resources to write this thesis. I also extend my gratitude to my academic supervisor, *Prof. Dr. Paul Kirchberg* for his guidance throughout the process; my professional advisor, *Emily Celen* for helping me with the scope and realistic expectations for this work; and my direct supervisor and friend *Ralph Thomaßen* for his wise advice and relentless support throughout my academic journey.

Abstract

[Placeholder, will write it once the thesis is finished.]

Listings

Contents

Abbreviations

A2A	Agent to Agent
DSR	Design Science Research
GaaS	Governance as a Service
GenAI	Generative Artificial Intelligence
IPA	Intelligent Process Automation
ISD	Information Systems Design
LLM	Large Language Model
MAS	Multi-Agent System
MBSE	Model-Based Systems Engineering
MCP	Model Context Protocol
OpEx	Operational Excellence
QCA	Qualitative Content Analysis
RE	Requirements Engineering
ReAct	Reason & Action
RPA	Robotic Process Automation
SLA	Service Level Agreement
SysML	Systems Modeling Language
WfMS	Workflow Management Systems

1 Introduction

Organizations across industries continue to face persistent challenges in achieving operational excellence (OpEx). Fragmented processes, manual interventions, and inconsistent data quality undermine efficiency and decision-making. Legacy workflows and siloed systems exacerbate these inefficiencies, while traditional automation approaches often lack the adaptability needed in dynamic business environments. For companies, this translates into slower response times, higher compliance risks, and limited scalability—issues that directly threaten competitiveness.

Agentic AI, building on the advances of generative artificial intelligence (GenAI), opens new possibilities to extend automation beyond deterministic scripts. While GenAI provides the cognitive and generative capabilities, agentic AI leverages these to create adaptive, tool-using agents that can plan, act, and coordinate—thereby supporting governance, decision quality, and organizational agility. Despite this potential, both practice and academic literature lack structured strategies and conceptual frameworks for embedding such agentic capabilities into operational workflows in a scalable and value-driven way. This gap motivates the present research.

In this context, multi-agent systems (MAS) can serve as a reference architecture for integrating GenAI-enabled agentic AI into enterprise workflow automation. The central research question is:

How can a MAS architecture be designed to integrate GenAI capabilities into workflow automation, in order to enhance agility, compliance, and decision quality to achieve OpEx?

To answer this question, the study addresses the following sub-questions:

- *Which design requirements are necessary to align a multi-agent architecture with the goals of OpEx?*
- *How should a MAS be architected to fulfill these requirements?*
- *Under which conditions is deploying a generative multi-agent architecture justified over traditional automation approaches?*

Methodologically, the thesis applies Design Science Research (DSR) to develop a conceptual reference architecture. The approach synthesizes requirements from academic literature and OpEx principles, models agent roles and interactions, and derives applicability conditions for real-world deployment.

The core contribution of this work is a conceptual design of a MAS that leverages GenAI to support OpEx in enterprise workflows. Specifically, it delivers:

- *A structured synthesis of system requirements derived from academic literature and OpEx principles.*
- *A conceptual architecture detailing agent roles, interactions, and integration points.*
- *A set of applicability conditions and design considerations to guide future deployment and evaluation of generative multi-agent architectures in practice.*

The scope is limited to conceptual design; formal evaluation and technical implementation are proposed as future work. Although the architecture is designed to remain industry-agnostic, a use case from the financial services sector is introduced to illustrate how the conceptual model can be instantiated in a regulated, legacy-intensive environment.

After this introduction in Section 1, the thesis is structured as follows: Section 2 outlines the research methodology, including the use of Design Science Research (DSR) and supporting methods. Section 3 presents a literature review on operational excellence, workflow automation, and agentic AI. Section 4 details the synthesis and modeling of requirements. Section 5 develops the conceptual multi-agent architecture. Section 6 discusses the applicability of MAS in workflow automation use cases, and Section 7 concludes with reflections and directions for future research.

2 Methodology

This thesis applies DSR methodology to create a conceptual artifact—a multi-agent architecture for workflow automation. Practically, the approach unfolded in three steps: (1) *reviewing the literature* on OpEx, workflow automation, and agentic AI; (2) *deriving and structuring requirements* from literature and case material into a requirements model; and (3) *designing a conceptual system architecture* using Systems Modeling Language (SysML).

Supporting methods included Mayring-style qualitative content analysis (QCA) for the review, requirements engineering (RE) and systems analysis for the requirements model, and information systems design (ISD) to structure the architecture and ensure requirement-to-design traceability, supported by SysML modeling practices from Model-Based Systems Engineering (MBSE). Within DSR, the work focuses on

problem identification, objective definition, and conceptual design, while instantiation/demonstration and formal evaluation are out of scope given the bachelor-thesis format and resource constraints. This scoping maintains methodological rigor while keeping the contribution focused: a well-argued reference architecture ready for subsequent implementation and empirical evaluation.

2.1 Qualitative Content Analysis

To ensure a systematic and structured literature review, this thesis employed QCA following the principles of Mayring and Fenzl (2022). As a rule-based method for synthesizing insights from textual sources, QCA was used within the DSR framework to support the problem identification and objective definition phases (Hevner et al. 2004; Peffers et al. 2007). In this thesis, it was applied in a literature-focused manner to structure the review and provide a traceable basis for subsequent RE.

The analysis was scoped along three dimensions. The *analysis unit* was defined as the overall body of literature addressing operational excellence, workflow automation, and agentic AI. The *context unit* consisted of individual publications (books, peer-reviewed articles, industry reports, standards). The *coding unit* was defined as discrete statements or conceptual claims relevant to the intersection of OpEx, automation paradigms, and AI-based multi-agent systems.

A mixed deductive-inductive approach was used. Deductive categories were derived from established theory, including OpEx dimensions such as adaptability, compliance, and decision quality, as well as prior automation frameworks (e.g., Robotic Process Automation, RPA; Intelligent Process Automation, IPA). Inductive categories emerged from the material itself, capturing issues highlighted repeatedly in the sources, such as observability, traceability, and governance in agentic AI. This balance ensured that both established and novel concerns were systematically reflected.

The outcome of this categorization was not a formal codebook, but a set of thematic clusters that guided the narrative structure of Section 3. Each subsection of the literature review is organized around these categories, which in turn serve as the input for the elicitation lists presented at the beginning of Section 4. In this way, QCA provides both a conceptual ordering of the literature and a direct bridge into the requirements engineering process.

2.2 Requirements Engineering

Following the IEEE (1990) definition, a requirement is a *condition or capability needed by a user to solve a problem or achieve an objective.* RE provides the systematic means to derive such objectives. In this thesis, RE was applied in the early phases to ensure that the conceptual architecture rests on precise, validated needs rather than general aspirations.

The synthesis of requirements followed a structured but literature-driven process. Recurring design concerns were identified from the results of the QCA, documented in *elicitation lists*, and then consolidated into a unified set of requirement candidates. Consistent with Glinz et al. (2020), each candidate was reformulated into an atomic, unambiguous, and verifiable "shall" statement and classified into *functional requirements* (system behaviors), *quality requirements* (non-functional attributes such as performance or compliance), or *constraints* (technological or regulatory limits).

While this procedure draws on the phases described by Herrmann (2022) (elicitation, documentation, analysis, management), it was adapted to the scope of this thesis: instead of stakeholder workshops, the primary elicitation source was the systematically coded literature. To situate the requirements, a complementary systems analysis defined the system boundary, identified stakeholders and external actors, and clarified interface obligations—helping prevent scope creep and omissions.

Requirements were then represented in SysML requirement diagrams. Trace links connect each documented requirement to the respective architecture elements, enabling full requirement-to-design traceability via ≪`satisfy`≫ and ≪`verify`≫ relationships. This model-based approach ensures that design decisions can always be traced back to validated needs and that no requirement was overlooked.

In summary, the integration of RE and systems analysis provided a structured, traceable, and quality-assured requirement set. This foundation anchors the subsequent conceptual architecture in rigorously defined objectives, ensuring consistency with both DSR methodology and operational excellence goals.

2.3 Information Systems Design

In line with the DSR approach, this thesis models the architecture in SysML v2 and applies MBSE practices to ensure requirements-to-design traceability. Although MBSE originates in systems engineering, its discipline transfers well to information systems and supports the systematic development of conceptual architectures. As

MAS grow in scope and complexity, a formal modeling framework becomes essential for maintaining control, observability, and consistency. SysML v2 is explored here as that framework, providing precise semantics and a version-controllable, analyzable representation of the artifact.

MBSE is used to transform the synthesized requirements into a coherent, analyzable system model and to validate the design at the conceptual level. Concretely, each requirement is represented as a SysML «`requirement`» element and linked via «`satisfy`» and, where applicable, «`verify`» relations to architectural elements (agent roles, interactions, and policies). This establishes requirements-to-design traceability and enables early checks against stakeholder needs and constraints. The resulting model offers a unified view of structure and behavior, making interface obligations, coordination patterns, and exception paths inspectable before implementation.

The conceptual architecture is captured as a SysML v2 model using a plain-text, Eclipse-based workflow. SysML v2's textual notation enables precise, tool-agnostic definitions of structure and behavior under version control. Compared with SysML v1 and informal diagramming, SysML v2 offers clearer semantics and richer constructs for specifying MAS concerns such as coordination patterns, policy enforcement points, interfaces, and exception pathways. In this work, the SysML v2 model functions as a machine-interpretable blueprint of the architecture: it supports early validation against the requirements, consistent terminology and interfaces, and a stable foundation for subsequent instantiation and evaluation.

3 Literature Review

The literature review is organized into three subsections that together frame the problem context of this thesis. It begins with operational excellence, which defines the strategic objectives—adaptability, compliance, decision quality—that guide enterprise transformation efforts. The second subsection addresses workflow automation, as the established technological approach for operationalizing these objectives in practice. The third subsection turns to agentic AI, a rapidly emerging paradigm that extends automation beyond deterministic scripts toward adaptive, tool-using agents. This sequence—objectives, established solutions, emerging solutions—provides a logical progression from strategic goals to current practice and then to prospective innovations. It ensures that the requirements synthesized in Section 4 are grounded

in both enduring management principles and the latest technological developments relevant to workflow design.

3.1 Operational Excellence

OpEx originated as a management philosophy in the manufacturing sector, particularly in the automotive industry to optimize quality and efficiency. In this classical context, OpEx focused on minimizing defects, eliminating waste, and embedding continuous improvement practices into organizational routines (Womack and Jones 1997). While these roots remain important, they provide only a partial foundation for understanding OpEx in today's IT-driven enterprises, which operate in volatile environments shaped by rapid technological change, regulatory complexity, and global competition.

One central dimension of OpEx is ADAPTABILITY AND AGILITY. In IT-driven firms, where OpEx is defined less by physical production flows and more by the ability to execute strategies effectively while maintaining innovation, adaptability refers to the capacity of processes to be reconfigured in response to volatility, ensuring resilience in dynamic environments. Agility emphasizes change-readiness and rapid adaptation, qualities increasingly recognized as indicators of organizational excellence in globalized and unpredictable markets. While adaptability enhances resilience, it can reduce efficiency if frequent changes disrupt standardization; conversely, a strong focus on efficiency can make processes rigid and less responsive to unexpected events. In practice, organizations must reconcile continuous improvement with agility, balancing stability and flexibility in their operational routines (cf. Carvalho et al. 2023, p. 1599).

Another recurring theme in the literature is COMPLIANCE AND RISK MANAGEMENT. OpEx in regulated industries demands that workflows embed mechanisms for ensuring transparency, auditability, and regulatory adherence. Compliance safeguards minimize operational risk but can introduce bureaucratic overhead and slow decision-making. The key challenge is balancing strict rule enforcement with the flexibility needed to respond to novel business conditions, a tension especially visible in digital service environments with evolving legal frameworks (cf. Owoade et al. 2024, p. 687).

A further category is DECISION QUALITY. A core aim of OpEx is not only to accelerate decision-making but to improve its reliability and evidential grounding (cf.

Owoade et al. 2024, p. 685). Automation can help by aggregating relevant data and reducing errors, but it also risks opacity and overconfidence when human oversight is limited. The central trade-off is speed versus quality: excessive automation may produce faster but less accountable outcomes, while excessive oversight slows operations. For sustainable excellence, systems must therefore support evidence-based choices and make decision pathways transparent (Carvalho et al. 2023).

Another key dimension is EFFICIENCY AND CONTINUOUS IMPROVEMENT. Rooted in Lean and TQM traditions, efficiency emphasizes minimizing waste and reducing manual effort, while continuous improvement institutionalizes iterative refinements in processes. Together, these principles enhance operational reliability and cost-effectiveness, yet they can conflict with the need for flexibility in volatile environments. The challenge is to design workflows that are optimized for today while remaining adaptable for tomorrow (cf. Juran and Godfrey 1999, p. 14.16 & p. 8.1).

Customer-facing outcomes are captured by CUSTOMER-CENTRICITY. OpEx emphasizes that processes must be aligned with user needs and service-level commitments, ensuring reliability, responsiveness, and satisfaction. A strong customer focus can create pressure to customize and accelerate processes, which may undermine efficiency or compliance. The literature stresses that sustainable excellence requires balancing external demands with internal consistency (Womack and Jones 1997; Juran and Godfrey 1999). The literature stresses that sustainable excellence requires balancing external demands with internal consistency, often formalized through service-level agreements (SLAs) (cf. Duan et al. 2015, p. 115).

The category of USER EMPOWERMENT AND CULTURE recognizes that operational excellence is not solely technical but also organizational. Effective improvement requires systems that support transparency, collaboration, and employee engagement. Empowerment fosters ownership and participation, but decentralizing authority can introduce inconsistency and conflict with standardization goals. Culture therefore functions as both an enabler and a constraint for process excellence, shaping how automation is accepted and leveraged by human actors (cf. Juran and Godfrey 1999, p. 15.2-3) (cf. Womack and Jones 1997, p.3).

Finally, TECHNOLOGY INTEGRATION AND SCALABILITY reflects the increasing role of digital platforms in enabling OpEx. Modern enterprises rely on architectures that integrate automation, AI, and cloud services to achieve scalable and resilient operations (Owoade et al. 2024). Integration enables end-to-end process coverage and agility, but also increases dependency on heterogeneous systems and external

vendors. Scalability promises growth and innovation, yet without careful governance it can amplify complexity and risk.

3.2 Workflow Automation

Building on the need to balance stability and agility in operational processes, workflow automation emerged as a means to systematically coordinate and streamline business workflows. It refers to the use of software systems (workflow management systems, WfMS) to orchestrate tasks, information flows, and decisions along a predefined business process model. According to industry standards, workflow automation entails routing documents, information, or tasks between participants according to procedural rules, with the goal of reducing manual effort and variability in execution (cf. Basu and Kumar 2002, p. 2). Early WfMS in the late 20th century were designed to make work more efficient, to integrate heterogeneous applications, and to support end-to-end processes even across organizational boundaries (cf. Stohr and Zhao 2001, p. 281).

One foundational aspect of workflow automation is PROCESS ORCHESTRATION. By encoding business procedures into formal process models that are executed by a *workflow engine*, organizations could enforce consistent process flows, improve speed and accuracy, and embed compliance checks into routine operations. In essence, pre-AI workflow automation provided a structured, deterministic way to implement business processes in software, directly addressing chronic issues like fragmented manual tasks and data silos in pursuit of OpEx (cf. Gadatsch 2012, p. 231). Orchestration denotes the centralized coordination of tasks and activities according to a defined process logic. In a typical WfMS, a workflow engine enacts the process model, dispatching tasks to the right resources (human or machine) in the correct sequence and enforcing the business rules at each step (Basu and Kumar 2002). This engine-driven coordination brings predictability and repeatability to workflows: tasks are executed in a fixed, optimized order with minimal ad-hoc variation. By systematically controlling task flow, early workflow systems could eliminate many manual hand-offs and delays, thereby boosting efficiency and consistency in outcomes. The orchestration approach essentially translated managerial routines into software: for example, an order processing workflow would automatically route an order through credit check, inventory allocation, shipping, and billing steps without needing human coordination at each transition. Such deterministic sequencing was crucial for

achieving the quality and reliability targets of OpEx in an era before adaptive AI capabilities (cf. Stohr and Zhao 2001, pp. 283-285).

A closely related design concern is INTEGRATION. Workflow automation inherently requires linking together diverse people, departments, and IT systems into an end-to-end process. Literature emphasizes that WfMS must integrate heterogeneous application systems and data sources to allow seamless information flow across functions (cf. Stohr and Zhao 2001, pp. 289-290). For instance, a procurement workflow might connect an ERP inventory module, a supplier's database, and a financial system so that each step can automatically consume and produce the necessary data. This integration extends beyond technical connectivity; it also encompasses coordinating work across organizational boundaries. As e-business initiatives grew in the 1990s and early 2000s, workflows increasingly spanned multiple organizations (suppliers, partners, customers), demanding inter-organizational process integration (Basu and Kumar 2002). Research in this period identified the need for distributed workflow architectures that could bridge independent systems and companies. Georgakopoulos et al. (1995) noted that existing workflow tools had limitations in complex environments, calling for infrastructure to handle heterogeneous, autonomous, and distributed information systems. In practice, this led to the development of interoperability standards (e.g., XML-based process definitions, web service interfaces) and process choreography protocols to ensure that a workflow could progress smoothly even when multiple organizations or platforms were involved. Effective integration was thus an essential condition for workflow automation, enabling the end-to-end automation of processes that formerly stopped at organizational or system boundaries.

To manage complexity and change, MODULARITY in workflow design became another important principle. Rather than hard-coding monolithic process flows, architects sought to break workflows into modular components or sub-processes that could be reused and reconfigured as needed. This component-based approach was accelerated by the rise of service-oriented architectures and e-business "workflow of services" concepts. For example, composite e-services and e-hubs allow organizations to construct complex workflows by composing smaller service modules. A modular workflow architecture improves maintainability: if a business rule changes or a new subprocess is required, one can update or insert a module without redesigning the entire workflow from scratch. Modularity also underpins adaptability. Ideally, a workflow systematizes routine functions but can be adjusted to accommodate new

requirements or variations in the process (cf. Basu and Kumar 2002, p. 10). In other words, the literature suggests that well-designed workflow automation should combine standardization with flexibility: processes are structured into clear modules for the "happy path" of routine operations, yet those modules can be reorchestrated or overridden in exceptional cases. This design philosophy reflects an early recognition that no single process model can anticipate all future conditions, so a degree of configurability must be built in (Georgakopoulos et al. 1995).

Despite efforts to introduce flexibility, traditional workflow automation faced notable challenges with EXCEPTION HANDLING. Exception handling refers to the ability of a system to cope with deviations, errors, or unforeseen scenarios that fall outside the predefined process flow. Basu and Kumar (2002) candidly observe that existing WfMS "tend to fall short whenever workflows have to accommodate exceptions to normal conditions"—i.e., when something unexpected occurs that was not explicitly modeled, the system often cannot resolve it autonomously, forcing human intervention. Typically, designers might anticipate a limited number of exception scenarios and build alternate paths for those (e.g., an approval escalation if a manager is absent). However, if a novel exception arises (say, a new regulatory requirement or an unplanned system outage affecting a step), the rigid workflow cannot handle it, and manual workarounds are needed. This problem of early workflows under dynamic conditions was widely acknowledged (Georgakopoulos et al. 1995). In the pre-AI era, most workflow automation remained predominantly rule-driven and inflexible outside of predefined contingencies. Exception handling thus stood out as a critical limitation of classical automation approaches, highlighting a gap between the desire for end-to-end automation and the reality of complex, ever-changing business environments.

Another salient theme in the literature is WORKFLOW GOVERNANCE—the structures and mechanisms for overseeing automated workflows and aligning them with organizational policies. As companies entrusted core business processes to software, ensuring the correct and intended execution of those processes became vital. Key governance considerations include monitoring, auditing, and controlling workflows (cf. Georgakopoulos et al. 1995, p. 131). A WfMS typically provides monitoring dashboards and logs so that managers can track the state and performance of process instances (e.g., to identify bottlenecks or errors). It also enforces role-based access control, ensuring that only authorized personnel perform certain tasks or approvals, which is essential for compliance in regulated industries. Basu and Kumar (2002)

highlight the importance of organizational "metamodels" and control mechanisms that tie workflows to an enterprise's structure—for example, defining which organizational roles are responsible for each task and how escalation or overrides should happen under specific conditions. Additionally, governance extends to establishing standards and best practices for workflow design and deployment. Industry coalitions and standards bodies (such as the WFMC in the 1990s) issued reference models and interface standards to promote consistency and interoperability in workflow implementations. In the context of inter-organizational workflows, governance also means agreeing on protocols and service-level commitments between partners so that automated interactions remain trustworthy and transparent. Overall, robust governance in workflow automation ensures not only efficiency but also accountability, security, and compliance. It addresses the managerial and oversight challenges that arise once processes are no longer directly handled by individuals but by software agents following prescribed logic.

3.3 Agentic Artificial Intelligence

Agentic AI refers to systems composed of multiple interacting generative agents that autonomously collaborate to achieve complex goals. It represents a shift beyond single AI agents toward orchestrated multi-agent ecosystems, enabled largely by recent advances in the field. Whereas traditional automation (e.g. rule-based RPA) executes predefined steps, an agentic architecture features adaptive, goal-directed agents that can perceive context, make decisions, and act with minimal hard-coded instructions (cf. Jennings et al. 1998, pp. 8-9). This idea builds on classic MAS principles of autonomy and social action (Castelfranchi 1998; Ferber 1999), but now agents are augmented with learning and reasoning capabilities from large language models (LLMs). Sapkota et al. (2025) highlight this paradigm shift by highlightin the difference between AI agents and agentic AI, framing the later as "multi-agent collaboration, dynamic task decomposition, persistent memory, and orchestrated autonomy" in pursuit of flexible problem-solving.

A defining trait of agentic AI is a higher degree of AUTONOMY. Agents can operate without constant human or central control, making and executing decisions in real time. Early MAS research already emphasized agent autonomy—e.g. agents as entities with independent control over their actions and state. Modern generative agents greatly amplify this autonomy by leveraging LLM-based reasoning to plan

multi-step actions toward goals. For instance, frameworks like AutoGPT (Yang et al. 2023) demonstrated that a single LLM-based agent can iteratively break down objectives, choose actions, and adjust based on feedback without human intervention. This autonomy promises agility and decision quality (agents can respond to situational changes or large search spaces beyond rigid scripts), but it also introduces risks. As Russell et al. (2015) caution, each additional decision delegated to an opaque AI agent shifts "ethical control" away from human operators. In enterprise settings, uncontrolled autonomous decisions might lead to policy violations or unsafe actions (Gaurav et al. 2025). Thus, an architectural challenge is balancing agent freedom with mechanisms to supervise or constrain critical decisions. In practice, this means designing agents with clearly scoped authorities, fail-safes, or escalation paths (e.g. requiring human confirmation for high-impact actions) to align autonomy with organizational policies.

tool-use capability is another hallmark of agentic AI architectures. Simply put, agents are not limited to their built-in knowledge; they can invoke external tools, APIs, or other services as part of their reasoning loop. This extends an agent's functionality—for example, an AI agent might call a database, run a code snippet, or query web services to gather real-time information. Research shows that augmenting LLM agents with tool integration significantly improves their problem-solving scope and accuracy. Notably, the ReAct paradigm interleaves an agents chain-of-thought with tool calls, allowing it to perceive (via queries) and act (via external operations) iteratively (Yao et al. 2023). Such designs transform static LLMs into dynamic cognitive agents that can perceive, plan, and adapt, a critical capability for complex, multi-step workflows. For workflow automation, this means an agent can not only parse instructions but also execute parts of a workflow (e.g. trigger an RPA bot or send an alert) and then reason over the results. Architecturally, enabling tool use requires adding interface layers for the agent to safely interact with enterprise systems (APIs, databases, RPA scripts), along with policies on allowed tools. It's worth noting that tool integration adds orchestration complexity and potential error propagation paths (Sapkota et al. 2025). Therefore, designs often include an orchestration layer or planner agent that manages when and how tools are invoked, checks tool outputs, and handles exceptions (e.g. what if a tool fails or returns unexpected data). In summary, tool-use greatly enhances agent capabilities, but it demands careful architectural planning to manage the added complexity and ensure robust tool-agent interaction.

Agentic AI systems are inherently multi-agent—they comprise not one but many

agents, often with specialized roles, that must coordinate their efforts. This multi-agent approach stems from the insight that complex workflows can be decomposed: instead of one monolithic AI agent trying to do everything, a team of agents can each handle subtasks and then combine results. Such specialization aligns with principles of OpEx (e.g. division of labor and expertise) and has been shown to improve performance. For example, Shu et al. (2024) report that a collaborative team of LLM-based agents achieved up to 70% higher success rates on complex tasks compared to a single-agent approach. Architecturally, coordination mechanisms are crucial to harness these gains. Agents need to communicate their intentions, share data/results, and synchronize plans. The literature distinguishes coordination structures along two dimensions: hierarchy vs. flat and centralized vs decentralized decision-making. In a centralized hierarchical design, a top-level planner/manager agent delegates tasks to subordinate agents and integrates their outputs (akin to a project manager overseeing specialists). This can simplify global coordination and ensure alignment with a single source of truth (the planner's goal), at the cost of a single point of failure or bottleneck. Conversely, decentralized teams use peer-to-peer negotiation or voting; all agents are more equal and collectively decide on task assignments or conflict resolution (drawing on concepts from distributed AI and game theory). For instance, one recent system had developer agents jointly agree on a solution design without a central boss, mimicking consensus decision-making (Qian et al. 2024). Each approach has trade-offs: hierarchical control can be more efficient for well-structured processes, while decentralized collaboration may be more robust to single-agent failure and better for ill-structured problems. Inter-agent communication protocols (what messages agents send and when) are another design facet—simple cases use direct message passing or shared memory, whereas more complex setups might use an event-bus or blackboard architecture for asynchronous communication. Importantly, specialization means each agent can be bounded in scope (e.g. a compliance checker agent vs. a data retrieval agent), which helps with scalability: each agent's LLM or reasoning module can operate within a focused context window, and different team members can even use different model types suited to their niche. This modularity and specialization, orchestrated through well-defined coordination logic, is a key architectural strength of agentic AI. It mirrors how human organizations structure teams for efficiency, and indeed is crucial for aligning multi-agent AI workflows with complex enterprise processes.

As agent behaviors become more autonomous and distributed, ensuring observ-

ability of the system is vital. Observability here means that the internal states, decisions, and actions of agents can be monitored and understood by humans or supervisory systems. Traditional MAS literature often dealt with observability in terms of state visibility (e.g. in partially observable environments), but in an enterprise context it translates to runtime transparency and traceability of what agents are doing and why. One challenge is that LLM-driven agents reason in natural language (or latent vectors), making their decision process somewhat opaque. Sapkota et al. (2025) highlight that AI agents "lack transparency, complicating debugging and trust", and they advocate for robust logging and auditing pipelines to make agent operations inspectable. In practice, agentic architectures include components to log key events: each prompt an agent generates, each tool API call and its result, each decision or plan the agent commits to, etc. Such audit logs enable post- hoc analysis, error tracing, and explanations—for example, if a workflow failed or a compliance issue occurred, developers can replay the agent interactions to pinpoint the cause. Some frameworks even expose an agent's chain-of-thought (the intermediate reasoning steps) in a controlled way for debugging or compliance review. Beyond logging, observability can be enhanced through dashboarding and alerts: e.g. real-time monitors that track agent performance metrics or detect anomalies (like an agent taking too long on a task or generating an out-of-bounds output). The end goal is to treat an agentic AI system not as a "black box" automation, but as an observable workflow that operations teams can supervise akin to any critical IT system. This also ties into explainability: by capturing the rationale behind decisions (even if only in approximate form, such as storing the intermediate reasoning text), the system can later provide explanations for its actions, which is invaluable for trust and for continuous improvement. Overall, the literature suggests that designing for transparency—instrumenting agents with logging, and perhaps even designing agents to self-report their status—is a best practice to ensure agentic AI doesn't become an inscrutable tangle of automations. High observability supports OpEx principles by enabling traceability, accountability, and faster incident response when something goes wrong.

Finally, a recurrent theme is the need for strong governance mechanisms in agentic AI architectures to ensure alignment with rules, ethics, and organizational policies. By their nature, autonomous agents may produce unexpected or undesired outcomes—a risk amplified in multi-agent settings where interactions are complex and no single agent has full oversight. Without proper governance, an agentic system

could easily violate compliance requirements or strategic constraints, undermining OpEx goals (e.g. a well-intentioned agent might inadvertently expose sensitive data or execute an unauthorized transaction). In fact, Gaurav et al. (2025) warn that the "absence of scalable, decoupled governance remains a structural liability" in today's agentic AI ecosystems. To address this, researchers are exploring policy-enforcement layers that sit between the agents and the outside world. One such approach is Governance-as-a-Service (GaaS), a framework that intercepts agent actions at runtime and checks them against explicit rules or constraints. Rather than trusting each agent to self-regulate, an external governance layer can block or redirect high-risk actions, log rule violations, and even adapt penalties or restrictions on agents that exhibit misbehavior over time. This effectively creates an oversight controller for the MAS—analogous to a "compliance officer" in a human organization—that ensures no single agent can compromise the system's integrity. Key design elements include declarative policy rules (defining allowable vs. disallowed outputs or tool uses), a mechanism to monitor all agent outputs (to flag violations), and possibly a trust score or reputation model to quantify an agent's reliability based on past behavior. Beyond automated enforcement, governance also encompasses human oversight: for example, requiring human approval for certain agent decisions (human-in-the-loop checkpoints) or having a fallback where a human operator can intervene if the agents encounter an ambiguous ethical situation. In classical MAS research, analogous concepts existed like normative agents and electronic institutions that enforce "rules of engagement" among agents; the new twist is that with LLM-based agents we must often treat the models as black boxes, so governance can't be injected into their internal logic easily and must surround them instead. The overarching recommendation is that any architecture for generative multi-agent workflows should bake in governance from the start—not as an afterthought—to manage risk. As one author succinctly put it, such a system "does not teach agents ethics; it enforces them". This governance emphasis aligns tightly with OpEx goals of compliance, risk management, and trustworthiness (Gaurav et al. 2025).

In summary, the literature portrays agentic AI as a powerful paradigm for workflow automation that, if well-designed, can dramatically enhance agility, adaptability, and decision quality in operations. By combining autonomous, tool-using agents into coordinated architectures, organizations can automate complex processes that previously required human judgment. At the same time, achieving sustainable excellence with such systems demands careful attention to transparency and control:

architects must ensure agents remain observable and governable to uphold compliance and reliability standards. These insights set the stage for the next section of this thesis, which will integrate OpEx principles, workflow automation requirements, and agentic AI capabilities into a unified reference architecture. The themes of autonomy, coordination, and governance identified here directly inform the design choices and requirements elaborated in the subsequent chapters.

4 Requirements Modeling

The literature review identified recurring design concerns across operational excellence, workflow automation, and agentic AI. Synthesizing these insights yields *elicitation lists* which represent the initial outcome of requirements documentation based on systematically coded sources. A subsequent *clustering* step consolidated overlapping items into a unified set of requirement candidates.

Each candidate was then reformulated into an atomic, unambiguous, and verifiable "shall" statement, following the best-practice formulation rules of Glinz et al. (2020). The final requirements were organized into *functional requirements*, *quality requirements*, and *constraints*, in line with the Glinz taxonomy.

Requirements were then represented in SysML v2 as dedicated requirement elements, with their textual statements captured in the description field. Trace links connect each requirement to its source in the elicitation lists and to the architecture elements that «`satisfy`» it. This model-based representation ensures that design decisions remain traceable to validated needs and that requirement coverage can later be verified systematically.

4.1 Clustering and Consolidation

In order to derive actionable system requirements from the literature, a structured clustering process was applied. This eliminated redundancies, consolidated overlapping issues, and normalized vocabulary across disciplines. The process began by translating *coded statements* from each domain into elicitation items—discrete, *design-relevant units* derived from the coding units identified in the QCA process described in Section 2.1.

O — OPERATIONAL EXCELLENCE

1. ADAPTABILITY AND AGILITY — processes must remain reconfigurable in response to volatile conditions, ensuring resilience in dynamic environments.
2. COMPLIANCE AND RISK MANAGEMENT — regulatory adherence and transparency must be embedded into workflows to minimize compliance risks.
3. DECISION QUALITY — automation should enable data-driven, timely, and well-informed decisions rather than simply increasing speed.
4. EFFICIENCY AND CONTINUOUS IMPROVEMENT — workflows should reduce manual effort, eliminate waste, and institutionalize iterative refinements.
5. CUSTOMER-CENTRICITY — operations must align with user needs and service-level commitments to sustain value delivery.
6. USER EMPOWERMENT AND CULTURE — systems should support collaboration, transparency, and employee engagement in improvement processes.
7. TECHNOLOGY INTEGRATION AND SCALABILITY — architectures must accommodate automation, AI, and cloud services to enable sustainable innovation.

W — WORKFLOW AUTOMATION

1. PROCESS ORCHESTRATION — workflow engines must enforce task sequences and business rules to guarantee reliable execution.
2. INTEGRATION AND INTEROPERABILITY — automation must seamlessly connect heterogeneous applications, data sources, and organizational boundaries.
3. MODULARITY AND REUSABILITY — workflows should be composed of modular tasks or subprocesses that can be reused and reconfigured with minimal effort.
4. EXCEPTION HANDLING AND FLEXIBILITY — systems must detect, manage, and escalate deviations rather than failing in unforeseen scenarios.
5. WORKFLOW GOVERNANCE — monitoring, audit trails, and role-based controls must ensure accountability and compliance throughout automated processes.

A — AGENTIC AI

1. AUTONOMY IN DECISION-MAKING — agents should operate independently within clearly scoped authority to enhance agility while managing risks.
2. TOOL USE AND INTEGRATION — agents must invoke external tools, APIs, or services reliably, requiring robust interfaces and safeguards.
3. COORDINATION AND SPECIALIZATION — multi-agent systems should divide labor through explicit roles and structured coordination mechanisms.

4. OBSERVABILITY AND TRANSPARENCY — all agent actions and decisions must be logged and explainable to support trust, debugging, and compliance.
5. GOVERNANCE AND COMPLIANCE — oversight mechanisms, including policy enforcement layers and human-in-the-loop checkpoints, are essential to align agent behavior with organizational and ethical standards.

To reduce redundancy and ensure conceptual clarity, the elicitation items from each domain were compared and consolidated based on semantic similarity and functional overlap. Particular attention was given to cross-domain intersections—such as governance appearing in both workflow automation and agentic AI—and to areas where multiple coding units aligned on a shared concern. The result of this consolidation step is a reduced set of requirement candidates, each traceable to one or more domain-specific clusters. These are summarized in Table 4.1, which maps the original coded clusters to the consolidated requirement areas used for formulation.

CLUSTER IDS	SOURCE DOMAINS	CONSOLIDATED REQ. AREA
O1	OPEX	ADAPTABILITY AND AGILITY
O4	OPEX	CONTINUOUS IMPROVEMENT
O5	OPEX	CUSTOMER-CENTRICITY
O6	OPEX	USER EMPOWERMENT
O7	OPEX	TECH. INTEGRATION & SCAL.
O2, W5, A5	ALL THREE	GOVERNANCE & COMPL.
O3, A1	OPEX, AGENTIC AI	DECISION QUALITY
W1, W3	WORKFLOW AUTOMATION	ORCHEST. & MODULARITY
W4, A3	WORK. AUT., AGENTIC AI	EXCEP. HANDLING & COORD.
A4, O2	AGENTIC AI, OPEX	OBSERVABILITY & TRAC.
W2, A2	WORK. AUT., AGENTIC AI	TOOL INTEGRATION

Table 4.1: Mapping of domain-specific clusters to consolidated requirement areas (in-scope areas highlighted in gray).

This structured consolidation establishes a coherent foundation for refining architecture-level requirements. To maintain architectural focus and avoid inflation of the requirement set, areas originating solely from managerial OpEx—namely, adaptability and agility, continuous improvement, customer-centricity, user empowerment,

and technology integration and scalability—were *not* modeled. Their architecturally relevant aspects were absorbed into cross-domain clusters, while non-architectural concerns were acknowledged as higher-level managerial frameworks, guidelines, and principles rather than system requirements.

4.2 Reformulation and Classification

The logical next step, consistent with RE methodology, is to dissect each consolidated requirement area individually and distill it into a set of atomic "shall" statements. This transformation emphasizes clarity, necessity, and verifiability, ensuring that each requirement stands alone as an actionable directive. At this stage, the focus remains on precision and alignment with design intent; formal classification into functional, quality, or constraint types follows only after this refinement process. Several requirement areas share overlapping architectural concerns. This reflects systemic interdependencies, particularly in adaptive, agent-based architectures and, more broadly, in complex systems. To avoid redundancy, each requirement is assigned to the cluster that most directly motivates it.

However, to address the tradeoff between best-practices and readability, a deliberate compromise was made in the formulation of requirements. While the requirements adhere to the core principles of clarity, verifiability, and design relevance, strict atomic decomposition was applied selectively. In cases where overly granular formulation would hinder readability or inflate the requirement set without clear architectural benefit, semantically related concerns were consolidated into cohesive statements. Furthermore, the scope of formal requirements was restricted to those directly affecting system architecture—such as structure, behavior, and coordination—while non-architectural concerns (e.g., organizational or cultural aspects) were acknowledged in the literature but excluded from the formal specification.

GOVERNANCE AND COMPLIANCE The literature consistently frames governance and compliance as architectural, not merely legal, concerns. Compliance should be embedded as a first-class constraint, with governance mechanisms preventing compliance drift as systems optimize or adapt. Classical workflow automation contributes monitoring, audit trails, and escalation, and has evolved toward policy-driven designs that decouple governance rules from core logic. Agentic AI requires a decoupled policy layer capable of intercepting and vetoing high-risk actions, complemented

by human-in-the-loop checkpoints and comprehensive auditability. Across domains, governance and compliance must operate as active, adaptable components—enforcing policy at runtime while preserving transparency and alignment with organizational and legal constraints (Basu and Kumar 2002; Gaurav et al. 2025).

- [FR] POLICY ENGINE
- [C] SEGREGATION OF DUTIES
- [FR] RISK-BASED APPROVALS AND ESCALATION
- [C] AUDIT LOGGING AND RETENTION
- [FR] COMPLIANCE MAPPING AND DRIFT CHECKS

DECISION QUALITY The literature portrays decision quality as producing reliable, evidence-based outcomes rather than merely increasing speed. In operational excellence, this means balancing rapid execution with transparency and accountability. Classical workflow automation contributes rule enforcement, role clarity, monitoring, and escalation to standardize and audit decision points. Agentic AI extends decision reach but must be bounded and observable: agents should operate within scoped authority, log the context of their choices (prompts, tool interactions, plan commitments), and escalate high-impact actions to human review. Together, these insights imply that architecture must combine evidence aggregation with rule-based consistency, human checkpoints for risk, and traceability of decision pathways to sustain quality under change. Based on these insights, the following requirements were derived:

- [FR] EVIDENCE-BASED DECISION SUPPORT
- [FR] RULE-ENFORCED DECISION POINTS
- [FR] DECISION TRACE AND RATIONALE
- [FR] RISK-BASED HUMAN APPROVAL
- [C] BOUNDED DECISION AUTONOMY

ORCHESTRATION AND MODULARITY The literature describes classical workflow automation as the enactment of formal process models by a central engine that coordinates human and machine tasks in a defined order and enforces business rules for predictable, repeatable execution. To manage change and complexity, workflows should be composed of modular subprocesses or services that can be reused and reconfigured without redesign, often realized through service-oriented compositions.

Interoperability standards and clear interfaces enable module composition across heterogeneous systems and organizational boundaries. In OpEx terms, parameterized designs allow variation without structural overhaul, preserving efficiency while enabling adaptation. Based on these insights, the following requirements were derived:

- [FR] PROCESS ORCHESTRATION ENGINE
- [C] MODULAR PROCESS UNITS
- [C] INTERFACE-BASED COMPOSITION
- [FR] PARAMETERIZED SUBPROCESSES
- [FR] HUMAN-MACHINE TASK ORCHESTRATION

EXCEPTION HANDLING AND COORDINATION The literature notes that classical workflow systems struggle with unforeseen exceptions; predefined alternates help, but novel cases often require human intervention. Architectures should therefore detect and route deviations to defined exception paths with appropriate escalation. In agentic AI, multi-agent designs decompose complex work into specialized roles, which demands explicit coordination structures—ranging from hierarchical planner-specialist patterns to decentralized negotiation—plus clear communication protocols (message passing or shared memory, with event-bus/blackboard for asynchronous cases). Together, robust exception pathways and well-specified coordination ensure that workflows remain reliable when reality diverges from the "happy path," while agent teams synchronize decisions without conflict or drift. Based on these insights, the following requirements were derived:

- [FR] EXCEPTION DETECTION AND ROUTING
- [FR] ESCALATION TO HUMAN AUTHORITY
- [C] EXPLICIT COORDINATION MODEL
- [FR] INTER-AGENT COMMUNICATION PROTOCOL
- [FR] TOOL FAILURE HANDLING IN EXCEPTIONS
- [FR] CONFLICT RESOLUTION

OBSERVABILITY AND TRACEABILITY The literature emphasizes that enterprise automation must be inspectable at runtime and explainable post hoc. Classical workflow systems contribute monitoring and logs for process execution visibility. Agentic AI adds opacity risks; to mitigate them, architectures should log agent

prompts, tool interactions, and plan/decision commitments, expose dashboards and alerts for anomaly detection, and support replay so that failures and outcomes can be reconstructed and explained. Across domains, observability and traceability provide accountability, faster incident response, and the basis for explaining why a particular decision or action occurred. Based on these insights, the following requirements were derived:

- [FR] EVENT LOGGING PIPELINE
- [FR] DECISION TRACE AND RATIONALE
- [FR] DASHBOARDS AND ALERTS
- [FR] REPLAY FOR POST-HOC ANALYSIS
- [C] AGENT STATUS SELF-REPORTING

TOOL INTEGRATION The literature presents integration as foundational to both classical workflow automation and agentic AI. Workflow systems must connect heterogeneous applications and data sources, often across organizational boundaries, using clear interfaces and interoperability standards. Agentic AI adds tool-use: agents invoke enterprise APIs, databases, or RPA scripts through interface layers, while an orchestration component manages when and how tools are called and checks results. Architecturally, tool integration therefore centers on well-defined adapters and contracts, protocol-level interoperability, and data transformation to enable end-to-end workflows without brittle coupling. Based on these insights, the following requirements were derived:

- [FR] INTEGRATION CONNECTORS
- [FR] AGENT TOOL ADAPTERS
- [C] INTERFACE CONTRACTS AND SCHEMAS
- [FR] INTER-ORGANIZATIONAL INTEROPERABILITY
- [FR] DATA TRANSFORMATION LAYER
- [C] INVOCATION SAFEGUARDS

The requirements were then consolidated into a final list, grouped by class—*functional requirements* and *constraints*—to provide a single, unambiguous baseline for the subsequent SysML formalization. Here an example of each requirement class:

FR-09 POLICY ENGINE The system shall provide a decoupled policy evaluation component that can validate, veto, or redirect workflow executions and agent actions

at runtime.

C-03 BOUNDED DECISION AUTONOMY The system shall constrain agent decision-making to clearly scoped authority levels aligned with organizational objectives.

The complete list of requirements can be found in Appendix A.1. As the reader can see, *quality requirements* were left out of scope at this stage because this thesis focuses on specifying *what* the system must do and *which* governance and operational constraints it is subject to. Moreover, quality attributes vary significantly by deployment context and would require domain-specific targets, which remain out of scope here. Some quality-related concerns (e.g., auditability, interoperability) were operationalized as *constraints* rather than standalone quality requirements.

While the list optimizes precision, it is not yet visually informative nor operationally easy to manage. In the next section, the requirements are recast as SysML «`requirement`» elements and diagrams to improve visibility, traceability, and change control—linking them explicitly to their sources and preparing `satisfy`/`verify` relations for the architectural and evaluation work that follows.

4.3 Model-Based Representation

The requirement model formalizes the subject system (MAS), the two requirement types (functional requirements and constraints), and the consolidated group MAS_SPEC which requires FR01..21 and C01..09. We use the figure below as a compact, reader-facing view; the full lists are provided in the appendix.

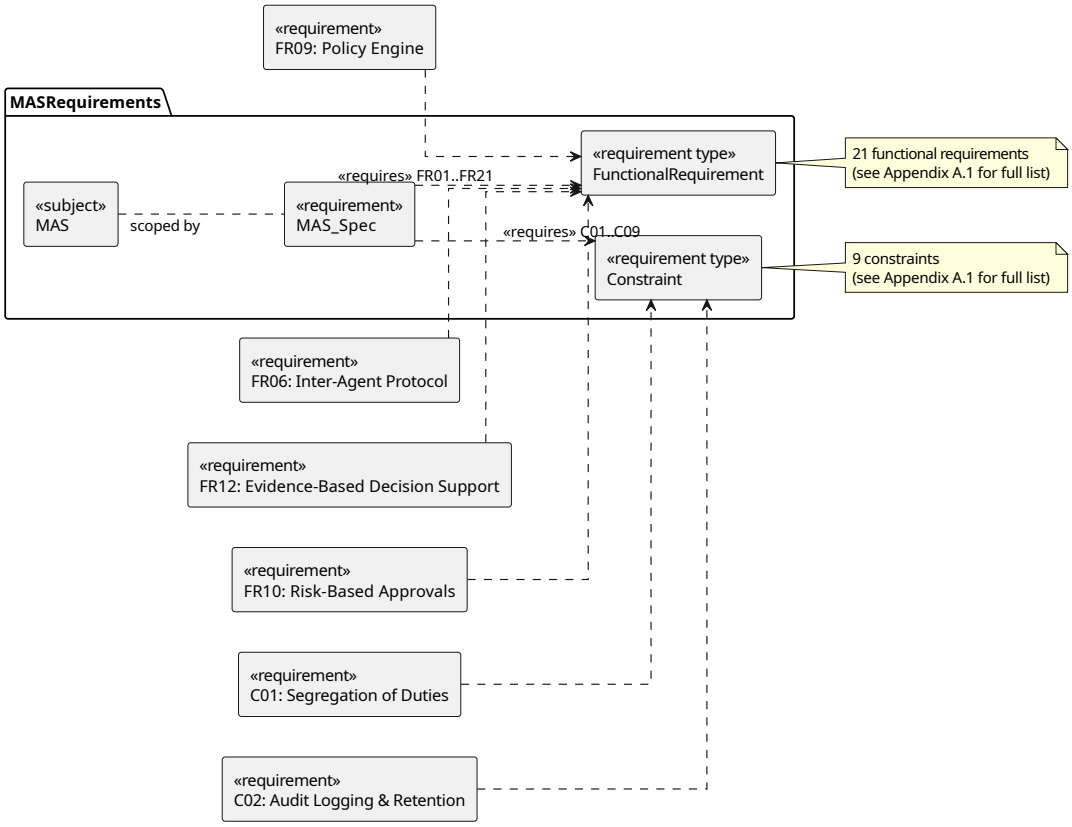


Figure 4.1: Model-based representation of the requirement set (subject, requirement types, consolidated group).

5 Architecture Modeling

This section presents a general, industry-agnostic design that balances agent autonomy with governance, observability, and integration. The architecture is intentionally simple: a central orchestrator coordinates a small set of specialized agents and vetted tools; complex patterns (e.g., multi-tier supervisor trees or free-for-all peer meshes) are avoided to preserve transparency and debuggability (LangChain 2024; Cognition AI 2025).

5.1 Architecture Overview

To keep the design industry-agnostic, the integration layer is bound to *open* interfaces: *Model Context Protocol (MCP)* (Anthropic 2024) standardizes agent-tool/data access via schema-driven servers/clients, while *Agent-to-Agent (A2A)* (Google 2025) stan-

dardizes secure inter-agent discovery and messaging across runtimes. MCP handles agent ↔ tool; A2A handles agent ↔ agent—preserving portability without giving up centralized orchestration.

Interface	Primary responsibility	Why here
MCP	Agent ↔ Tool/Data contracts (capabilities, schemas, auth)	Uniform adapters across SaaS/on-prem; reduces glue code; enforces I/O validation (FR-18..21; C-08, C-09).
A2A	Agent ↔ Agent messaging, discovery, long-running tasks	Vendor-neutral multi-agent collaboration across teams/orgs; complements MCP (FR-20).

Table 5.1: Open protocols used by the integration layer and their scope boundaries.

DESIGN DECISIONS A central supervisor with agents-as-tools pattern was adopted: the orchestrator invokes specialists as function/tool calls and merges results, rather than letting agents freely call one another (LangChain 2024). This yields predictable control flow and clearer accountability (Cognition AI 2025) (satisfies C-06, supports FR-04, FR-08, FR-14, FR-17).

- PROMPTING AND TOOL USE Agents follow a *reason+act* loop (ReAct): interleaving lightweight reasoning with tool calls and state updates, which improves faithfulness and reduces ungrounded steps in complex tasks (Yao et al. 2023). Tool exposure is *scoped per role* ("tool loadouts") to minimize cognitive overhead and context bloat (FR-19; C-09) (Breunig 2025a; Breunig 2025b). Prompt/context was treated as a constrained resource. The orchestrator curates per-step, minimal context; older details are summarized or offloaded to external memory; retrieval is used for relevance (Breunig 2025a; Breunig 2025b) (supports FR-12, FR-14, FR-15; C-07).
- COORDINATION MODEL Orchestrator-mediated hand-offs with explicit task payloads and expected outputs (*handoff contracts*) ensure shared understanding and avoid agent conflict (FR-06, FR-08). Only key outcomes are published to a global scratchpad; low-level traces remain local unless needed (FR-15; C-07) (Cognition AI 2025; LangChain 2024).
- PARALLELIZATION Breadth-first tasks (e.g., multiple independent lookups) may run in parallel; generation tasks (long-form writing/code) remain se-

rial or are merged under Orchestrator/Reviewer control to prevent incoherence (LangChain 2024; Cognition AI 2025). The orchestrator enforces per-tool concurrency caps and rate-limit-aware backoff (jittered retries) to avoid API storms; merge operations use deterministic ordering (e.g., timestamp/key sort) and a reviewer-gated reconcile-before-write for generative outputs (FR-04, FR-08, FR-15).

- EXCEPTION HANDLING Deviation detection → route to *Exception Subprocess* (fallback tools, re-prompt, human escalation) → resume or terminate (FR-04, FR-05, FR-07).

PATTERN	WHY/WHEN	TRADE-OFFS
Central orchestrator, agents-as-tools	default for clarity, traceability, testability.	Single coordination point; easiest to govern/observe.
Decentralized mesh	Only for highly exploratory, research-heavy tasks.	Complex coordination, harder to debug.
Hierarchical supervisors	For very large, multi-team workflows.	Adds layers and latency, out of scope for simplicity.

Table 5.2: Chosen pattern (in gray) vs. alternatives and scope boundaries.

INTEGRATION (INDUSTRY-AGNOSTIC)

- TOOL ADAPTERS with contracts/schemas, timeouts/retries/idempotency (FR-19, FR-21; C-08, C-09).
- OPEN CONNECTORS (MCP) Standardize agent access to tools/data via MCP servers/clients (capabilities, JSON schemas, auth), avoiding bespoke adapters and easing on-prem + cloud integration (Anthropic 2024) (FR-18..21; C-05, C-08, C-09).
- AGENT INTEROPERABILITY (A2A) Use A2A for secure agent discovery, messaging, long-running tasks, and cross-org coordination over web standards (Google 2025) (FR-20). External partners' agents become callable *as if* tools—without breaking the orchestrator pattern.

CONTEXT MANAGEMENT Right-sized prompts; retrieve-then-read; periodic summarization; external scratchpad for verbose/intermediate artifacts (Breunig 2025a; Breunig 2025b) (supports FR-12, FR-14, FR-15; C-07).

FAILURE MODES TO AVOID Very long prompts and indiscriminate context sharing degrade reliability ("context is not free"); prefer retrieval, summarization, and offloaded scratchpads to keep step-specific context tight (Breunig 2025a; Breunig 2025b).

TOP-LEVEL STRUCTURE The system comprises five logical layers (Fig. **??**) each of which maps the primary requirement ownership; several requirements (e.g., FR-04, FR-07, FR-10) are cross-cutting and also appear in DESIGN DECISIONS and GOVERNANCE/INTEGRATION. Because progress in agentic AI is recent and evolving, the literature review establishes the broader scholarly and technical context, while this design section proritizes primary sources (official specifications and technical reports); practitioner posts are cited only to illustrate engineering trade-offs, not as authoritative evidence.

1. INTERACTION LAYER (user/client, human-in-the-loop checkpoints) — entry point, human-machine coordination surfaces, and approval/escalation gates (FR-03, FR-05, FR-10).
2. ORCHESTRATION LAYER (orchestrator agent) — decomposes tasks, sequences steps, performs policy checks; manages parameterized subprocesses and exception routing (FR-01, FR-02, FR-04, FR-09, FR-10; C-04, C-06).
3. AGENT LAYER (specialists) — narrow-scope agents such as researcher, analyst, reviewer/verifier, writer (FR-06, FR-08, FR-12..14).
4. TOOL/INTEGRATION LAYER — adapters to enterprise systems via standard connectors; validates tool outputs and supports recovery/fallbacks (FR-07, FR-18..21; C-05, C-08, C-09).
5. OPERATIONS AND GOVERNANCE LAYER — logging, tracing, dashboards, replay, compliance mapping and drift checks, policy store (FR-11, FR-15..17; C-01, C-02, C-03, C-07).

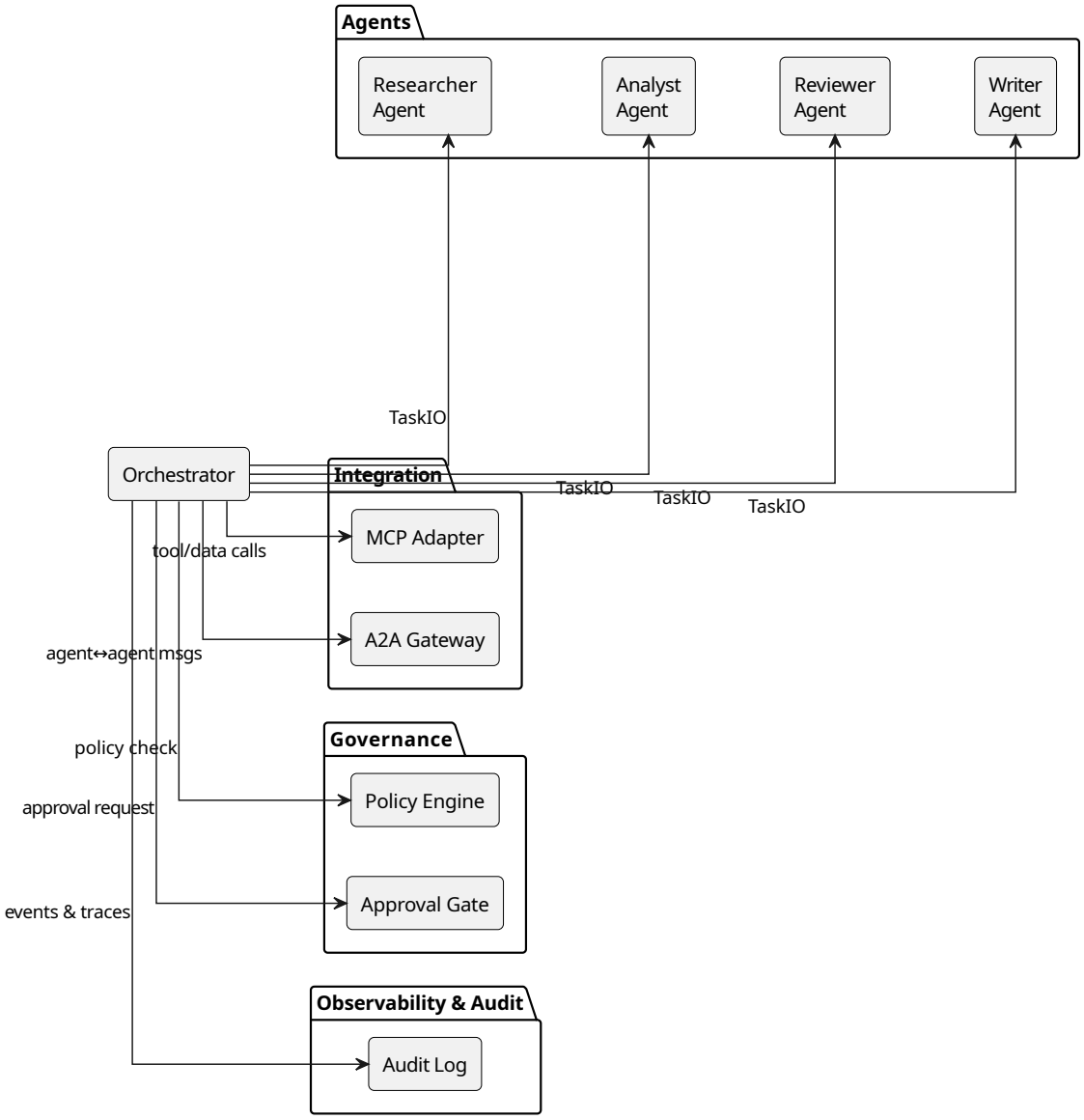


Figure 5.1: High-level architecture with layered responsibilities.

ROLES AND RESPONSIBILITIES

- ORCHESTRATOR — plans, delegates, validates, and gates risky actions; owns workflow state and shared scratchpad (FR-01, FR-04, FR-09..11, FR-15).
- RESEARCHER — retrieves evidence using RAG/tools; writes salient findings to shared state (FR-12, FR-18..21).
- ANALYST — transforms evidence (calculations, checks, comparisons); flags uncertainty and triggers escalation if needed (FR-07, FR-12, FR-10).
- REVIEWER/VERIFIER — critiques outputs against rules and policies; requests

fixes; records rationale (FR-13, FR-14, FR-11).

- WRITER — synthesizes final artifacts to required format; applies review feedback; preserves decision trace (FR-14).

Because MCP (Anthropic 2024) and A2A (Google 2025) are living industry protocols, core properties are documented primarily in official announcements/specs and arXiv technical reports; I triangulate these with emerging analyses. They are included here (design section) because they determine interface contracts and runtime behavior, not just background theory. Where possible, I cite primary protocol docs and recent surveys to mitigate gray-literature risks.

5.2 Governance and Observability

GOVERNANCE (POLICY-FIRST) A decoupled *policy engine* validates proposed actions before execution (allow/deny/escalate), with rules that reference risk thresholds, data sensitivity, or regulatory constraints (FR-09..11; C-01..03). High-impact actions pause at *approval gates* (human-in-the-loop) with resumable state (LangGraph 2024) (FR-10).

SECURITY TOUCHPOINTS Protocol-level controls complement the policy engine: (i) *A2A* supplies authenticated agent identity, authorization, and scoped message exchange for cross-runtime collaboration; (ii) *MCP* enforces schema-validated I/O, capability-scoped tool permissions, and externalized secrets (no secrets in prompts); (iii) least-privilege credentials and per-adapter allowlists bind high-impact actions. These guardrails integrate with approval gates and audit trails (FR-09..11, FR-15..17; C-01..C-03, C-05) (Google 2025; Anthropic 2024).

THREATS AND LIMITATIONS Known risks include prompt/tool injection and data exfiltration via adapters; agent impersonation or message tampering in cross-runtime links; schema drift and capability misalignment between agent intent and MCP servers; and evaluation brittleness for multi-agent coordination. We mitigate via input sanitization, capability allowlists, signed agent identities, per-adapter least-privilege, and replayable traces; residual risk remains and is monitored operationally.

OBSERVABILITY Agents and the orchestrator emit structured events (prompts, tool calls/results, decisions, state changes) (FR-15). Operators get dashboards and alerts for anomalies; the recorded trace enables audits and deterministic replay (FR-16, FR-17), while concise decision rationales are persisted for accountability (FR-14). This trace (prompts, tool I/O, decisions) is the basis for audits and regression

evaluation, and can be replayed deterministically (FR-16, FR-17) (LangChain 2024).

5.3 Architecture Model

6 Discussion: Applicability Criteria

7 Conclusion

References

Anthropic (2024). Introducing the Model Context Protocol (MCP).

Basu, Amit and Akhil Kumar (2002). "Research Commentary: Workflow Management Issues in e-Business". *Information Systems Research* 13, pp. 1–14.

Breunig, Daniel (2025a). How (Long) Contexts Fail—and How to Fix Them.

Breunig, Daniel (2025b). How to Fix Your Context.

Carvalho, André M. et al. (2023). "Operational excellence, organizational culture, and agility: bridging the gap between quality and adaptability". *Total Quality Management & Business Excellence* 34, pp. 1598–1628.

Castelfranchi, Cristiano (1998). "Modelling social action for AI agents". *Artificial Intelligence* 103, pp. 157–182.

Cognition AI (2025). Don't Build Multi-Agents.

Duan, Qing et al. (2015). Data-Driven Optimization and Knowledge Discovery for an Enterprise Information System. 1st. Cham: Springer International.

Ferber, Jacques (1999). Multi-agent systems: an introduction to distributed artificial intelligence. 1st ed. Harlow Bonn: Addison-Wesley, pp. 479–498.

Gadatsch, Andreas (2012). Grundkurs Geschäftsprozess-Management: Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker. 7. Wiesbaden: Vieweg+Teubner Verlag.

Gaurav, Suyash et al. (2025). Governance-as-a-Service: A Multi-Agent Framework for AI System Compliance and Policy Enforcement.

Georgakopoulos, Diimitrios et al. (1995). "An overview of workflow management: From process modeling to workflow automation infrastructure". *Distributed and Parallel Databases* 3, pp. 119–153.

Glinz, Martin et al. (2020). Handbook for the CPRE Foundation Level according to the IREB Standard. International Requiremenst Engineering Board.

Google (2025). A2A: A New Era of Agent Interoperability.

Herrmann, Andrea (2022). Grundlagen der Anforderungsanalyse: Standardkonformes Requirements Engineering. Wiesbaden & Heidelberg: Springer Vieweg.

Hevner, Alan R. et al. (2004). "Design Science in Information Systems Research". *MIS Quarterly* 28, pp. 75–105.

IEEE (1990). Standard Glossary of Software Engineering Terminology. Corrected ed. New York: Institute of Electrical and Electronics Engineers.

Jennings, Nicholas R. et al. (1998). "A Roadmap of Agent Research and Development". *Autonomous Agents and Multi-Agent Systems* 1, pp. 7–38.

Juran, Joseph Moses and A. Blanton Godfrey (1999). Juran's quality handbook. 5th ed. McGraw Hill.

LangChain (2024). How and when to build multi-agent systems.

LangGraph (2024). Human-in-the-Loop Patterns for Agent Workflows.

Mayring, Philipp and Thomas Fenzl (2022). "Qualitative Inhaltsanalyse". Handbuch Methoden der empirischen Sozialforschung. Springer Fachmedien Wiesbaden, pp. 691–706.

Owoade, Samuel et al. (2024). "Systematic Review of Strategic Business Administration Practices for Driving Operational Excellence in IT-Driven Firms". *International Journal of Scientific Research in Science and Technology* 11, pp. 680–700.

Peffers, Ken et al. (2007). "A design science research methodology for information systems research". *Journal of Management Information Systems* 24, pp. 45–77.

Qian, Chen et al. (2024). ChatDev: Communicative Agents for Software Development.

Russell, Stuart et al. (2015). "Research Priorities for Robust and Beneficial Artificial Intelligence". Version 1.

Sapkota, Ranjan et al. (2025). "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges". *Information Fusion* 126, pp. 103–599.

Shu, Raphael et al. (2024). Towards Effective GenAI Multi-Agent Collaboration: Design and Evaluation for Enterprise Applications.

Stohr, Edward A. and J. Leon Zhao (2001). "Workflow Automation: Overview and Research Issues". *Information Systems Frontiers* 3, pp. 281–296.

Womack, J P and D T Jones (1997). "Lean Thinking—Banish Waste and Create Wealth in your Corporation". *Journal of the Operational Research Society* 48.11, pp. 1148–1148.

Yang, Hui et al. (2023). Auto-GPT for Online Decision Making: Benchmarks and Additional Opinions.

Yao, Shunyu et al. (2023). ReAct: Synergizing Reasoning and Acting in Language Models.

Appendix

A.1: Requirements List

FUNCTIONAL REQUIREMENTS

FR-01 PROCESS ORCHESTRATION ENGINE — The system shall execute workflow models by dispatching tasks to human or software actors according to model control flow and business rules.

FR-02 PARAMETERIZED SUBPROCESSES — The system shall support parameterized subprocesses and rules so behavioral variants can be expressed without altering the underlying process structure.

FR-03 HUMAN-MACHINE TASK ORCHESTRATION — The system shall coordinate both human and automated tasks within the same orchestrated process model.

FR-04 EXCEPTION DETECTION AND ROUTING — The system shall detect deviations from the nominal process and route cases to defined exception subprocesses.

FR-05 ESCALATION TO HUMAN AUTHORITY — The system shall escalate unresolved or unmodeled exceptions to designated human roles with clear ownership.

FR-06 INTER-AGENT COMMUNICATION PROTOCOL — The system shall specify message formats and interaction rules for agent collaboration (e.g., direct messaging or shared memory; event-bus/blackboard where asynchronous exchange is required).

FR-07 TOOL FAILURE HANDLING IN EXCEPTIONS — The system shall validate tool/API outputs and trigger defined recovery or fallback steps when invocations fail.

FR-08 CONFLICT RESOLUTION — The system shall provide mechanisms for agents to resolve task or decision conflicts (e.g., delegation to a planner, negotiation, or voting).

FR-09 POLICY ENGINE — The system shall provide a decoupled policy evaluation component that can validate, veto, or redirect workflow executions and agent actions at runtime.

FR-10 RISK-BASED APPROVALS AND ESCALATION — The system shall require human approval and/or escalation for actions or decisions exceeding defined risk or impact thresholds.

FR-11 COMPLIANCE MAPPING AND DRIFT CHECKS — The system shall map actions and policies to applicable regulations or internal rules and perform runtime checks to detect and block compliance drift.

FR-12 EVIDENCE-BASED DECISION SUPPORT — The system shall aggregate relevant data at decision points to support evidence-based choices and reduce errors.

FR-13 RULE-ENFORCED DECISION POINTS — The system shall enforce business rules and role responsibilities at decision points to ensure consistent, auditable outcomes.

FR-14 DECISION TRACE AND RATIONALE — The system shall persist a human-readable decision trace for automated or assisted decisions, including inputs, tool/API calls and results, and concise rationale summaries.

FR-15 EVENT LOGGING PIPELINE — The system shall instrument the workflow engine and agents to emit structured, timestamped events for key actions (e.g., prompts, tool/API invocations and results, plan/decision commits, and state changes).

FR-16 DASHBOARDS AND ALERTS — The system shall provide real-time dashboards and alerting for observability metrics (e.g., execution latency, error/anomaly rates, blocked actions).

FR-17 REPLAY FOR POST-HOC ANALYSIS — The system shall support reconstruction and replay of workflow and agent interactions from logged events to enable root-cause analysis and explanation of outcomes.

FR-18 INTEGRATION CONNECTORS — The system shall provide connectors to integrate heterogeneous applications and data sources required by the workflows.

FR-19 AGENT TOOL ADAPTERS — The system shall expose a uniform adapter interface for agents and workflows to invoke external tools, APIs, databases, or RPA scripts.

FR-20 INTER-ORGANIZATIONAL INTEROPERABILITY — The system shall support protocol and interface interoperability suitable for cross-organizational workflows.

FR-21 DATA TRANSFORMATION LAYER — The system shall provide mapping and transformation capabilities to reconcile data across integrated systems.

CONSTRAINTS

C-01 SEGREGATION OF DUTIES — The system shall enforce role-based access control and segregation of duties for configuration changes and sensitive actions.

C-02 AUDIT LOGGING AND RETENTION — The system shall produce tamper-evident audit trails of agent actions, policy decisions, and configuration changes, retained according to the applicable compliance policy.

C-03 BOUNDED DECISION AUTONOMY — The system shall constrain agent decision-making to clearly scoped authority levels aligned with organizational objectives.

C-04 MODULAR PROCESS UNITS — The system shall structure workflows as modular subprocesses or services that can be reused and reconfigured without redesign.

C-05 INTERFACE-BASED COMPOSITION — The system shall expose clear process and service interfaces compatible with established interoperability standards to enable composition across heterogeneous systems and organizations.

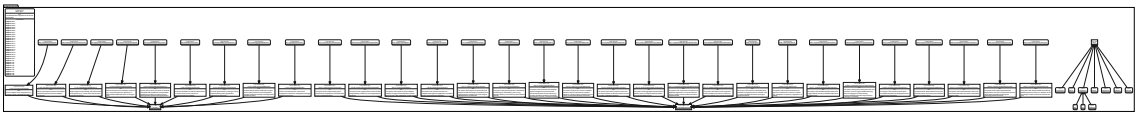
C-06 EXPLICIT COORDINATION MODEL — The system shall define and enforce a coordination structure (e.g., hierarchical planner–specialists or decentralized collaboration) for multi-agent work.

C-07 AGENT STATUS SELF-REPORTING — The system shall require agents to periodically self-report status and progress (e.g., current task, step outcome, next planned action) to improve runtime transparency.

C-08 INTERFACE CONTRACTS AND SCHEMAS — The system shall define input/output contracts and validate request/response schemas at adapter boundaries.

C-09 INVOCATION SAFEGUARDS — The system shall enforce adapter-level safeguards (e.g., timeouts, retries, and idempotency keys) to limit side effects of failed or repeated tool calls.

A.2 Requirements Model



A.3 Requirements SysML v2 Code

```
package MASRequirements {
  private import Requirements::*;

  // --- Subject system referenced by requirements ---
  part def MAS {
    part orchestrator;
    part agents[*];
    part integrations {
      part mcp; // MCP adapter layer
      part a2a; // A2A gateway
      part auditLog; // immutable audit trail
    }
    part planner;
    part compliance;
    part toolProxy;
    part observer;
  }

  // --- Categories used in thesis ---
  requirement def FunctionalRequirement;
```

```
21  requirement def Constraint;
22
23  // --- Functional requirements (FR-01..FR-21) ---
24  requirement def <'FR-01'> ProcessOrchestrationEngine :>
        FunctionalRequirement {
25    doc /* The system shall execute workflow models by dispatching tasks to
          human or software actors according to model control flow and business
          rules. */
26  }
27  requirement def <'FR-02'> ParameterizedSubprocesses :> FunctionalRequirement
        {
28    doc /* The system shall support parameterized subprocesses and rules so
          behavioral variants can be expressed without altering the underlying
          process structure. */
29  }
30  requirement def <'FR-03'> HumanMachineTaskOrchestration :>
        FunctionalRequirement {
31    doc /* The system shall coordinate both human and automated tasks within
          the same orchestrated process model. */
32  }
33  requirement def <'FR-04'> ExceptionDetectionAndRouting :>
        FunctionalRequirement {
34    doc /* The system shall detect deviations from the nominal process and
          route cases to defined exception subprocesses. */
35  }
36  requirement def <'FR-05'> EscalationToHumanAuthority :>
        FunctionalRequirement {
37    doc /* The system shall escalate unresolved or unmodeled exceptions to
          designated human roles with clear ownership. */
38  }
39  requirement def <'FR-06'> InterAgentCommunicationProtocol :>
        FunctionalRequirement {
40    doc /* The system shall specify message formats and interaction rules for
          agent collaboration (e.g., direct messaging or shared memory; event-
          bus/blackboard where asynchronous exchange is required). */
41  }
```

```
42  requirement def <'FR-07'> ToolFailureHandlingInExceptions :>
        FunctionalRequirement {
43    doc /* The system shall validate tool/API outputs and trigger defined
          recovery or fallback steps when invocations fail. */
44  }
45  requirement def <'FR-08'> ConflictResolution :> FunctionalRequirement {
46    doc /* The system shall provide mechanisms for agents to resolve task or
          decision conflicts (e.g., delegation to a planner, negotiation, or
          voting). */
47  }
48  requirement def <'FR-09'> PolicyEngine :> FunctionalRequirement {
49    doc /* The system shall provide a decoupled policy evaluation component
          that can validate, veto, or redirect workflow executions and agent
          actions at runtime. */
50  }
51  requirement def <'FR-10'> RiskBasedApprovalsAndEscalation :>
        FunctionalRequirement {
52    doc /* The system shall require human approval and/or escalation for
          actions or decisions exceeding defined risk or impact thresholds. */
53  }
54  requirement def <'FR-11'> ComplianceMappingAndDriftChecks :>
        FunctionalRequirement {
55    doc /* The system shall map actions and policies to applicable regulations
           or internal rules and perform runtime checks to detect and block
          compliance drift. */
56  }
57  requirement def <'FR-12'> EvidenceBasedDecisionSupport :>
        FunctionalRequirement {
58    doc /* The system shall aggregate relevant data at decision points to
          support evidence-based choices and reduce errors. */
59  }
60  requirement def <'FR-13'> RuleEnforcedDecisionPoints :>
        FunctionalRequirement {
61    doc /* The system shall enforce business rules and role responsibilities
          at decision points to ensure consistent, auditable outcomes. */
62  }
```

```
63  requirement def <'FR-14'> DecisionTraceAndRationale :> FunctionalRequirement
        {
64    doc /* The system shall persist a human-readable decision trace for
          automated or assisted decisions, including inputs, tool/API calls and
          results, and concise rationale summaries. */
65  }
66  requirement def <'FR-15'> EventLoggingPipeline :> FunctionalRequirement {
67    doc /* The system shall instrument the workflow engine and agents to emit
          structured, timestamped events for key actions (e.g., prompts, tool/
          API invocations and results, plan/decision commits, and state changes).
           */
68  }
69  requirement def <'FR-16'> DashboardsAndAlerts :> FunctionalRequirement {
70    doc /* The system shall provide real-time dashboards and alerting for
          observability metrics (e.g., execution latency, error/anomaly rates,
          blocked actions). */
71  }
72  requirement def <'FR-17'> ReplayForPostHocAnalysis :> FunctionalRequirement
        {
73    doc /* The system shall support reconstruction and replay of workflow and
          agent interactions from logged events to enable root-cause analysis
          and explanation of outcomes. */
74  }
75  requirement def <'FR-18'> IntegrationConnectors :> FunctionalRequirement {
76    doc /* The system shall provide connectors to integrate heterogeneous
          applications and data sources required by the workflows. */
77  }
78  requirement def <'FR-19'> AgentToolAdapters :> FunctionalRequirement {
79    doc /* The system shall expose a uniform adapter interface for agents and
          workflows to invoke external tools, APIs, databases, or RPA scripts.
          */
80  }
81  requirement def <'FR-20'> InterOrganizationalInteroperability :>
        FunctionalRequirement {
82    doc /* The system shall support protocol and interface interoperability
          suitable for cross-organizational workflows. */
83  }
```

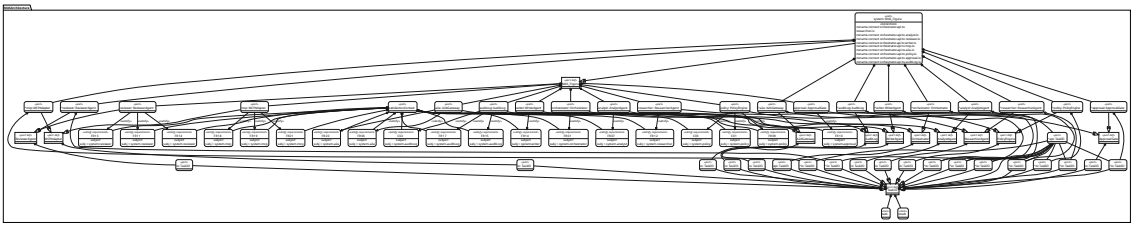
```
84  requirement def <'FR-21'> DataTransformationLayer :> FunctionalRequirement {
85    doc /* The system shall provide mapping and transformation capabilities to
           reconcile data across integrated systems. */
86  }
87
88  // --- Constraints (C-01..C-09) ---
89  requirement def <'C-01'> SegregationOfDuties :> Constraint {
90    doc /* Enforce role-based access control and segregation of duties for
           configuration changes and sensitive actions. */
91  }
92  requirement def <'C-02'> AuditLoggingAndRetention :> Constraint {
93    doc /* Produce tamper-evident audit trails of agent actions, policy
           decisions, and configuration changes, retained according to the
           applicable compliance policy. */
94  }
95  requirement def <'C-03'> BoundedDecisionAutonomy :> Constraint {
96    doc /* Constrain agent decision-making to clearly scoped authority levels
           aligned with organizational objectives. */
97  }
98  requirement def <'C-04'> ModularProcessUnits :> Constraint {
99    doc /* Structure workflows as modular subprocesses or services that can be
           reused and reconfigured without redesign. */
100 }
101 requirement def <'C-05'> InterfaceBasedComposition :> Constraint {
102   doc /* Expose clear process and service interfaces compatible with
          established interoperability standards to enable composition across
          heterogeneous systems and organizations. */
103 }
104 requirement def <'C-06'> ExplicitCoordinationModel :> Constraint {
105   doc /* Define and enforce a coordination structure (e.g., hierarchical
          plannerspecialists or decentralized collaboration) for multi-agent
          work. */
106 }
107 requirement def <'C-07'> AgentStatusSelfReporting :> Constraint {
108   doc /* Require agents to periodically self-report status and progress (e.g
          ., current task, step outcome, next planned action). */
109 }
```

```
110  requirement def <'C-08'> InterfaceContractsAndSchemas :> Constraint {
111    doc /* Define input/output contracts and validate request/response schemas
           at adapter boundaries. */
112  }
113  requirement def <'C-09'> InvocationSafeguards :> Constraint {
114    doc /* Enforce adapter-level safeguards (e.g., timeouts, retries,
           idempotency keys). */
115  }
116
117  // Package-scope usages (for grouping)
118  requirement FR01 : ProcessOrchestrationEngine;
119  requirement FR02 : ParameterizedSubprocesses;
120  requirement FR03 : HumanMachineTaskOrchestration;
121  requirement FR04 : ExceptionDetectionAndRouting;
122  requirement FR05 : EscalationToHumanAuthority;
123  requirement FR06 : InterAgentCommunicationProtocol;
124  requirement FR07 : ToolFailureHandlingInExceptions;
125  requirement FR08 : ConflictResolution;
126  requirement FR09 : PolicyEngine;
127  requirement FR10 : RiskBasedApprovalsAndEscalation;
128  requirement FR11 : ComplianceMappingAndDriftChecks;
129  requirement FR12 : EvidenceBasedDecisionSupport;
130  requirement FR13 : RuleEnforcedDecisionPoints;
131  requirement FR14 : DecisionTraceAndRationale;
132  requirement FR15 : EventLoggingPipeline;
133  requirement FR16 : DashboardsAndAlerts;
134  requirement FR17 : ReplayForPostHocAnalysis;
135  requirement FR18 : IntegrationConnectors;
136  requirement FR19 : AgentToolAdapters;
137  requirement FR20 : InterOrganizationalInteroperability;
138  requirement FR21 : DataTransformationLayer;
139
140  requirement C01 : SegregationOfDuties;
141  requirement C02 : AuditLoggingAndRetention;
142  requirement C03 : BoundedDecisionAutonomy;
143  requirement C04 : ModularProcessUnits;
144  requirement C05 : InterfaceBasedComposition;
```

```
  requirement C06 : ExplicitCoordinationModel;
  requirement C07 : AgentStatusSelfReporting;
  requirement C08 : InterfaceContractsAndSchemas;
  requirement C09 : InvocationSafeguards;

  // Group
  requirement MAS_Spec {
    doc /* Consolidated MAS requirement set for the architecture. */

    require FR01; require FR02; require FR03; require FR04; require FR05;
    require FR06; require FR07; require FR08; require FR09; require FR10;
    require FR11; require FR12; require FR13; require FR14; require FR15;
    require FR16; require FR17; require FR18; require FR19; require FR20;
        require FR21;

    require C01; require C02; require C03; require C04; require C05;
    require C06; require C07; require C08; require C09;
  }
}
```

Listing 1: Requirements model (MASRequirements.sysml)

B.1 Architecture Model



B.2 Architecture SysML v2 Code

```
package MASArchitecture {

  // Use the same requirements you defined in your thesis/model
  private import MASRequirements::*;
```

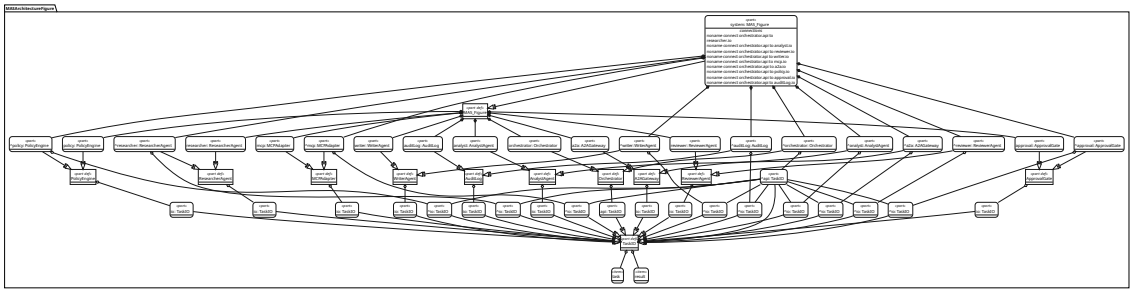
```
5
6   // Tiny, untyped port (same style as your main model)
7   port def TaskIO { in item task; out item result; }
8
9   // --- Elements named exactly like in the thesis ---
10  part def Orchestrator { port api : TaskIO; }
11
12  // specialist roles
13  part def ResearcherAgent { port io : TaskIO; }
14  part def AnalystAgent { port io : TaskIO; }
15  part def ReviewerAgent { port io : TaskIO; }
16  part def WriterAgent { port io : TaskIO; }
17
18  // integration & governance
19  part def MCPAdapter { port io : TaskIO; } // MCP
20  part def A2AGateway { port io : TaskIO; } // A2A
21  part def PolicyEngine { port io : TaskIO; } // FR09 (policy)
22  part def ApprovalGate { port io : TaskIO; } // FR10 (human approval)
23  part def AuditLog { port io : TaskIO; } // observability & audit
24
25  // --- Minimal high-level system (no extra nesting) ---
26  part def MAS_Figure {
27    part orchestrator : Orchestrator;
28
29    // roles
30    part researcher : ResearcherAgent;
31    part analyst : AnalystAgent;
32    part reviewer : ReviewerAgent;
33    part writer : WriterAgent;
34
35    // integration & governance
36    part mcp : MCPAdapter;
37    part a2a : A2AGateway;
38    part policy : PolicyEngine;
39    part approval : ApprovalGate;
40    part auditLog : AuditLog;
41  }
```

```
42
43  // --- One instance + obvious connections only ---
44  part system : MAS_Figure {
45    // Orchestrator coordinates specialists
46    connect orchestrator.api to researcher.io;
47    connect orchestrator.api to analyst.io;
48    connect orchestrator.api to reviewer.io;
49    connect orchestrator.api to writer.io;
50
51    // Orchestrator touches integration / governance / audit
52    connect orchestrator.api to mcp.io;
53    connect orchestrator.api to a2a.io;
54    connect orchestrator.api to policy.io;
55    connect orchestrator.api to approval.io;
56    connect orchestrator.api to auditLog.io;
57  }
58
59  // --- Small, manager-friendly satisfy set (maps to your thesis text) ---
60  part satisfactionContext {
61    // orchestration core
62    satisfy requirement FR01 by system.orchestrator;
63
64    // roles (one-liners, easy to defend)
65    satisfy requirement FR12 by system.researcher; // evidence at decisions
66    satisfy requirement FR07 by system.analyst; // tool failure handling
67    satisfy requirement FR11 by system.reviewer; // compliance mapping & drift
          checks
68    satisfy requirement FR13 by system.reviewer; // rule-enforced decision
        points
69    satisfy requirement FR14 by system.reviewer; // decision trace in review
70    satisfy requirement FR14 by system.writer; // trace carried into outputs
71
72    // governance & risk
73    satisfy requirement FR09 by system.policy; // decoupled policy engine
74    satisfy requirement C01 by system.policy; // segregation of duties
75    satisfy requirement C03 by system.policy; // bounded decision autonomy
76    satisfy requirement FR10 by system.approval; // human approval/escalation
```

```
77
78      // interoperability
79      satisfy requirement FR18 by system.mcp; // integration connectors
80      satisfy requirement FR19 by system.mcp; // agent tool adapters
81      satisfy requirement FR21 by system.mcp; // data transformation
82      satisfy requirement FR06 by system.a2a; // inter-agent protocol
83      satisfy requirement FR20 by system.a2a; // inter-org interoperability
84
85      // observability & audit
86      satisfy requirement FR15 by system.auditLog; // event logging pipeline
87      satisfy requirement FR17 by system.auditLog; // replay
88      satisfy requirement C02 by system.auditLog; // audit logging & retention
89    }
90
91  } // end package
```

Listing 2: Architecture model (MASArchitecture.sysml)

B.2 Architecture Figure



B.3 Architecture Figure SysML v2 Code

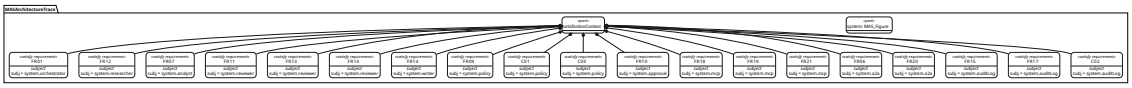
```
1  package MASArchitectureFigure {
2
3    // (No requirements imported here  this file is just the structure for the
         figure)
4
5    // Tiny untyped port (same style as your main model)
6    port def TaskIO { in item task; out item result; }
```

```
7
8   // --- Elements named exactly as in the thesis ---
9   part def Orchestrator { port api : TaskIO; }
10
11  // Specialist roles
12  part def ResearcherAgent { port io : TaskIO; }
13  part def AnalystAgent { port io : TaskIO; }
14  part def ReviewerAgent { port io : TaskIO; }
15  part def WriterAgent { port io : TaskIO; }
16
17  // Integration & governance
18  part def MCPAdapter { port io : TaskIO; } // MCP
19  part def A2AGateway { port io : TaskIO; } // A2A
20  part def PolicyEngine { port io : TaskIO; }
21  part def ApprovalGate { port io : TaskIO; }
22  part def AuditLog { port io : TaskIO; }
23
24  // --- Simple high-level system (no nesting tricks) ---
25  part def MAS_Figure {
26    part orchestrator : Orchestrator;
27
28    // roles
29    part researcher : ResearcherAgent;
30    part analyst : AnalystAgent;
31    part reviewer : ReviewerAgent;
32    part writer : WriterAgent;
33
34    // integration & governance
35    part mcp : MCPAdapter;
36    part a2a : A2AGateway;
37    part policy : PolicyEngine;
38    part approval : ApprovalGate;
39    part auditLog : AuditLog;
40  }
41
42  // --- One instance + obvious connections only ---
43  part system : MAS_Figure {
```

```
    // Orchestrator coordinates specialists
    connect orchestrator.api to researcher.io;
    connect orchestrator.api to analyst.io;
    connect orchestrator.api to reviewer.io;
    connect orchestrator.api to writer.io;

    // Orchestrator interfaces with integration/governance/observability
    connect orchestrator.api to mcp.io;
    connect orchestrator.api to a2a.io;
    connect orchestrator.api to policy.io;
    connect orchestrator.api to approval.io;
    connect orchestrator.api to auditLog.io;
  }

} // end package
```

Listing 3: Architecture figure (MASArchitecture_Figure.sysml)

C.1 Traceability Diagram



C.2 Traceability SysML v2 Code

```
package MASArchitectureTrace {

  // Reuse your clean figure parts + your requirement IDs
  private import MASArchitectureFigure::*;
  private import MASRequirements::*;

  // Reuse the same simple system from the figure
  part system : MAS_Figure;

  // Minimal, jury-friendly mapping (only the key FR/Cs)
  part satisfactionContext {
```

```
12
13      // Core orchestration
14      satisfy requirement FR01 by system.orchestrator;
15
16      // Roles
17      satisfy requirement FR12 by system.researcher; // evidence at decisions
18      satisfy requirement FR07 by system.analyst; // tool failure handling
19      satisfy requirement FR11 by system.reviewer; // compliance checks
20      satisfy requirement FR13 by system.reviewer; // rule-enforced gates
21      satisfy requirement FR14 by system.reviewer; // decision trace during
            review
22      satisfy requirement FR14 by system.writer; // trace carried into outputs
23
24      // Governance & risk
25      satisfy requirement FR09 by system.policy; // policy engine
26      satisfy requirement C01 by system.policy; // segregation of duties
27      satisfy requirement C03 by system.policy; // bounded autonomy
28      satisfy requirement FR10 by system.approval; // human approvals/escalation
29
30      // Interoperability (integration layer)
31      satisfy requirement FR18 by system.mcp; // integration connectors
32      satisfy requirement FR19 by system.mcp; // agent/tool adapters
33      satisfy requirement FR21 by system.mcp; // data transformation
34      satisfy requirement FR06 by system.a2a; // inter-agent protocol
35      satisfy requirement FR20 by system.a2a; // inter-org interoperability
36
37      // Observability & audit
38      satisfy requirement FR15 by system.auditLog; // event logging pipeline
39      satisfy requirement FR17 by system.auditLog; // replay
40      satisfy requirement C02 by system.auditLog; // audit logging & retention
41    }
42  }
```

Listing 4: Traceability mapping (MASArchitecture_Trace.sysml)

C. 3 Traceability